PROCEEDINGS OF IPSC 2019 - 2ND INTERNATIONAL PLANT SPECTROSCOPY CONFERENCE

EDITED BY: Lisbeth Garbrecht Thygesen, Andras Gorzsas and Hartwig Schulz
PUBLISHED IN: Frontiers in Plant Science

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

PROCEEDINGS OF IPSC 2019 - 2ND INTERNATIONAL PLANT SPECTROSCOPY CONFERENCE

Topic Editors:
Lisbeth Garbrecht Thygesen, University of Copenhagen, Denmark
Andras Gorzsas, Umeå University, Sweden
Hartwig Schulz, Julius Kühn-Institut, Germany

Table of Contents

Check for
updates

Editorial: Proceedings of IPSC 2019 - 2nd International Plant Spectroscopy Conference

Lisbeth G. Thygesen[1]*, Andras Gorzsás[2] and Hartwig Schulz[3]

[1] Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark,
[2] Department of Chemistry, Umeå University, Umeå, Sweden, [3] Independent Researcher, Stahnsdorf, Germany

Editorial on the Research Topic

Proceedings of IPSC 2019 - 2nd International Plant Spectroscopy Conference

The 2nd International Plant Spectroscopy Conference took place in Berlin 2018, where 138 experts from 26 countries gathered to exchange new knowledge on the use of many different types of spectroscopy and microspectroscopy for the study and evaluation of plants and plant products within both academia and industry. The current Research Topic comprises 12 highlights from the conference, illustrating the broadness of the field, expanding into chemometrics.

The main part of the contributions illustrates how different types of spectroscopy can be used to explore plant tissues and how they develop with the aim of understanding the numerous ways in which physiology and anatomy are linked. Xiao et al. used X-ray tomography, infrared spectroscopy and Raman imaging to follow the maturation of walnut shells, and found that cell wall thickening and lignification was followed by an accumulation of fluorescent compounds. Piqueras et al. also studied the development of a specialized plant tissue, namely heartwood in a larch species, in this case by use of synchrotron infrared microspectroscopy and atomic force microscope infrared spectroscopy. They found that the heartwood formation observed is consistent with a process where phenolic precursors to extractives accumulate in the sapwood rays, and subsequently are oxidized and/or condensed in the transition zone and spread to the neighboring tracheids in the heartwood. Cuello et al. used infrared imaging to study the cell wall chemistry of different cell types in poplar tension wood, and confirmed that cellulose is abundant in the so-called G-layer of the xylem cells within this tissue, but also found differences within the S-layer of the cell walls. Kendel and Zimmermann used Raman and infrared spectroscopy to compare the chemistry of pollen from more than 200 plant species showing the different strengths of the techniques and demonstrating that they are ideally best used together when studying pollen. In particular, FT-Raman spectra were found to be strongly biased toward the chemical composition of pollen wall constituents, while FTIR over-represented chemical constituents of the grain interior. In the study by Füchtner et al., gas chromatography coupled with mass spectroscopy was employed to study heartwood extractives of a spruce and a larch species with regard to their wood protection properties toward a brown rot fungus. The two tree species were found to rely on two different defense strategies.

Another group of contributions used spectroscopy in studies with an ecological focus, i.e., interactions between plants and their environment. The study by Karimi et al. showed that near infrared spectroscopy coupled with gas chromatography/flame ionization and gas chromatography/mass spectrometry could classify *Zataria multiflora* plants into chemotypes, and that these types could be linked to growth conditions. Diehn et al. used infrared spectroscopy, Raman spectroscopy, surface enhanced Raman scattering, and matrix-assisted laser desorption/ionization mass spectrometry to study pollen samples collected from populations of

the grass *Poa alpine* in three different European countries. The molecular information from the four types of spectroscopy was combined with phenotypical and environmental information of the parent plants and populations. Correlations were generally not strong, but it was found that the mass spectroscopy method could differentiate between pollen from the three populations, while the three other techniques could separate between pollen produced under different growth conditions. In the study by Staková et al., identification and quantification of (groups of) peatland plant species by use of infrared spectroscopy of plant roots was tested with some success.

A third group of studies concerned the use of spectroscopy for quality control of plant-derived products. The study by Yan et al. discussed the use of hand-held spectrometers within herbal medicine, and presented two examples, where a hand-held device based on near infrared spectroscopy was successfully used to differentiate similar species. The study also illustrates that a reliable calibration model is pivotal to such applications. In another study, still within herbal medicine, Brangule et al. successfully tested the use of two different variants of infrared spectroscopy for species identification of herbal products and for separation of leaves from flowers. Still within this group of studies, Kanski et al. used gas chromatography based methods to evaluate the quality of tomatoes after two different storage regimes.

The last contribution included in this Research Topic is an up-to-date review by Bec et al. on the use of spectroscopic imaging for the study of plants.

The wide span of the topics included in the present compilation demonstrates the versatility of different types of spectroscopy both within plant sciences and for fast, non-destructive quality control of plant products, especially when combined with chemometrics. The data analysis approaches used in the studies varied with the spectroscopic techniques and the purpose of the studies, and covered a wide span of multivariate techniques. Naturally, principal component analysis (PCA), the "workhorse" of chemometrics, features heavily, along with classical statistical tools (e.g., t-tests). However, with more directional analyses and especially with combined datasets (from e.g., multimodal approaches), more guided and specialized techniques were skillfully applied, from analysis of variance (ANOVA) to multivariate curve resolution (MCR), the latter being especially useful for improving interpretability and performing dimension reduction in less abstract terms than PCA (working with spectral and concentration profiles instead of scores and loadings).

