Universiteit
Leiden
The Netherlands

From the root of variation: A metabolomics perspective to plant soil-feedback

Huberty, M.D.

Citation
Huberty, M. D. (2020, November 24). *From the root of variation: A metabolomics perspective to plant soil-feedback*. *NIOO-thesis*. Retrieved from https://hdl.handle.net/1887/138402

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/138402

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/138402 holds various files of this Leiden University dissertation.

Author: Huberty, M.D.
Title: From the root of variation: A metabolomics perspective to plant soil-feedback
Issue date: 2020-11-24

Chapter 5

Herbivory effects on the metabolome of *Taraxacum officinale* monitored in multi analytical platforms and their quality comparison

Martine Huberty, Lina M. Bayona, T. Martijn Bezemer, Young Hae Choi

Manuscript

Abstract:

Organisms interact with a multitude of environmental factors in their surroundings. These interactions are influenced by chemicals within the organisms, which are highly diverse. Conventional targeted chemical analysis can only describe a limited fraction of this diversity. Systems biology offers a suitable framework to explore the influence of these factors on the state of an organism. Metabolomics, in systems biology, aims to monitor the metabolome to reveal the state of organisms. The ultimate goal of metabolomics is the extraction, detection and identification of all metabolites within an organism. To overcome the limitation of a single analytical platform, it was suggested to combine various analytical platforms and integrate them. Prerequisite for this, is a thorough understanding of the individual methods, e.g. coverage and sensitivity of metabolites, which allow to choose the most suitable method. Here we analysed the response of *Taraxacum officinale,* as a model plant*,* towards herbivory with multiple analytical platforms and compared the outcome. With ^{1}H NMR we show that primary metabolites, e.g. sugars and amino acids, changed upon herbivory. Some abundant secondary metabolites such as flavonoids were detected, but identification of individual compounds requires a further step. We then analysed the samples by LC-MS to clarify changes in secondary compounds of plants in response to herbivory. With molecular networking based on LC-MS-MS spectra we found that mainly compounds within the groups of lipids and lipid-like molecules and phenolics were changed. In the last step, we used HPTLC as a supplementary tool to aid identification of metabolites. The metabolite that changed most was isolated and further identified as campesterol by GC-MS. In conclusion, we show that ^{1}H NMR is suitable to investigate the response of plants but mainly for metabolites present in high concentrations. LC-MS can be used to investigate changes in structurally-diverse low concentrated secondary metabolites. HPTLC offers a cost and time effective method to investigate targeted groups of compounds and can be used as preparative tool to concentrate certain compounds which would be present in too low concentrations otherwise. Up to today one optimal method for metabolomics does not exist and different methods give different insides which all together improve our understanding of how the metabolome changes.

Key Words-Ecometabolomics, integration, analytical platforms, HPTLC, ^{1}H NMR, LC-MS, GC-MS.

Introduction

All living organisms continuously interact with others in their surrounding or even within or on their own body. Most interactions are triggered or mediated by chemicals released or contained within an organism (Dyer et al., 2018). Especially for sessile organisms, such as plants, these are far more essential and inescapable. Basically, plants interact through any tissue with their surroundings, both above and belowground, therefore the chemicals involved in plant interactions can also maintain communication within the plant, transmitting signals from below to aboveground tissues and vice versa. These triggered interactions can lead to a whole cascade of physiological changes and influence bioprocessing revealed in genes, proteins and metabolites (Huberty et al., 2020; Zhu et al., 2018) .

As soon as systems biology emerged, attempts were initiated to analyse the profound impact of those interactions. Along with genomics, transcriptomics, and proteomics, metabolomics, as a part of systems biology, has its own contribution to this, aiming at a comprehensive characterisation of the metabolome of an organism at a certain stage in a specific condition (Stobiecki and Kachlicki, 2005). In particular, metabolomics can serve as a powerful tool to unveil environmental effects on organisms by investigating the final products of gene expressions. Despite recent efforts in the development of methodologies, the high diversity of compounds within a metabolome is still one of the biggest challenges in modern metabolomics. Even though the goal of any metabolomics investigation is comprehensive profiling, the outcomes usually contain a paucity of identified metabolites.

In a recent trend of metabolomics, multiple analytical platforms are implemented to address specific biological questions in an integrated manner, admitting that every metabolomics platform has its own advantages and disadvantages in terms of sensitivity and coverage. In general, the most common methods in metabolomics analysis are liquid chromatography coupled with mass spectrometry (LC-MS), gas chromatography coupled with MS (GC-MS) and nuclear magnetic resonance spectroscopy (NMR) (Peters et al., 2018). Each of them has its own realm for detection and identification of metabolites. Despite various positive properties of LC-MS, regarding its sensitivity and powerful analytical features to identify secondary metabolites, its low data-reproducibility, difficulty of ionization of certain metabolic groups, and low reliability of the databases are thought to be barriers to achieve a

real metabolomics goal. Another MS-based technology, GC-MS, has a specialty for the analysis of primary metabolites *vis-á-vis* LC-MS ones. This leads to a preference in the field of metabolomics to use GC-MS as the first choice of method when details of amino acids, sugars and organic acids are of interest.

Inherently, GC-MS works require a derivatization to increase volatility of the samples to go through the columns at high temperature, in which some labile metabolites could be degraded. The greatest drawback of GC-MS is its limitation of molecular weight of metabolites to 600 – 700 MW.

Compared with the MS-based methods, NMR, particularly ^{1}H NMR, covers a broad range of metabolites for detection – especially, every proton in organic molecules no matter the position can be easily detected by ^{1}H NMR. This easiness facilitates any kind of overall metabolic description, especially of organisms for which there is no prior knowledge about the compounds in the organism. However, it is still confined to a rather rough metabolic profiling tool owing to its low sensitivity and large dependence on spectral databases for identification. Although the sensitivity has been improved recently, e.g. by cryo (cold)- and/or micro-probe, in the case of the requirement of spectral database there are not yet reliable databases, particularly for the large number of secondary metabolites. ^{1}H NMR therefore offers a great tool to analyse unknown samples and get a general overview of the metabolome. However, for complex mixtures it often suffers from a strong overlap of signals caused by the complexity of samples. This makes the identification of specific metabolites difficult. An overview of the strengths and weaknesses of all mentioned methods is found in Table 5.1

Table 5.1. Strength and weaknesses of each of the used metabolomic methods.

Method	Strength	Weakness	Metabolic Coverage
^{1}H NMR	• Broad range of compounds can be detected • Absolut quantification possible • High reproducibility	• Identification of compounds in mixtures requires experience • Low sensitivity -high detection limit • Low concentrated compounds (small peaks) often overlapped by other peaks.	• High concentrated metabolites (µmol level)
GC-MS	• Simple identification based on database • High sensitivity	• Samples need to be heated and derivatized • Required internal standards for quantification • Difficulty in the detection of glycosides	• Primary metabolites • Volatiles • Limitation to molecules with MW not over 600-700 MW
LC-MS	• High sensitivity • Simple detection of MW	• Required internal standard for quantification • Low reproducibility • Difficulty to build up database	• Most of secondary metabolites • Limitation of the detection of non-polar metabolites
HPTLC	• Low costs • Parallel analysis • Broad coverage of metabolite detection	• Low resolution • Semi-quantitation	• Most of metabolites can be detected by various chemical reactions

There have been many attempts to combine multi-analytical platforms for metabolomics research. In present-day metabolomics research, there are two kinds of approaches to this attempt: mathematical or statistical integration of different analytical results carried out in parallel analyses and targeting distinct groups of compounds with different methods. Recently, several publications have attempted to combine NMR and MS analysis to obtain a broader coverage of compounds which can be detected and identified. For this, samples were first analysed with ^{1}H NMR, TOCSY, HSQC and then with LC-MS and the data matrixes received by both methods were then combined with statistical tools such as correlation analysis (CoNaM) (Li et al., 2019). One example for these applications is a method called statistical heterospectroscopy (SHY) with which NMR and LC-MS can be combined and this was used in multiple studies to identify biomarkers in biological samples (Crockford Dj et al., 2006). The application could extend to other combinations like GC-MS and LC-MS data which can be fused and analysed as one data matrix (Tikunov et al., 2010). However, this statistical connection of two methods comes with certain limitations. Information might be lost during the process of combining data from different methods and the individual strengths of each method might be missed or diluted. Especially the combination of MS data and NMR data can

be difficult since the sizes of the data matrixes differ strongly. The weight of the MS data on the outcome of the analysis might be higher due to its bigger size compared to NMR data.

Another way to exploit the potential of the different techniques available for metabolomics is to use each platform for the identification of specific compounds and use the individual strength of each method. Previously, (Kusano et al., 2011) assessed the diversity of the metabolome of genetically modified tomato plants and used for primary metabolites detection GC-MS, and for polar secondary metabolites LC-MS and for ionic compounds CE-MS. In another study they used as untargeted approach UPLC-FTMS and as targeted approach NMR for sugar amino acids and organic acids, LC/BFI-MS/MS for acylcamolines glycerophospoholipids sphingolipids and HPLC-UV for phenolics, LC MS/MS for alkaloids to investigate the metabolome of 20 taxonomically related benzylisoquinoline alkaloid-producing plants (Hagel et al., 2015). All these studies showed that the coverage of metabolites identified increases if certain targeted platforms are used.

Previous studies have therefore shown that the combination of platforms can yield a higher coverage however the limitations of these studies are that often only MS based platforms were used and therefore these studies often neglect the potential changes in primary metabolites. In this study we aimed to examine the feasibility of different metabolomics tools for the model plant *Taraxacum offincale.* Although some studies have investigated which method would be best used for which part of the metabolome, for our model species it was still unclear which compounds we could exactly identify by each method and how efficient each method is to discover differences between plants grown when exposed to stress. Herbivory is such a stress that is known to have a strong effect on the metabolome of plants (Papazian et al., 2019) and it therefore represents an ideal treatment to investigate the combinations of the platforms in our laboratory.

In a previous study where we grew plant species in sequential orders, we detected that the metabolome of 12 tested species greatly changed depending on which plant grew in the soil previously and whether the test plant was exposed to herbivory (Huberty et al. 2020). The treatments mainly influenced primary metabolites such as amino acids and sugars inferring that primary metabolites are essential for plant adaptations to these stresses or changes in the soil. Most secondary metabolites could not be attributed to the treatment of soils and herbivory even though major proximate roles of them are linked to a plant's survival and fitness upon

environmental challenges. This might be because a single analytical tool (^{1}H NMR in the case) could pose a barrier to the detection of the low-concentrated secondary metabolites.

Taraxacum officinale was chosen as model plant for the current study since it is globally distributed and occurs in many different ecological systems (CABI, 2020) and it is often used in ecological studies. Furthermore, in our previous study, it was one of the plant species that showed the strongest reaction in its metabolome towards herbivory, and it has a high metabolic diversity and therefore a high potential to react to environmental stresses in different molecular ways.

In this study we investigated the metabolomic changes in *T. officinale* due to herbivory by *Mamestra brassicae*. *Mamestra brassicae* is a generalist herbivore, native to the palearctic and can feed on a wide range of plant families and species (Rojas et al., 2000). It is commonly used in ecological studies.

To get a broad overview of the metabolomic differences between the samples we used ^{1}H NMR. ^{1}H NMR offers a great tool to get an overall impression of the metabolomic changes in the plants and therefore represents a good first step for studies that investigate ecological interactions which influence on the metabolome was not investigated before. If this overall metabolomics approach shows that the metabolomes differ it is worth to further investigate the differences. The next step in this study was then to use methods for a more detailed metabolomics analysis. The samples were analysed with LC-MS/MS for secondary compounds and we used the ^{1}H NMR data to investigate changes in primary metabolites. LC-MS/MS was chosen as a tool to identify secondary compounds due to its high sensitivity. Finally, targeted analysis for more identifications of compounds were carried out. For this we combined LC-MS and GC-MS with HTPLC. HPTLC offers the possibility to split an extract into certain fractions, visually depict differences in the extracts and offers a quantitative and qualitative method for metabolomics. TLC offers many advantages over other metabolomic techniques. The initial costs for the equipment and the running costs are low and since it is possible to run several samples at the same time it offers a short handling time (Morlock et al., 2014). It can split and detect a broad range of chemicals which makes it an ideal tool for metabolomics. Moreover, it can be used as preparative tool to split up an extract and use different fractions for identification with other tools at hand or for biological tests. A range of different metabolomic groups can be visualised with different derivatisations methods (Fichou

et al., 2016). Similar to other techniques, TLC has its disadvantages such as its relatively low resolution and reproducibility. However, in the last years those limitations were minimized by the development of more effective sorbents and an automatization of the crucial steps of the TLC (e.g. sample application, development and derivatisation as well as visualisation). This automated TLC is also known as high-performance thin-layer chromatography (HTPLC) (Salomé-Abarca et al., 2018). With this new approach we display an easy and reproducible way to separate, detect and visualise metabolomic variation of plant samples due to environmental challenges which they face.

One approach to tackle the diversity of metabolites in plants is to combine different chemical platforms to get information about the metabolites within a plant from different methods. Together, liquid chromatography and/or gas chromatography coupled with mass spectrometric detection (MS or MS/MS) and ^{1}H NMR represent a great toolbox to analyse plants metabolomes and offers detection of a broad range of metabolites with suitable resolution and sensitivity (Kim et al., 2010). However, all those methods have limitations and often do not lead to conclusive results since the data matrix is too complex to identify single compounds. In a traditional setup the extracts obtained from experimental plants would be split into different fractions with highly time-consuming preparative work, and it often leads to losses of extract during the process or the extracts would have to be purified of all compounds which might interfere with the analysis. With this study we highlight a comprehensive way to metabolic profiling of this specific interaction of a plant and its herbivore. We also will shed light on the different methods used by comparing their outcome for the same samples and with show the strengths and weaknesses of each method.

Material and Methods

Plant material

The plants were grown in soils in which previously other plants were grown. However, for this chapter we only investigated the effect of herbivory. The soils in which other plant species were previously grown were mixed for each of the plant species. One part of soil with two parts of sterile soil (1:2 conditioned: sterile v/v) were mixed to reduce potential differences in nutrient content of the soils. These mixed soils were then used to fill 2 pots (9 x 9cm; 650 g). *Taraxacum officinale* was grown in all soils in duplicate and one of the replicates was randomly assigned to one of two herbivory levels (absent or present). This led to 36 replicates x 2 treatments = 72 pots. After four weeks of growth all pots were caged with a transparent plastic tube (9 cm diameter, 30 cm height) closed with a mesh on top. We then put one freshly hatched *M. brassicae* in one of the two cages. After 7 days of feeding the larvae were removed and the shoots of all plants were immediately wrapped in aluminium foil and flash frozen in liquid nitrogen and stored at -80 °C.