The works of this special issue highlighted that no single technique (e.g., vibrational spectroscopic, mass spectrometric, or chromatographic data) can be considered the ultimate tool. Each of them has specific advantages and disadvantages (e.g., the versatility of vibrational spectroscopy vs. the specificity of mass spectroscopy vs. the resolution and speed of fluorescence imaging, etc.), and the key for research groups working at the intersection of plant sciences and spectroscopy is to have a versatile toolbox, where each tool has a well-defined purpose and application area in which it excels, and is complemented by other tools in other areas. The challenge (especially for principal investigators) in this respect is to know the tools well enough to select the optimal one for each task, to be able to coordinate a multitool approach and last but not least, to effectively combine the results with scientific rigor and statistical accuracy. As the contributions show, the biannual International Plant Spectroscopy Conference series provides an excellent forum to strengthen skills along these lines, as it promotes the exchange of knowledge and ideas.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

Hand-Held Near-Infrared Spectroscopy for Authentication of Fengdous and Quantitative Analysis of Mulberry Fruits

Hui Yan[1]*, Yi-Chao Xu[1], Heinz W. Siesler[2], Bang-Xing Han[3] and Guo-Zheng Zhang[1]

[1] School of Biotechnology, Jiangsu University of Science and Technology, Zhenjiang, China, [2] Department of Physical Chemistry, University of Duisburg-Essen, Essen, Germany, [3] College of Biological and Pharmaceutical Engineering, West Anhui University, Lu' an, China

Recently, miniaturization of Raman, mid-infrared (MIR) and near-infrared (NIR) spectrometers have made substantial progress, and marketing companies predict this segment of instrumentation a significant growth rate within the next few years. This increase will be based on a more frequent implementation for industrial quality and process control and a broader adoption of spectrometers for in-the-field testing, on-site measurements, and every-day-life consumer applications. The reduction in size, however, must not lead to compromises in measurement performance and the hand-held instrumentation will only have a real impact if spectra of comparable quality to laboratory spectrometers can be obtained. The present communication will, on the one hand, explain the instrumental reasons why NIR spectroscopy is presently the most advanced technique regarding miniaturization and on the other hand, it will emphasize the impact of NIR spectroscopy for plant analysis by discussing in some detail a qualitative and a quantitative application example.

Keywords: hand-held spectrometers, instrumentation, near-infrared (NIR), qualitative and quantitative analysis, authentication of fengdous, nutritional parameters of mulberry fruits

INTRODUCTION

Miniaturization of vibrational spectrometers started more than two decades ago, but only within the last decade real hand-held Raman, MIR and NIR scanning spectrometers have become commercially available and have been utilized for a broad range of analytical applications (Sorak et al., 2012; Guillemain et al., 2017; Crocombe, 2018; Karunathilaka et al., 2018; Soriano-Disla et al., 2018; Vargas Jentzsch et al., 2018). While the weight of the majority of Raman and MIR spectrometers is still in the s1 kg range, the miniaturization of NIR spectrometers has advanced down to the ~100 g level and developments are underway to integrate them into mobile phones (Tino et al., 2016). Furthermore, miniaturized NIR systems have recently reached the <1,000 US$ level. Therefore, only the acquisition of NIR systems can be taken into consideration for private use whereas hand-held Raman and MIR spectrometers will be restricted to industrial, military or homeland security applications and public use by first responders, customs or environmental institutions.

Because of the substantial progress in the miniaturization of near-infrared spectrometers in combination with a drastic cost reduction, marketing experts predict this type of instrumentation a significant growth rate. These trends have made hand-held NIR spectroscopy also attractive for everyday life consumer applications of a new, non-expert user community ranging from food testing to the detection of fraud

and adulteration in a broad area of materials. Notwithstanding this wide-spread application range of hand-held NIR spectroscopy, the focus of this communication will be for plant analytical aspects only. The discussion of a qualitative and a quantitative analytical problem shall serve as examples, to demonstrate the vital role that hand-held NIR spectroscopy will play in the near future for plant analysis.

Before these selected qualitative and quantitative case studies are discussed, however, an overview of the various instrumental features of the most frequently used hand-held NIR spectrometers will be given.

INSTRUMENTATION

The recent progress in miniaturization of hand-held NIR spectrometers has taken advantage of new micro-technologies such as MEMS (Micro-Electro-Mechanical Systems), MOEMS (Micro-Opto-Electro- Mechanical Systems), DMD™ (digital mirror device), or LVFs (Linear Variable Filters) and has led to a drastic reduction of spectrometer size (the weight of the spectrometers discussed in this communication varies between 100 and 200 g) while allowing excellent performance due to the high-precision implementation of essential elements in the final device (Wolffenbuttel, 2005). High-volume manufacturability will further reduce costs and thereby contribute towards broader dissemination of such instruments. In what follows the specific instrumental features of four different hand-held NIR spectrometers will be shortly outlined.

Based on the type of detector, the hand-held NIR spectrometers can be classified in the two categories of array-detector and single-detector instruments (Wolffenbuttel, 2005). Probably the first commercial, real hand-held NIR spectrometer (VIAVI MicroNIR 1700 (formerly JDSU), Santa Rosa, CA, USA) has an array detector that covers the wavelength range from 908 to 1,676 nm and uses an LVF as a monochromator. It has so far been used for a multiplicity of applications ranging from authentication of seafood and determination of food nutrients to the analysis of hydrocarbon contaminants in soil and authentication and quantitative determination of pharmaceutical drugs (Altinpinar et al., 2013; O'Brien et al., 2013; Jantra et al., 2017; Yan and Siesler, 2018b). However, compared to an array detector, the price for a single detector is much lower, and in an attempt to further reduce the hardware costs, new developments focus on systems with single detectors. Thus, the DLP NIRscan Nano EVM (Dallas, TX, USA), for example, is based on Texas Instruments' DMD™ in combination with a grating and a single-element detector and also covers the wavelength range from 900 to 1,701 nm. Very recently a MEMS-based FT-NIR instrument, that contains a single-chip Michelson interferometer with a monolithic opto-electro-mechanical structure has been introduced by Si-Ware Systems (Cairo, Egypt). Contrary to most of the other handheld spectrometers, this instrument can scan FT-NIR spectra over the extended range from 1,298 to 2,606 nm. Finally, Spectral Engines (Helsinki, Finland) developed miniaturized NIR spectrometers, that are based on a tunable Fabry-Perot interferometer. In order to cover the NIR wavelength region 1,350–2,450 nm, however, four spectrometers are required.