Insects

M.brassicae (Lepidoptera: Noctuidae) is a generalist herbivore, native to the palearctic and that can feed on a wide range of plant families and species (Rojas, Wyatt, & Birch, 2000). Eggs were received from the Department of Entomology at Wageningen University (Wageningen, The Netherlands).

Metabolomics ^{1}H NMR analysis

The samples were first analysed with ^{1}H NMR. Samples were lyophilized for 5 days and grounded with a Retsch Mixer Mill MM 400 for 4-5 minutes, depending on the structure of the plant material, at a frequency of 30 Hz. These ground powered samples were then aliquoted and weighed (20.21 ± 0.05 mg). To the extracts we added 300 µL of CH_3OH-d_4 (Merck, Darmstadt, Germany) and 300 µL of D_2O buffer with 0.01% TSP to the samples. The vials were then sonicated for 10 minutes and centrifuged for 10 minutes at 13000 ppm in a Heraeus

Pico 17 Microcentrifuge. 250 µL of the clear supernatant was then transferred to a 3mm-NMR tube (Bruker, Karlsruhe, Germany) using a glass pipette.

The ^{1}H NMR spectra were recorded on a Bruker AV-600 MHz NMR spectrometer (Bruker, Karlsruhe, Germany) operating at an NMR frequency of 600.13 MHz. CH_3OH-d_4 was used as internal lock. Each ^{1}H NMR spectrum consisted of 128 scans taking 10 min and 26 sec acquisition time with the following parameters: 0.16 Hz/point, pulse width (PW) = 30° (11.3 µs), Relaxation delay (RD) = 1.5 sec.

To supress the residual H_2O signal, we used a presaturation sequence by low power selective irradiation at the H_2O frequency during the recycle delay. FIDs were Fourier transformed by a line broadening of 0.3 Hz. The spectra were phased manually, baseline corrected and calibrated to TSP at 0.00 ppm, using TOPSPIN (v 3.0. Bruker).

Bucketing was done with AMIX software (v. 3.9.12, Bruker BioSpin GmbH, Reinstetten, Germany) with scaling to total intensity and a bucket width of 0.04 ppm. Due to residual signals of solvents, the regions between 4.70 - 4.9 ppm and between 3.28 - 3.34 ppm were excluded. The ^{1}H NMR data matrix consisted of 246 buckets per analysed sample. The identification of the signals was done by examining the splitting pattern and the chemical shift of signals and comparing it with an internal database and as described in (Kim et al., 2010).

HPTLC analysis

For the HPTLC analysis 20 mg of the samples were weight and extracted with 600 µL of Methanol. Samples were then sonicated for 10 minutes and centrifuged at 13000 ppm for 10 minutes in a Heraeus Pico 17 Microcentrifuge. The supernatant was transferred into HPTLC glass vials. For the application, development, derivatization and the visualization we used an in house CAMAG HPTLC facility consisting of TLC sampler (version 4), Automatic Developing Chamber 2 (ADC 2), derivatiser (version 1.0 AT), TLC plate heater (version III), and TLC visualiser was used (CAMAG, Muttenz, Switzerland). We used Silica HPTLC plates (20 × 10 cm, F254, Merck, Darmstadt, Germany). We spotted 10 µL per sample on the plate in bands of 10 mm. Therefore, we applied 13 samples on one plate with a distance of 2 mm between bands. As a quality control we used a mix of all extracts and applied it in random positions on each plate. Therefore, each replicate was applied on a different plate. With a total

number of samples of 72 this led to 6 plates. The samples were applied with the TLC sampler. The plates were then dried and developed in an automatic devolving chamber with the respective mobile phase. In order to split up polar and non-polar compounds in the extract we used two different mobile phases e.g. Ethyl acetate, acetic acid, formic acid, water (100:11:11:27), Toluene, Ethyl acetate (8:2). The chamber saturation time was automated (± 20 min) and the migration front was at 20 mm from the application point. After the plates had developed, they were sprayed with 2mL of p-anisaldehyde/sulfuric acid (Wagner et al. 1984) with an automated derivatizer. They were then placed on a TLC plate heater at 100°C for 3 minutes. After that images were recorded with a TLC visualiser at 366, 254 nm and white light. Before developing the plates, pictures of the empty plates were taken with the TLC visualiser in order to correct for potential impurities of the plates before the start of the experiment. These impurities were corrected for by visionCATS. The pictures were then exported into rTLC (version 1.0) and the data was extracted according to Fichou et al. (2016). The dimensions for the data extraction were the same as the settings in the CAMAG system. Pixel width was set to 128 units and band alignment and parametric time warping was used. The program extracts values of intensity of the colour of the picture for red blue and green and averages intensities of colour in the grey channel.

HPTLC as preparative method

After multivariate data analysis the HPTLC with the mobile phase toluene-ethyl acetate (8:2) was run again and all the steps were the same except that the plate was not derivatized. Instead the silica was scraped out at the Rfs that were of interest based on the results of the multivariate data analysis. Four bands (Fraction 1: Rf 0.27-0.31, Fraction 2: Rf 0.55-0.57, Fraction 3: Rf 0.62-0.63, Fraction 4: Rf 0.73-0.75). The scrape out silica was stored in a microtube. The HPTLC was repeated 4 more times to achieve a higher concentration of the isolated compounds. The silica was then immediately extracted with 500 µL acetonitrile and sonicated for 30 min. To exclude measuring compounds related to the silica, a control which only included the scraped-out silica after development of the plate without a sample was included and measured in later steps. The extract was then put through a filter (0.02 µm) and stored in a glass vial upon use for the LC-MS. For the identification with GC-MS we dried the samples that were used for GC-MS in the speedVac and rediluted them with 100 µL Pyridine and 100

µL BSTFA and TMCS. For the derivatization the tubes were put in a water bath of 80°C for 50 min.

UHPLC-QTOF-MS

The UHPLC-DAD-MS was performed using a Bruker OTOF-Q II spectrometer with electrospray ionization (ESI). For separation a Kintex C18 column (2.1 x 150 mm, 2.6µm) (Phenomenex, Utrecht, the Netherlands) with a gradient of water (A) and acetonitrile with 0.1% formic acid (B) with a flow rate of 0.300 mL/min. The gradient was 0-29 min 10% B 90%A; 30-31 min 90% B, 31-27 min 98% B, 38 -40 min 10% B. The temperature of the column was 40°C. The injection volume was set at 1 µL. The MS parameters were: Nebulizer gas 2.0 bar, drying gas 10.0 mL/min, gas temperature 250°C, capillary voltage 3500 V. The MS was operated in positive mode and data were collected between 100 and 1650 m/z. Each chromatogram was processed in Brucker Daltonics Profile Analysis version 2.1 and calibrated using the zone between 0.0 and 0.25 minutes. As a calibrant sodium formate solution was used. Data was saved as mzXML file. Data was then processed with XCMS online (Tautenhahn et al., 2012). For the molecular networking, the same samples were run again but in MS/MS detection mode.

Gas chromatography–mass spectrometry (GC-MS)

The derivatized samples were analysed with a 7890A gas chromatograph with 7693 automatic samples coupled to a 5975C mass single-quadrupole detector (Agilent, Folsom, CA, USA). A DB-5 GC column (30 m × 0.25 mm, 0.25 µm film thickness, J&W Science, Folsom, CA, USA) with an ionisation energy in EI of 70 eV. As carrier gas, He (99.9% purity) at a flow rate of 1mL/min was used. The oven temperature started at 60 °C for one minute and increased to 290 °C for 5 minutes with a rate of 7 °C/min. After that it was heated up to 310°C for 3 minutes with a rate of 5°C/min. The injector was put on splitless mode and 280°C. Peaks were identified by comparison of their iron spectra with the NIST library (version 2008) or by comparisons of their retention time and spectra with standard compounds.

Data analysis and statistics

To reduce the variation between plates the HPTLC data extracted was normalised to the signal of the pooled quality control sample for each plate. Orthogonal partial least square-discriminant analysis (OPLSDA) were done with SIMCA-P software (v15.0.2, Umeå, Sweden). As scaling method UV was used expect for the LC-MS data which was scaled with pareto. UV scaling was chosen for ^{1}H NMR since it yielded better fitting models.

To compare the effectivity of the different methods for detecting metabolome differences induced by herbivory we conducted a t-test for every feature/bucket and compared the signal for herbivory and non-herbivory samples. P-values were FDR corrected and the percentage of significant signals was calculated. These analyses were performed in R Studio (R Studio Team, 2016).

A molecular network was created using the online workflow (https://ccms-ucsd.github.io/GNPSDocumentation/) on the GNPS website (http://gnps.ucsd.edu) (Wang et al., 2016). For that all mzXML data were uploaded. The data was filtered by removing all MS/MS fragment ions within ± 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the ± 50Da window throughout the spectrum. The precursor ion mass tolerance was set to 2.0 Da and a MS/MS fragment ion tolerance of 0.5 Da. A network was then created where edges were filtered to contain a cosine score higher than 0.7 and that included more than six matched peaks. Further, edges between two nodes were kept in the network when each of the nodes appeared in each other's respective top 10 of most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks.

To enhance chemical structural information within the molecular network, information from in silico structure annotations from GNPS Library Search, Network Annotation Propogation were incorporated into the network using the GNPS MolNetEnhancer workflow (https://ccms-ucsd.github.io/GNPSDocumentation/molnetenhancer/) on the GNPS website (http://gnps.ucsd.edu) (Ernst et al., 2019; Silva et al., 2018; Wang et al., 2016). Chemical class annotations were performed using the ClassyFire chemical ontology (Djoumbou Feunang et

al., 2016). The definitions of the chemical groups can be obtained from: http://classyfire.wishartlab.com/tax_nodes. The outcome of all molecular networks was processed and visualised in Cytoscape version 3.7.2 (Shannon et al., 2003). For a detailed view of the network it is recommended to refer to the online version which is found under: https://gnps-cytoscape.ucsd.edu/process?task=f971cbc9d4ab42229397ccda76748ab2

Results and Discussion

In this study we investigated the chemical response of *T. officinale* to herbivory by *M. brassicae*. *Taraxacum officinale* grows in a large range of habitats and is common all over the world (CABI, 2020). We therefore choose this ubiquitous species as an example to demonstrate the power of the different metabolomics platforms, which we used. *Taraxacum officinale* is used in traditional medicine to treat diabetes mellitus, hepatic disease, rheumatic arthritis cancer and jaundice (Hfaiedh, Brahmi, & Zourgui, 2016; Liu, Xiong, Ping, Ju, & Zhang, 2010; Warashina, Umehara, & Miyase, 2012). Furthermore it can have anti-inflammatory and anti-oxidant properties (Chen et al., 2012; Hu and Kitts, 2003). Not many compounds have been identified so far in *T. officinale* but it is known to contain phytosterols sesquiterpene lactones, flavonoids and phenolic acids (Choi et al., 2018; Díaz et al., 2018). From previous studies it is known to contain saponins, flavonoids, alkaloids, steroids and phenols (Amin Mir et al., 2013; Schütz et al., 2006). *Taraxacum officinale* has also extensively been studied due to its pharmacologically relevant metabolites. For example, sesquiterpenes such as eudesmanolides, tetrahydroridentin B and taraxacolide-*O*-β-glucopyranoside (Hänsel et al., 1980) have been detected. Furthermore sterols such as , taraxasterol, β -taraxasterol, their acetates and their 16-hydroxy derivatives arnidol and faradiol, - and β -amyrin, β -sitosterol, β -sitosterol- β -d-glucopyranoside and stigmasterol, have been reported (Akashi et al., 1994; Burrows and Simpson, 1938; Hänsel et al., 1980). In roots tissue of *T. officinale* several phenolic compounds (e.g. chlorogenic acid, vanillic acid, caffeic acid) have been detected (Williams et al., 1996). In the shoots of *T. officinale* the most highly concentrate phenolic compounds are caffeic acid esters such as chlorogenic, dicaffeoyltartaric (chicoric acid) and monocaffeoyltartaric acids(Williams et al., 1996). Cumarins have also been identified in the aboveground parts of *T. officinale* (Williams et al., 1996).With this diversity of metabolites, it is an excellent species to demonstrate the potential of different metabolomics platforms.

To investigate the effect of herbivory on the metabolome of *T. officinale* we analysed the foliar extracts with different metabolomic platforms. ^{1}H NMR was used to get a broad overview of the metabolic state of *T. officinale* and to analyse differences in primary metabolites and abundant secondary metabolites. LC-MS was used to target secondary metabolites. HPTLC was used for metabolic fingerprinting and as a preparative tool for LC-MS and GC-MS to separate compounds depending on their properties (Fig. 5.1). With this integration of different metabolomics methods, we aimed to maximize the outcome of the metabolomic analysis.

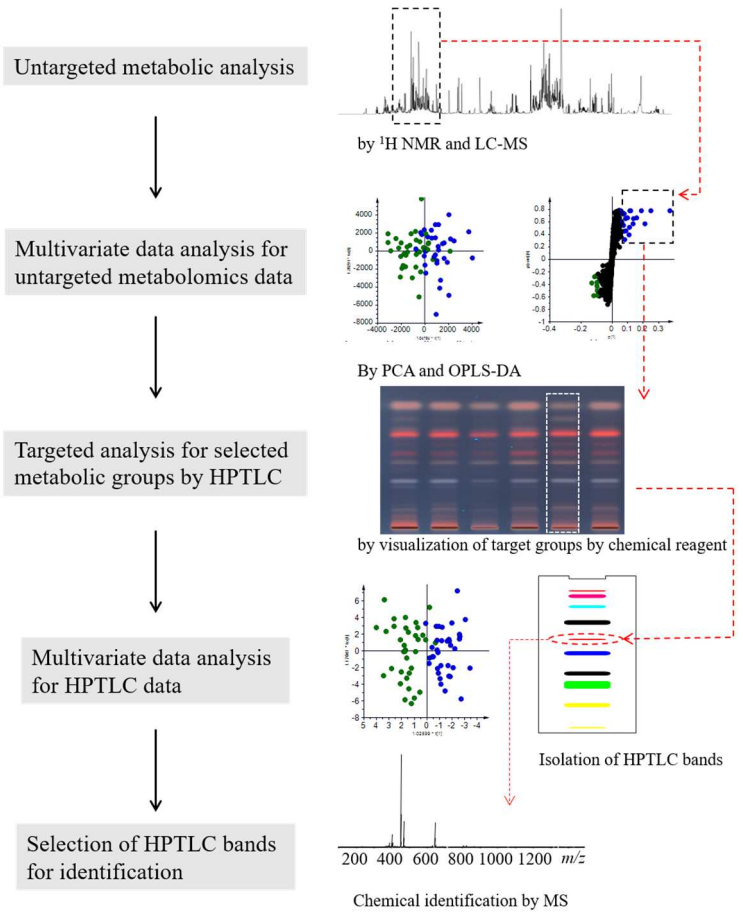

Figure 5.1 Integration plan of metabolomics tools. Figure is adapted from (Bayona et al., 2019)