The schematic principles of the different monochromator designs of the described NIR spectrometers are summarized in **Figure 1**.

FIGURE 1 | The optical schemes of hand-held NIR spectrometers based on different monochromator principles **(A)** VIAVI MicroNIR 1700, linear variable filter; **(B)** DLP NIRscan Nano EVM with Texas Instruments´ digital micromirror device (DMD™); **(C)** Si-Ware Systems, MEMS-based FT-NIR spectrometer; **(D)** Spectral Engines NIR spectrometer with tunable Fabry-Perot interferometer.
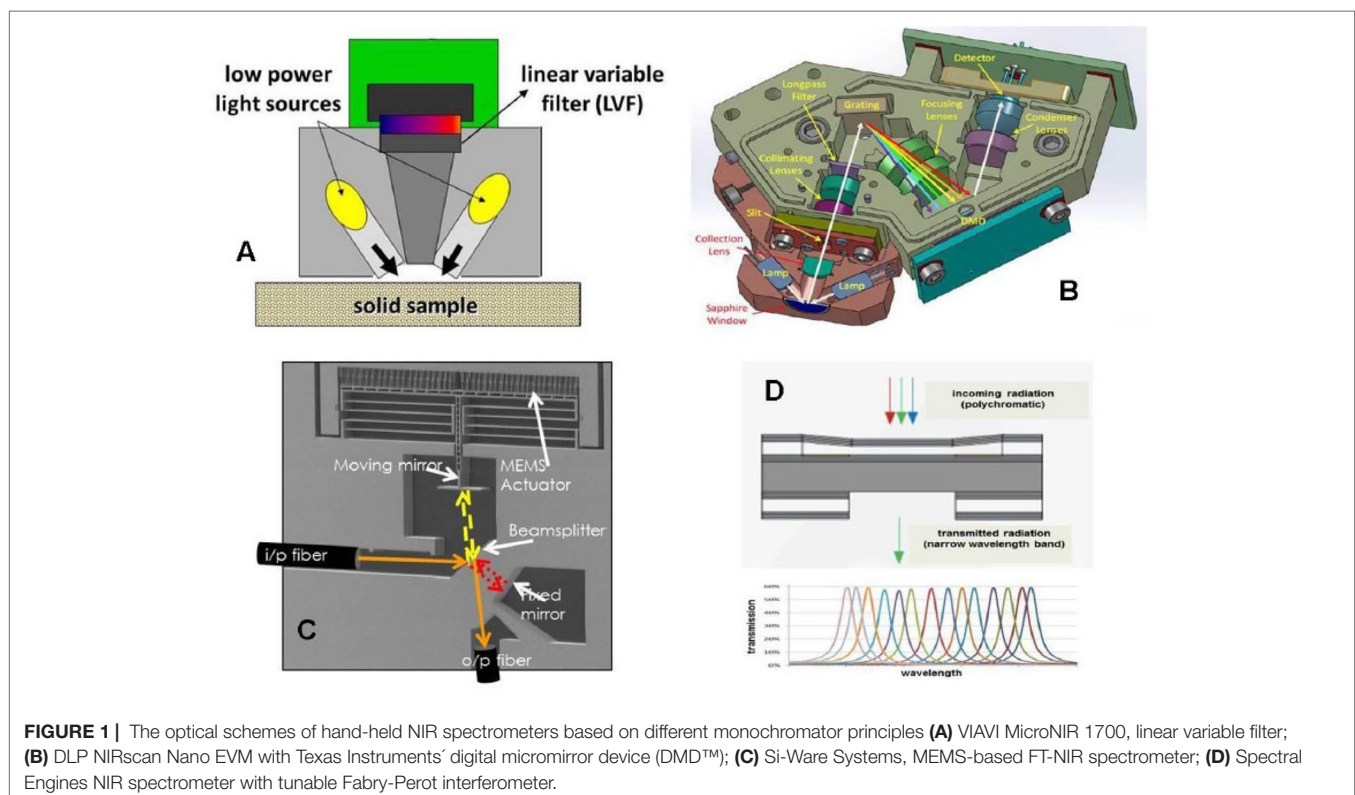
APPLICATIONS

Although the NIR technique is usually applied for a broad range of industrial material quality and control applications (Grassi et al., 2018; Piao et al., 2018; Silva et al., 2018; Yan and Siesler, 2018a; Yan and Siesler, 2018b), the present communication is targeted at practical, everyday life applications in order to attract the attention of a prospective non-expert user community. These days, qualitative and quantitative analysis is more than ever needed also by ordinary people. Because both fraud and adulteration are widely spread, and public health awareness has grown strongly over the last years, the control of nutritional parameters of everyday life food and pharmaceuticals has become an important issue. Therefore, the progress in miniaturization and increasing affordability of hand-held NIR spectrometers make them an attractive tool to fight the above evils efficiently in the public domain.

To demonstrate the potential of hand-held NIR spectrometers for plant analysis, a qualitative and a quantitative application example will be presented here.

Identification of Fengdous

In China, the stem of the Dendrobium is processed into a fengdou (**Figure 2**), that is considered a convenient dosage form of not only a valuable health-care food but also Chinese traditional medicine (TCM) with efficacy in liver protection, treatment of pharyngitis and many other diseases (Chen and Guo, 2001). Fengdou processed from *Dendrobium officinale Kimura et Migo* (DOK) have not only high medicinal value but are also in short supply, and, are very expensive. Therefore, it

would be desirable to discriminate them from fengdou based on *Dendrobium devonianum Paxt* (DDP) with lower efficacy and correspondingly much lower (1/4–1/5) price. However, this is not possible by visual inspection only (**Figure 2**).

Because of the high public interest, an analytical method based on hand-held NIR spectroscopy with the DLP NIRscan Nano EVM system in combination with a partial least squares discriminant analysis (PLS-DA) evaluation method was developed, to rapidly discriminate fengdou processed either from DOK or DDP.

Materials and Methods
Samples
A total of 468 fengdou samples based on DOK (288) and DDP (180) were collected from Luosiwan (Yunnan, China), and the calibration and validation sets were randomly distributed at a ratio of 2:1.

Measurement of Spectra
NIR spectra were collected with the DLP NIRScan Nano EVM spectrometer by accumulating 32 scans in the wavelength range of 909–1,649 nm (209 wavelength variables) in approximately 7 s. After each measurement, the sample was rotated for approximately 120°, and the average of three spectra was then used as the final raw spectrum (**Figure 3**). A certified reflection standard (Labsphere, North Sutton, NA, USA) was used to measure the reference spectrum.

Evaluation of Spectra
Spectral Pretreatment. Due to the fact that NIR spectra frequently contain interferences of background information, drift, and noise, the raw NIR spectra were subjected to spectral preprocessing.

FIGURE 2 | Schematic diagram of the fengdou processing.
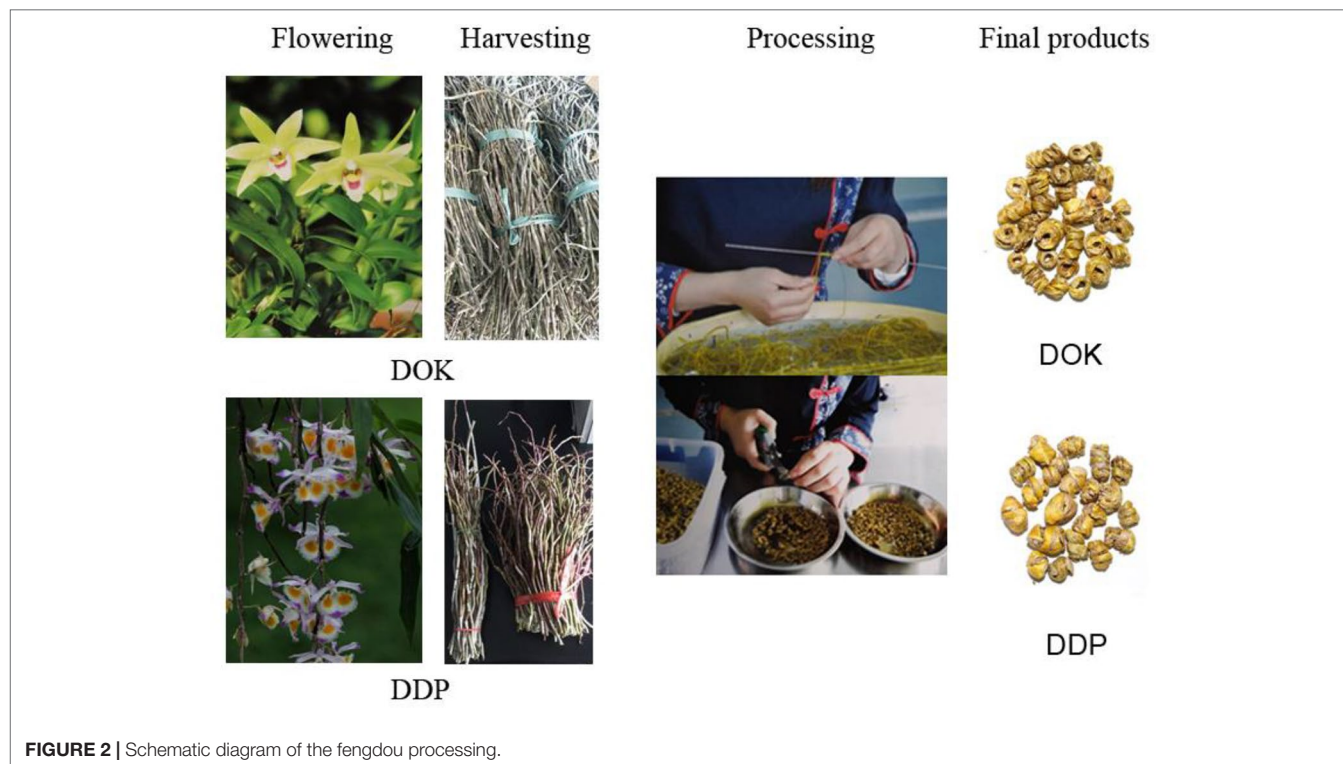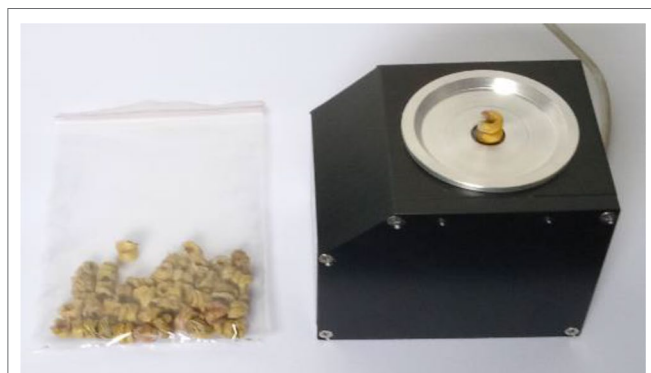
FIGURE 3 | The sample presentation to record NIR spectra of fengdous.