^{1}H NMR: monitoring overall metabolomic changes

The typical ^{1}H NMR spectrum of *T. officinale* is depicted in Fig. 5.3c and an expansion in the region between δ 4.5-8.0 is found in Fig. 5.2. The most intense signals corresponded to sugars. Of these signals many were overlapping but some characteristic signals for sugars could be detected such as CH-4 for α-glucose δ 3.20 (d, *J* = 3.8 Hz) and CH-1 at δ 5.20 (d), CH-1 for inositol at δ 4.02 (t, *J* = 8.3 Hz), CH-3´ for sucrose at δ 4.16 (d, *J* = 3.9 Hz) and CH-1 at 5.40 (d). As well as amino acids such as threonine at δ 1.48 (d, *J* = 7.3 Hz, 1H) and alanine at δ 1.36 (d, *J* = 6.9 Hz). Furthermore, CH-8´ of chlorogenic acid (3-caffeoylquinic acid) at δ 6.48 (d, *J* = 16.0 Hz, 4H), citrate at δ 2.55 (d, *J* = 13.6 Hz) and malate at δ 2.77 (dd, *J* = 16.4, 4.5 Hz and δ 4.31 (dd, *J* = 6.9, 4.5 Hz,) could be identified. Characteristic signals for flavonoids are found in the region of δ 7-8.

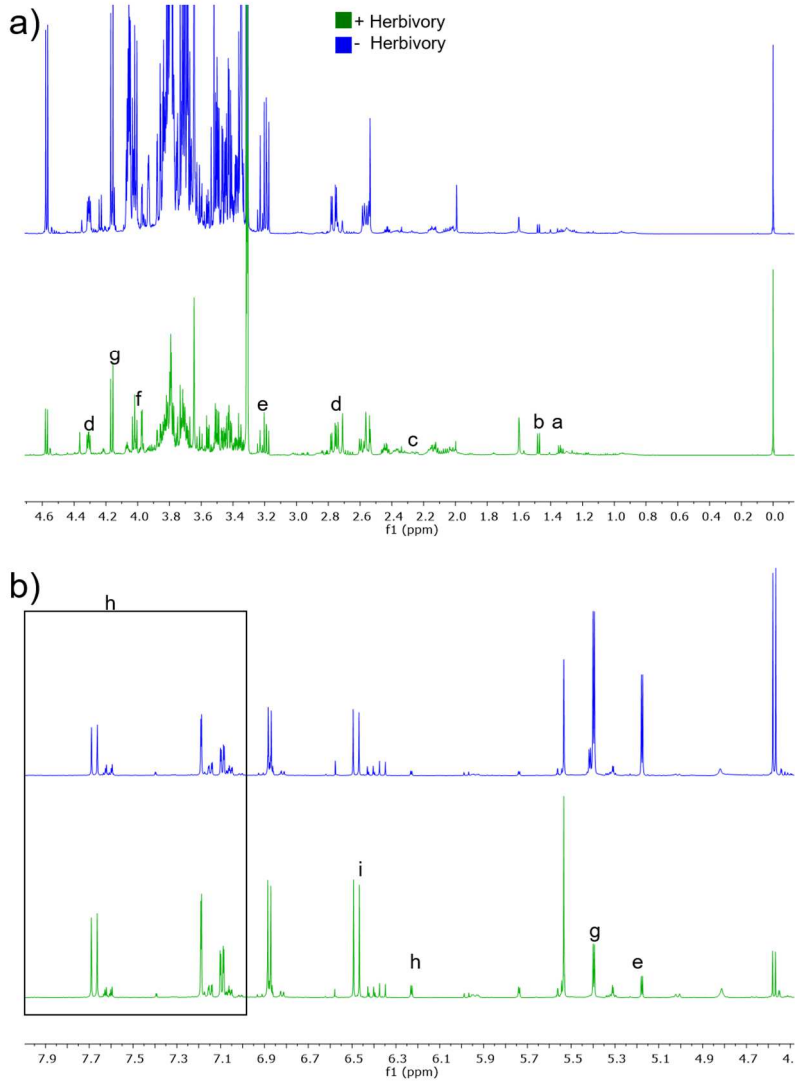

Figure 5.2 Expanded ^{1}H NMR spectra with signals related to a: alanine, b: threonine c: citrate, d: malate e: glucose, f: inositol, g: sucrose, h: flavonoids and i: chlorogenic acid (3-caffeoylquinic acid)

Principal component analysis (PCA) was used to investigate the influence of herbivory on the metabolome measured by ^{1}H NMR. PCA did not show any clear separation between the treatments and therefore we applied orthogonal partial least squares discriminant analysis (OPLS-DA) in the next step. The prediction power of the OPLS-DA was high with a cross-validation Q^2 (0.33) and a significant CV ANOVA ($p<0.001$) when data was UV-scaled. This scaling was used since it provided the best fit of the model. The score plot shows that the

metabolome measured with ^{1}H NMR differs between the two herbivore treatments along the first axis (Fig. 5.3a). To identify the metabolites discriminant for the treatments the S-plot was constructed (Fig. 5.3b). The S- plot showed that alanine, threonine and flavonoids were especially high in plants exposed to herbivory, while glucose, inositol and signals related to the sugar region (δ 3.2- δ 4.0) in ^{1}H NMR were low. The intensity of the signals in 50.07% of the buckets changed significantly upon herbivory.

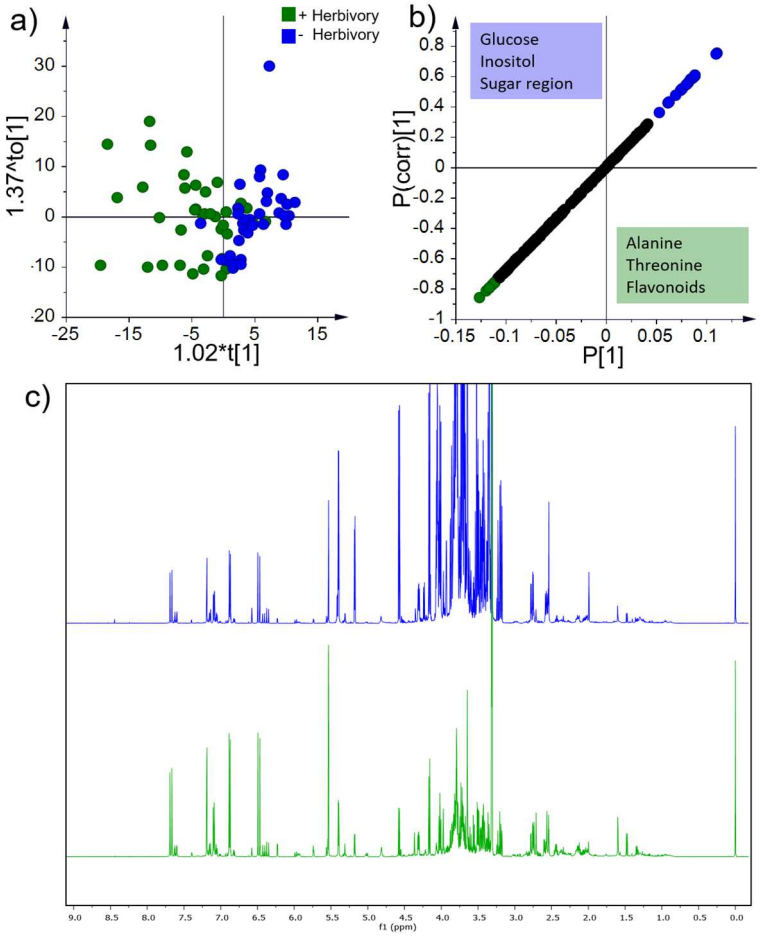

Figure 5.3 Orthogonal partial least squares discriminant analysis of the metabolomics data acquired with ^{1}H NMR. a) the sample plot and b) the S-plot. In a) colours depict the herbivory treatment (Yes/No) and in b) they depict the signals related with the corresponding treatment. Compounds associated with the treatments were tentatively identified. c) typical ^{1}H NMR of extracts from plants with or without herbivory. For expanded spectra in regions δ 0-4.6 and δ 4-8 refer to Figure 5.2

^{1}H NMR fingerprinting provided a broad overview and fingerprint of the metabolomic changes due to herbivory. However, the metabolites that differed between the treatments and that could be identified in the mixtures were all primary metabolites. Signals related to secondary metabolites were detected but could not be identified with ^{1}H NMR. In a biological framework differences in primary metabolites can be important for interactions of the plant with the environment. They were recognized in the last decades to influence herbivore performance especially of specialist herbivores that only feed on one plant (Berenbaum, 1995). For example, inositol concentrations were decreased upon herbivory, and inositol can act as a feeding stimulant for certain insects (Thorsteinson and Nayar, 1963). However, herbivory is known to also change secondary metabolites in plants. We assume that these changes in secondary metabolites might not be well visible with the current method because of potential overlap of signals.

^{1}H NMR is a suitable tool for an overall view of the metabolome changes induced by herbivory and metabolomic fingerprinting. It is a tool which is popular for profiling with the possibility to detect a large range of metabolites. However, ^{1}H NMR is mostly suitable to detect differences of abundant metabolites. The deconvolution of signals in mixtures and with that the identification of single metabolites is difficult. This was also the case in our study in which the identification was hindered by highly congested signals in ^{1}H NMR.

LC-MS: uncovering secondary metabolites

In the ^{1}H NMR analyses we found that flavonoids changed due to herbivory, however it was not possible to identify them in the ^{1}H NMR spectra. We therefore used LC-MS as a next step to analyse the metabolome focusing on secondary metabolites. We analysed all samples with LC-MS in order to achieve a higher coverage of different metabolites that can potentially be identified. Flavonoids were detected in the shoots of *T. officinale* in other studies (Amin Mir et al., 2013; Schütz et al., 2006).

A typical LC-MS chromatogram of a sample from a plant exposed to herbivory and a control plant is depicted in Fig. 5.4c. The PCA of the LC-MS data did not show any clear discrimination between the treatments. Therefore, an OPLS-DA was used. The model was validated with Q^2 =0.312 and a p-value in CV ANOVA lower than 0.01. The plants exposed to herbivory clearly sperate along the first axis in the score plot (Fig. 5.4 a). The S-plot showed

an increase and decrease of metabolites caused by the treatment. However, none of the signals which differed between herbivory and non-herbivory plants (Fig. 5.4 b) were present in our LC-MS library and therefore we were not able to identify the metabolites. 16.56% of the features in LCMS changed upon herbivory and 50 signals (m/z) had a high VIP values (>2.5) in the OPLS-DA.

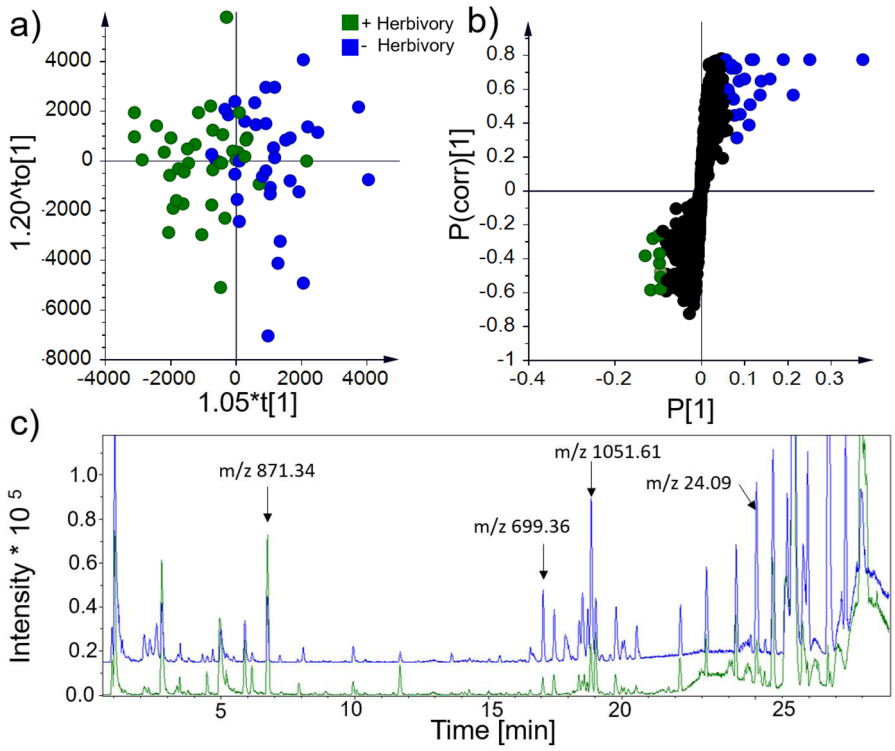

Figure 5.4 Orthogonal partial least squares discriminant analysis of the metabolomics data acquired with LC-MS. a) the sample plot and b) the S-plot. In a) colours depict the herbivore treatment (Yes/No) and in b) they depict the signals related with the corresponding treatment. C) A typical LC-MS chromatogram of plants treated with a herbivore or not. The signals having the strongest effect in b) are depicted in c).

Our results point at one of the major limitations of LC-MS, which is that only a limited number of libraries is available, making the identification of metabolites difficult, especially for plant species for which the metabolome has previously not been studied (Peters et al., 2018). One solution in the future for this problem can be molecular networking (Ramos et al., 2019; Silva et al., 2018; Wang et al., 2016).

A current trend in metabolomics is the use of molecular networks (MN). MN can be used to overcome the often-faced problem of LC-MS that the libraries only contain a small fraction of the detected metabolites. In molecular networks chemical space in MS/MS data is visually displayed. By this, spectra from related molecules (molecular network) can be put together in sets even if the spectras itself do not match with any known metabolite. In molecular networking each spectrum is represented in a node and the connections (edges) represent spectrum alignments. By comparison of the spectra with various databases they can be grouped into chemical groups, or depending on the spectra even identified within the network (Wang et al., 2016). The start of the Global Natural Products Social Molecular Networking (GNPS) in 2016 gave the scientific community the opportunity to share and curate data and use multiple Ms libraries. With that MN can help to increase the efficiency, in terms of identification, of LC-MS data.