For this purpose, the first derivative based on a Savitzky Golay smoothing procedure with a five data point window and a 2nd order polynomial followed by a standard normal variate (SNV) transformation as a scatter correction was used.

Competitive Adaptive Reweighted Sampling. In NIR spectroscopy, the spectral information is not evenly distributed over the whole wavelength range under investigation. Some data may be superimposed by noise or contain irrelevant information, that can decrease the performance of the calibration models. Therefore, the selection of the informative variables is a significant preprocessing step (Li et al., 2009; Li et al., 2013; Yun et al., 2013). In this work, the competitive adaptive reweighted sampling (CARS), based on the simple but effective principle "survival of the fittest" was applied to select the optimal combinations of spectral variables (Zhang et al., 2015). Compared to the moving window algorithm and Monte Carlo uninformative variable elimination procedure, CARS shows a strong capability of increasing the predictive accuracy (Li et al., 2009). For the present analysis, the CARS was run by the libPLS toolbox (http://www.libpls.net) based on the best combination of pretreated spectra.

PLSDA Analysis
The informative spectral variables determined by CARS were used to develop classification models with the PLS-DA. PLS-DA is a linear classification method that is based on the well-known partial least-squares (PLS) regression. In this work, the leave-one-out (LOO) method was applied to obtain the optimal number of latent variables (LVs) of each model, and the LVs with the lowest root mean square error (RMSE) of cross-validation set (RMSECV) were employed to establish the PLS-DA classification model.

The indices of class accuracy, which are described in the following equation, were calculated to evaluate the performance of each classification model. The higher the accuracy values, the better the predictive ability of the classification model:

$$Class\ accuracy = \frac{Number\ of\ correct\ assignments\ of\ each\ class}{Total\ sample\ number\ of\ each\ class\ tested}.$$

All calculations were performed in MATLAB environment (R2009, Mathworks, Natick, MA, USA) and PLS-DA models were built using the "PLS Toolbox 6.21" from Eigenvector Research (Manson, WA, USA).

Results and Discussion
NIR Spectra
In **Figure 4** the raw NIR spectra of the fengdou calibration set, the mean spectra of all DOK and DDP calibration samples, and the spectra of **Figure 4A** after the different pretreatment steps are shown. As can be seen from the pretreated NIR spectra in **Figures 4C**, **D**, the 1st derivative eliminates most of the baseline shift, whereas the SNV is applied for the scatter correction. The bands at 981 nm, 1,199 nm, and 1,450 nm can be assigned to the 2nd overtones of the N–H, C–H and O–H stretching vibrations, respectively, while the band at 1,568 nm is the 1st overtone of the N–H stretching vibration.

The diagrams of the wavelength optimization variable screening are shown in **Figures 5A–C**. As the number of sampling operations increase, the number of selected wavelength variables decreases first gradually, and then quickly. It embodies the algorithm's ability of an initial rough selection followed by a fine-tuning (**Figure 5A**). The gradual zone for the RMSECV screening process indicates that wavelength variables irrelevant to the type of fengdou were removed, and the growth zone indicates that the essential variables relating to the type of fengdou were excluded. Finally, the trend of the regression coefficient of each wavelength variable in the screening process was achieved. The position of "*" in the figure corresponds to the minimum value of the RMSECV (**Figure 5C**). The 65 selected variables, finally selected for the calibration procedure, are shown in **Figure 5D**.

Identification of DOK
After spectral pretreatment by 1st derivative, SNV and mean centering, the CARS wavelength optimization algorithm was used to filter out the wavelengths with high information, and then the optimized wavelength variables were used to develop a classification model with the PLS-DA method.

The results showed that for the calibration, cross-validation and prediction sets the accuracy is 93.9%, 89.6%, and 84.1%, respectively (**Figure 6A**). As shown by the blue dots (calibration set) and the red dots (test set) in this graph, the samples clearly cluster in two categories and can be readily discriminated. Furthermore, the probabilities of being identified as DOK were calculated and summarized in **Figure 6B**. For the majority of samples, the probability was 1 or 0, which means that these samples were either DOK or DDP. Probability values >0.5 or <0.5 refer to DOK or DDP, respectively.

Sensitivity and specificity are statistical measures of the performance of a binary classification test and are very important for qualitative analysis. Sensitivity (also called the true positive rate) measures the proportion of actual positives that are correctly identified as such. Specificity (also called the true negative rate), on the other hand, measures the proportion of actual negatives that are correctly identified. In this study, for the calibration set, cross-validation set, and test set, the
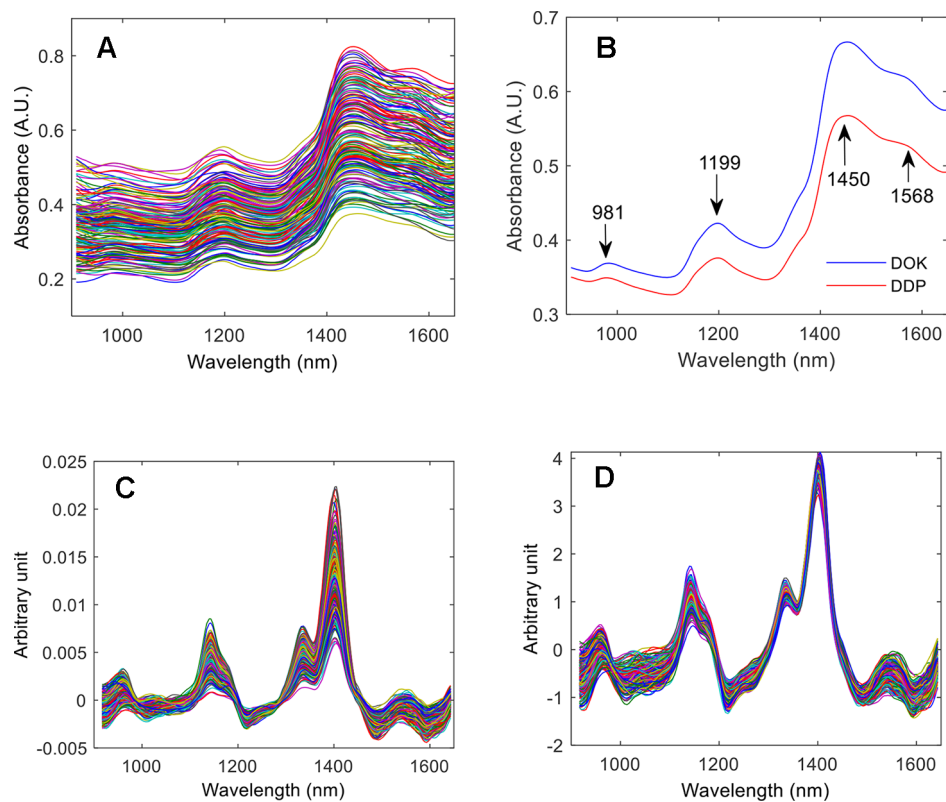
FIGURE 4 | The raw NIR spectra of the fengdou calibration set **(A)**, the mean spectra of the DOK and DDP calibration samples **(B)**, the NIR spectra after pretreatment by the 1st derivative **(C)**, and the NIR spectra pretreated by the 1st derivative and subsequent SNV **(D)**.

FIGURE 5 | Wavelength-variable screening by CARS: **(A)** number of sampled variables versus number of sampling runs; **(B)** RMSECV versus number of sampling runs; **(C)** regression coefficients path versus number of sampling runs; **(D)** wavelength variables selected by CARS.

FIGURE 6 | Classification results of the PLS-DA method: **(A)** DOK classification results of the calibration and test set, **(B)** classification probability of DOK of the calibration samples.

data evaluation, can be utilized for the rapid discrimination of fengdous produced from DOK or DDP.

Quantitative Analysis of Mulberry Fruits

The mulberry fruits have a bumpy surface, and because of the fruits' tightly-packed and seed-bearing ovaries, they have a superficial resemblance to blackberries (Huang et al., 2011). The mulberry fruits are eaten, mostly unprocessed, in their fresh state. As traditional Chinese medicine, the fresh mulberry fruit is used in the treatment of sore throats, fever, hypertension, and anemia (Kamiloglu et al., 2013); they are also used widely in the production of jams, pies, tarts, marmalades, juices, wines, and liquors, natural dyes and in the pharmaceutical, food and cosmetic industry (Huang et al., 2011; Khalifa et al., 2018).

Mulberry fruits contain high nutrient and bioactive contents, including soluble solids content (SSC), polyphenols, flavonoids, ascorbic acid, fatty acids, minerals, and anthocyanin (Lou et al., 2012). The SSC and dry matter content (DMC) are closely related to senses and nutrition. polyphenols and flavonoids (contained in polyphenols) have many pharmacological effects. polyphenols are naturally secreting, and biologically active substances and a wide range of polyphenols are provided by mulberry fruits such as flavanols, phenolic acids, derivatives, and anthocyanins. Polyphenols show activities of antioxidant, detoxification, induction of apoptosis, antiangiogenic and antiproliferation, and so on (Khalifa et al., 2018). Polyphenols in mulberry fruits and their corresponding functionalities vary considerably according to the genetic diversity, climatic, agricultural practices, processing conditions, and stability during storage (Khalifa et al., 2018). Flavonoids are found mostly in glycosylated form, and they have complex flavonol glycosides profiles including 13 quercetin derivatives, five kaempferol derivatives, and O-methylated flavonol-analogs, such as rhamnetin and isorhamnetin. Levels of quercetin glycoside are reported to increase as the fruit ripens from white to black stages (Sánchez-Salcedo et al., 2015). The flavonoids variation in different breeds of mulberries is significant (Sánchez-Salcedo et al., 2015).

Fruit quality has traditionally been determined by visual inspection of the external appearance and its internal content

sensitivities are 0.927, 0.875, 0.896, and the corresponding specificities are 0.950, 0.917, 0.783, respectively.

The sensitivity and specificity derived from the PLS-DA model for the test set samples are represented in **Figure 7**. In **Figure 7B**, the threshold value used to classify the DOK is drawn as a dashed line. With the increase of the threshold value, the specificity increases, i.e., the number of false-positives DECREASES. Likewise, a sensitivity decrease represents the INCREASE of the false-negatives. With the receiver operator characteristic curve (ROC) graph in **Figure 7A** similar information is provided in a different format. The presented results, clearly demonstrate that handheld spectroscopy, combined with CARS-PLS-DA

FIGURE 7 | Specificity and sensitivity of the calibration model: **(A)** predicted ROC, **(B)** predicted responses.

determined by destructive methods, which require operators with the expertise to perform the analysis in a professional laboratory. However, this is impractical for routine analysis by ordinary people. In recent times, consumers have grown conscious of the health benefits of the ingredients of this fruit and a new approach to determine their concentrations is required. In this context it has been reported, that NIR spectroscopy can be used to nondestructively analyze the internal contents, including the SSC, DMC, and total polyphenol content (TPC) of apples (Pissard et al., 2012). Furthermore, Chen et al. employed FT-NIR spectroscopy to determine the TPC in green tea (Chen et al., 2008). In view of this prior knowledge, the demand for a new analytical procedure of mulberry fruits, that will require little to no training originated. In the present work, this issue is addressed by applying the hand-held NIR spectrometer MicroNIR 1700 for a feasibility study of the fast determination of SSC, DMC, polyphenols, and flavonoids in fresh mulberry fruits.