LC-MS/MS was performed, and a feature-based network was constructed. The classification of the compounds was done according to ClassyFire. With this network we found that the compounds marked in the loading plot shown in Fig. 5.3b and related to the herbivory treatment mainly belonged to the group of lipids and benzenoids, including phenolics. Within the lipid class we found compounds that were glycerophospholipids, steroids and steroid derivates. In the group of the benzenoids, we detected phenolic glycosides and biphenyls and derivatives (Fig. 5.5). Furthermore, one signal could be associated to one triterpenoid and one trialkylamine. Compounds related to control plants were clustered as lineolic acid or one of its derivates and two glycerophospholipids.

Biologically these results show that there was an activation of plant defences upon herbivore. Triterpenoids are normally stored in a glycosylated form such as saponin. They are known to be defensive compounds against herbivores (Kuzina et al., 2009). Steroids and steroid derivatives were also influence by herbivory. Since they are produced from precursors of terpenoids the may be related to the changes in triterpenoids. Furthermore, a few of the

compounds were identified as glycerophospholipids which are a component of cell walls and they could therefore indicate that cell wall processes change upon herbivory. This has previously been shown for mechanical wounding and herbivory in plants (Galati et al., 2019; Ryu and Wang, 1996).

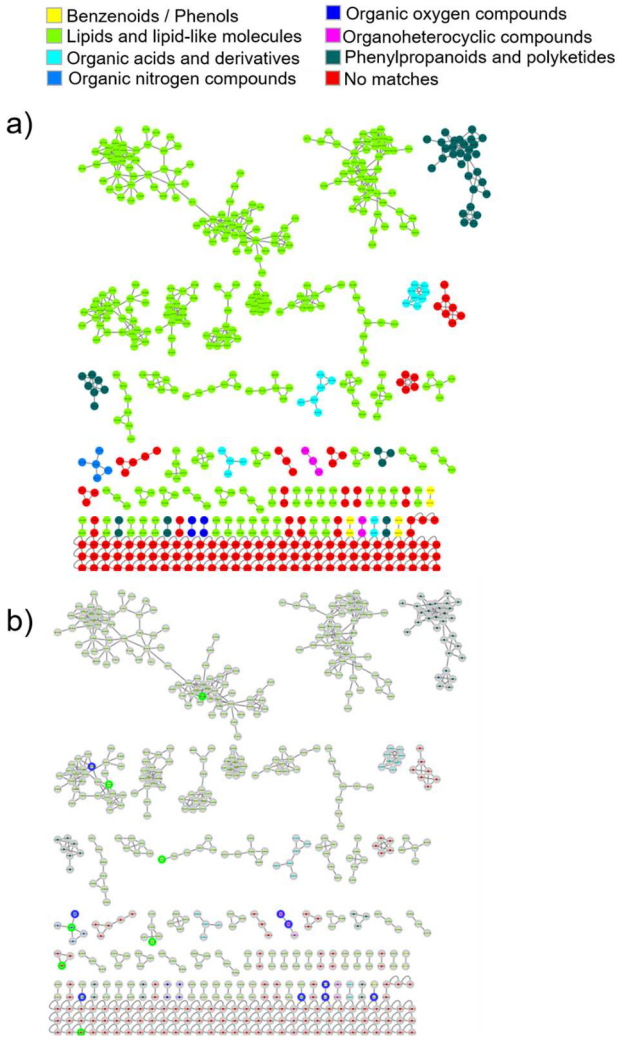

Figure 5.5 Feature based network. a) shows the network and colours depict the chemical class at which the compounds were assigned to. b) shows the same network but the compounds which were shown to differ between the treatments in Figure 5.4b. are depicted by the colour around the nodes. Green represents compounds associated with herbivory. Blue highlights compounds associated with the control plants.

The molecular networking made it possible to relate signals that changed due to the treatments with certain chemical groups. It is a highly promising tool and it is likely that this tool will be used more frequently in future studies. However, with this method it is not possible to identify single compounds.

HPTLC: metabolic fingerprinting and isolation of metabolites.

Traditionally facing many issues HPTLC has recently improved in its reproducibility and standardisation of the use (Audoin et al., 2014). These technical advances and the automatization have made it a suitable tool for metabolomics. It has been used for metabolomic fingerprinting of traditional medical plants (Ogegbo et al., 2012) and for chemical differentiation between different plant species (Maldini et al., 2019; Salomé-Abarca et al., 2018). It has also been used to differentiate between different collection periods of plants (Salomé-Abarca et al., 2018) and to examine the variation of the metabolome of rhubarb roots from different regions (Ge et al., 2018).. T our knowledge, HTPLC has not been used for the differentiation of plants experiencing environmental challenges. We therefore provide here the first study showing that HPTLC can be used as metabolomics fingerprinting platform in the context of ecometabolomics.

An advantage of HPTLC is that it can be used for metabolomic fingerprinting in a simple manner. In contrast to the other platforms it simultaneously offers the possibility to separate mixtures into fractions that can be analysed with other platforms. In order to do so we performed HPTLC with a polar and a non-polar mobile phase to target a large range of compounds with different polarities.

HPTLC offers the possibility to supplement the results from ^{1}H NMR and LC-MS by potentially increasing the concentrations of the compounds and by separating the compounds from the rest of the chemical matrix and thereby decreasing the overlap of signals in ^{1}H NMR. One of the advantages of HPTLC is that it produces highly reproducible data, in contrary to other metabolomics platforms (Jayachandran Nair et al., 2017). Furthermore, many samples can be analysed simultaneously on one plate and it can cover a wide range different metabolic groups that originate from numerous chemical reactions. In our example, the PCAs from the data of the two mobile phases did not show a separation between the treatments. Multivariate data analysis could only be validated for the HPTLC with the mobile phase (toluene: ethyl

acetate) for non-polar compounds (with a Q^2 of 0.31 and a p-value under 0.001 in CV-ANOVA). The metabolome measured with HPTLC did differ between the treatments as shown in the sample plot of the OPLS-DA (Fig. 5.6a). In the next step we identified the compounds responsible for this difference (Fig. 5.6b) by using HPTLC as a preparative tool for other metabolomics techniques (LC-MS and GC-MS). On its own it is difficult to identify compounds with HPTLC but it offers an easy possibility to fractionate extracts and analyse those with other available methods. We analysed four fractions of the HPTLC with a second metabolomics platform in order to identify the compounds in each fraction. A typical representation of a HPTLC chromatogram of one extract from a plant with herbivory and one without herbivory is found in Fig. 5.6c). 3.90% of all signals changed significantly in HPTLC due to the treatment.