Materials and Methods

Samples
The mulberry varieties applied in this work are Zhongmu 1, 8632, Mengchang 4, and Dashi. A total of 434 mulberry fruits (6–9 maturity) were collected from the conservation of mulberry germplasm resources of the Institute of Sericulture, Chinese Academy of Agricultural Sciences (Zhenjiang, Jiangsu, China).

Measurement of NIR Spectra
As shown in **Figure 8**, NIR diffuse reflection spectra of mulberry fruits were collected with the MicroNIR 1700 spectrometer by accumulating 50 scans with an integration time of 15 ms, and 125 wavelength variables in the range from 908 to 1,676 nm. Triplicate

FIGURE 8 | Presentation of the mulberry fruit for NIR spectra measurement with the MicroNIR 1700.

measurements were made at different spots, and the average of the three spectra was used as the final spectrum of the sample for further processing. The measurements were performed at an environmental temperature of 25 °C and a humidity of about 40%.

Reference Analysis
Determination of Soluble Solids Content. After collection of the NIR spectra, the SSC was determined immediately by a refractometer. First, the equipment was calibrated to zero with distilled water, then the detection surface was dried, and then a few drops of mulberry fruit juice were applied to the detection surface. The juice drops were spread on the prism surface by gently closing the cover of the refractometer, and the corresponding refractive index value was taken.

Determination of Dry Matter Content. The DMC was obtained by measuring the weight percentage of the dried fruit against the corresponding value of the fresh fruit. The weight of the fresh mulberry fruit was measured as m^1, and then the fruit was dried at 65 °C for 24 h and finally dried to constant weight m^2 at 105 °C. The DMC was calculated as DMC (%) = $(m^2/m^1) \times 100$ (%).

Determination of Total Polyphenol Content. The TPC of mulberry fruit was determined by the Folin–Ciocalteau method (Yu and Dahlgren, 2000).

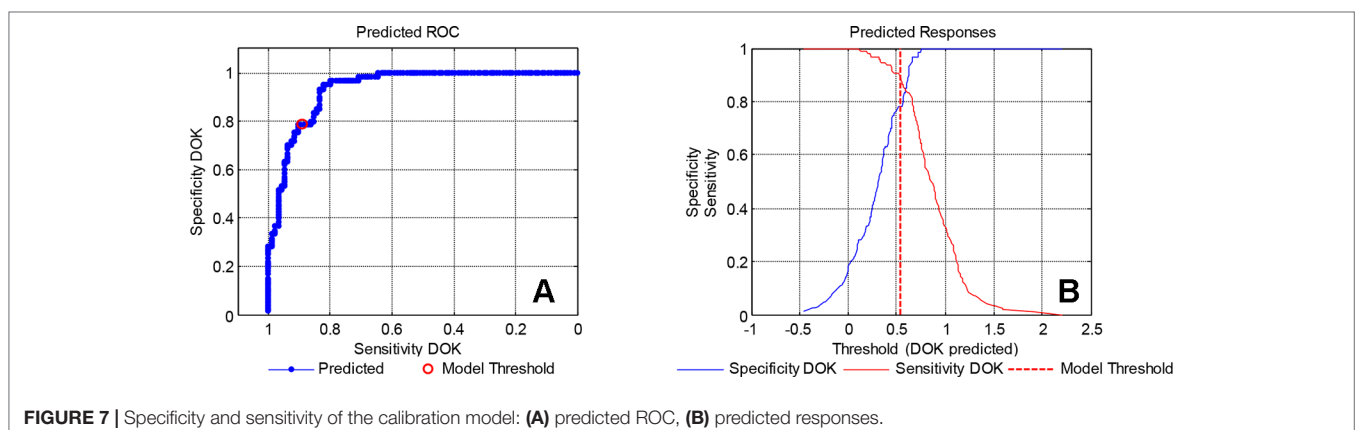**Determination of Total Flavonoid Content.** The content of total flavonoids content (TFC) in the investigated mulberry fruits was measured by colorimetry (Marinova et al., 2005).

Evaluation of Spectra
Spectral Pretreatment. The standard normal variate (SNV) transformation and the 1st derivative based on a Savitzky Golay smoothing procedure with a five data point window and a 2nd order polynomial were applied.

Wavelength Optimization. In this work, two kinds of wavelength selection methods have been applied: genetic algorithm (GA) and CARS. GA is an adaptive search procedure based on the mechanism of genetics and natural selection (Shao et al., 2004; Yan et al., 2011). At first, the GA algorithm randomly generates a population (each individual in the population represents a way of solving the problem) that is composed of a binary string (called chromosome). The bit value "1" represents a selected variable whereas "0" is a variable that is not selected. The fitness of an individual (its ability to adapt to the environment) is calculated; high-quality individuals are retained, low-quality individuals are out. New individuals are generated through inheritance and evolved through natural selection. In this way, eventually, the solution of the problem is achieved. In the present work, the parameters chromosomes 30, mutation 1% and cross-over 50% were adopted in the GA to optimize the variables. The principle of the CARS technique has been described for the previous application example and will not be repeated here.