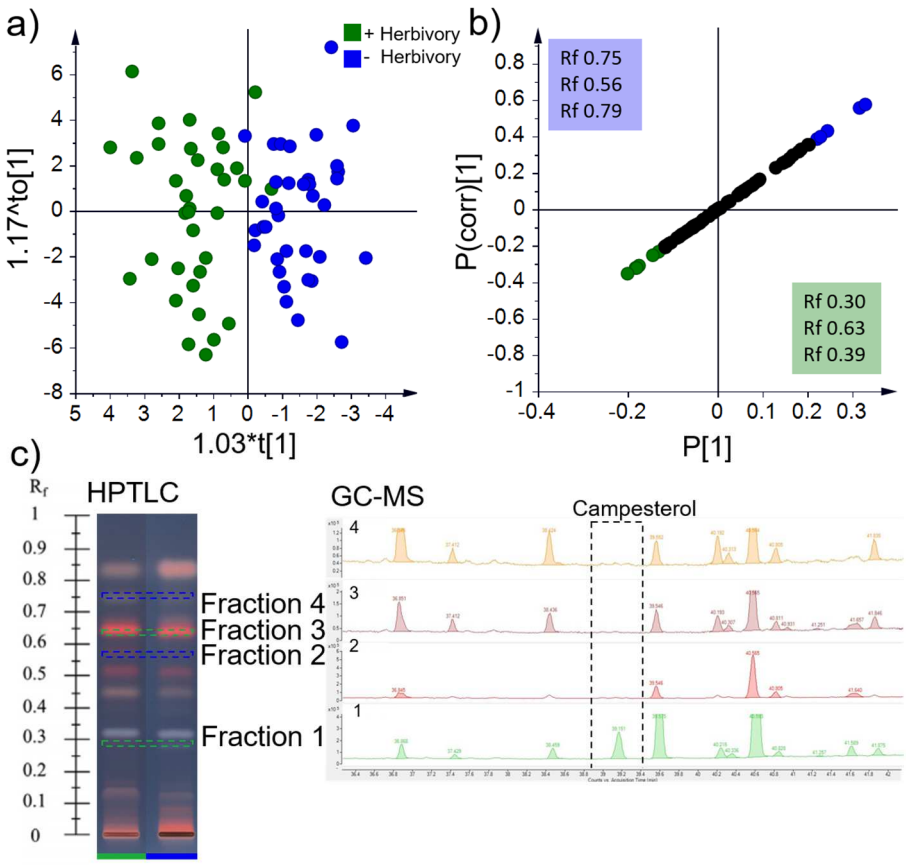

Figure 5.6 Orthogonal partial least squares discriminant analysis of the metabolomics data acquired with HPTLC with Toluene, Ethyl acetate (8:2) mobile phase. a) the sample plot and b) the S-plot. In a) colours depict the herbivore treatment (Yes/No) and in b) they depict the signals related with the corresponding treatment. Rf values for the compounds important for the separation in a) are given in b). c) A typical chromatogram of a high-performance thin layer chromatogram (HPTLC) of *Taraxacum officinalis* developed on a silica gel HPTLC plate using Ethyl acetate, acetic acid, formic acid, water (100:11:11:27) as mobile phase. For visualisation plates were sprayed with anisaldehyde-H_2SO_4 and the picture was taken under 366 nm light. The fraction was analysed by GC-MS separately and the chromatogram of each fraction is shown in c)

We analysed the scraped-out fraction with LC-MS and were able to detect peaks that were not detected in the crude extract, potentially due to a low concentration in the crude extract that we overcame by concentrating the compound through HPTLC. However, we were not able to identify the compounds which we detected in LC-MS. Therefore, we derivatized the same extracts for analysis in GC-MS.

We analysed the fractions of the extract differing between plants exposed to herbivory or not with GC-MS in order to identify the compounds responsible for the separation in the OPLS-DA. In the GC-MS we were able to identify several compounds in Fraction 1, indicating that the sample was not pure. We identified phytosterols including β-sitosterol, stigmasterol and campesterol. With this information we investigated our ^{1}H NMR and LC-MS spectra again and found a signal in the ^{1}H NMR related to sterols (δ 0.8 (s). In the mixture it would have not been possible to identify the specific sterols. However, this was not visible in the LC-MS, probability because of the non-polar properties of the sterols. With that we show that HPTLC can be hyphenated with GC-MS in order to identify compounds that might otherwise would have been missed out on.

We subsequently analysed one crude extract with GC-MS to see if the compounds identified in the fraction are visible in the crude extract. Stigmasterol and β-sitosterol were also found and could be identified in the crude extract. Campesterol could however not be identified in the crude extract because it the quantities in the crude extracts were too low. With the HPTLC step we managed to double the concentration of campesterol and with that concentration we were able to identify this compound.

From a biological view these results are highly interesting since arthropods cannot produce cholesterol de novo (Behmer and David Nes, 2003) and some caterpillars and grasshoppers are known to generate tissue cholesterol by metabolising sterols from plants that they feed on (Behmer et al., 1999). An increase in sterols in the plant would lead to an increase of sterols in the insect since the insect's uptake of sterols is non-selective (Behmer et al., 1999). High concentrations of sterols can lead to insect mortality (Behmer et al., 2011). In our study we now show that herbivory can lead to increased concentrations of sterols in plants. Feeding on the plant could therefore, in the long term, negatively affect the herbivore or even lead to the death of the insect. This needs to be tested in future studies. To our knowledge this is the first report of changes in sterol concentrations in *T. officinale* due to herbivory. This shows that the combination of methods that we propose enabled us to detect and identify compounds that would not have been visible without the HPTLC unless tedious fractionation steps were included.

Conclusion

In this study we show that herbivory by *M. brassicae* influenced the metabolome of *T. officinale*. To achieve a broad cover of different metabolites we combined different metabolomics platforms. ^{1}H NMR showed that herbivory leads to different concentrations of primary metabolites such as sugars and we also detected flavonoids in the ^{1}H NMR. ^{1}H NMR was therefore proven to be a suitable tool for metabolomic fingerprinting and provided an overview of the abundant secondary compounds. However, identifying secondary compounds with ^{1}H NMR still faces many difficulties. Therefore, the secondary metabolites were studied in more depth with LC-MS-MS, and with molecular networking we were able to group the signals from LC-MS into chemical groups. LC-MS-MS together with MN has great potential to make it possible to identify more metabolites in the future, especially since the libraries are getting more exhaustive. However even in LC-MS some metabolites which are only found in low concentrations in the plant, but potentially with a biological effect, cannot be detected. Lipid like metabolites mainly differed between the treatments. We then used HPTLC to combine metabolomic fingerprinting with the fractionation of the complex mixtures. HPTLC enabled us to detect campesterol which was not detected with the other methods. With that HPTLC represents a promising supplementary tool for metabolomics enabling scientists to simultaneously do metabolomic fingerprinting of many samples at once and the fractionation of complex mixtures.

From a biological perspective our results show that sugar concentrations, in general, decrease upon herbivory as has been previously shown in (Van dam and Oomen, 2008; Steinbrenner et al., 2011). Interestingly, we could show in our study by combining different platforms that many lipid-like compounds, such as steroids change upon herbivory.

Metabolomics is a relatively young research field in which the challenges are growing with each new publication. More and more complex systems are analysed, and the rise of ecometabolomics researching ecological interactions with metabolomics increase the need for methods to explore those complex chemistries. A way to tackle this challenge is to use multiple platforms - multiplatform metabolomics. This approach gives us the possibility to exploit the advantages of each method and provides a large coverage of different metabolites. With our study we show how to select and separate compounds influenced by the treatment of interest

in a quick and uncomplicated way by using HPTLC to separate compounds and combining this with LC-MS.

We show that all single methods showed different metabolites which differed depending on the herbivory treatment. This shows that it remains important to use more than on established method for metabolomics. However, if only one method can be used, NMR is probably the best choice to depict overall changes in the metabolome.

Identifications of compounds in mixtures can be tedious and difficult. Our method of combining HPTLC to separate compounds and LC-MS to analyse the compounds afterwards enables the users to identify compounds related to their treatment of interest quicker.

Acknowledgments

We thank Feng Zhu, Robin Heinen, Joost Schrieks, Freddy ten Hooven and Martijn Van der Sluijs for technical support during the greenhouse experiment and for planning this experiment. Further we want to thank the Entomology Department of Wageningen University for providing us *Mamestra brassicae* eggs. We thank Luis F Salomé-Abarca for the assistance with the GC-MS measurements. This study was supported by the Netherlands Organization for Scientific Research (NOW-VICI grant to TMB).