PLS Calibration. PLS calibration was developed using the PLS toolbox (version 6.21, Eigenvector Inc., Manson, WA, USA), and
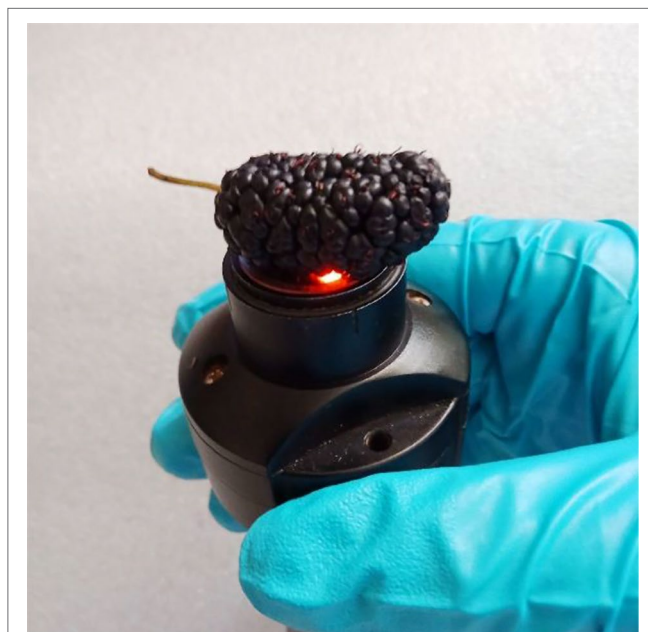
internal cross-validation (CV) was used to select the optimum number of factors. CV estimated the prediction error by splitting all samples into 20 segments, and one segment was reserved for validation, and the remaining (Næs et al., 2002) segments were used for calibration. This process was repeated until all segments were used for validation once.

Calibrations and Validation Statistics. Calibration and validation statistics included the RMSEof calibration set (RMSEC), RMSECV and RMSE of prediction set (RMSEP) and R-squares (Fan et al., 2016). The RMSEC, RMSECV, and RMSEP were used to evaluate the feasibility of the model and its predictive ability. The lower the RMSEP and the closer its value to the RMSEC, the stronger is the prediction ability, and the greater is the robustness of the model. The residual predictive deviation (RPD) defined by the Std Dev/RMSEC of the calibration set was also included to estimate how well the calibration model can predict the compositional data. Generally, an RPD value greater than three can be considered as very good for prediction purposes (Fearn, 2002).

Validation With Unknown Samples
Unknown mulberry fruit samples were collected as a test set to validate the prediction capability of the calibration models developed for SSC, DMC, TPC, and TFC.

Results and Discussion
Reference Values. The reference values of SSC, DMC, TPC, and TFC in mulberry fruits were determined after the spectra were recorded. As shown in **Table 1**, the mean of SSC, DMC, TPC, and TFC were 10.21 Brix, 11.92%, 3.06 mg/g, and 2.26 mg/g, respectively, and the corresponding standard deviation values were 3.16 Brix, 2.26%, 1.25 mg/g and 0.84 mg/g, respectively. The coefficients of variation (C.V.) were 30.96%, 18.94%, 40.95%, and 37.32%, respectively, which suggested that the parameters vary strongly, especially for the TPC and TFC. It is indicative that the collected samples are representative, and the calibration model will show good performance for the determination of unknown samples.

NIR Spectra
The raw NIR spectra of the calibration set are shown in **Figure 9**. The absorption bands at 990 nm and 1,450 nm are related to the 2nd and 1st overtones of the ν(OH) stretching vibration, respectively. The absorption bands from 1,110 nm to 1255 nm belong to the 2nd overtones of ν(CH) stretchng vibrations.

Spectral Pretreatment
Different methods were used to pretreat the spectral data. The spectra pretreated by SNV only, and a combination of SNV + 1st derivative are shown in **Figures 10A**, **B**, respectively, and specifically in the second pretreatment, an accentuation of spectral features can be observed.

The results in **Table 2** show that the pretreated spectra can significantly affect the prediction accuracy of the model. Because the SNV method corrects for scattering effects caused by sample roughness and particle heterogeneity (Yan and Siesler, 2018b) the prediction accuracy of the SSC and DMC calibration models is improved. For the TPC and TFC, the SNV followed by the 1st derivative yielded the best calibration performance. Obviously, besides the scatter correction effect of the SNV, the first derivative contributes spectral features that are beneficial for the calibration
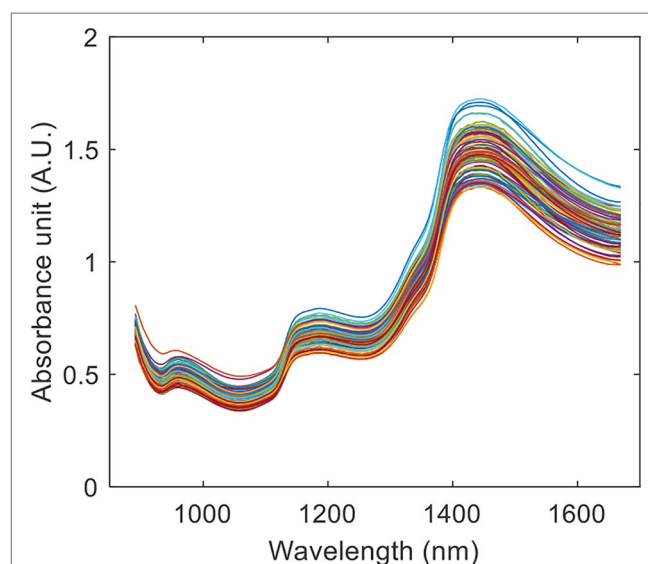
FIGURE 9 | The raw NIR spectra of the mulberry fruit calibration set.

TABLE 1 | Statistical analysis of the reference results of the 4 parameters of mulberry fruits.

Statistical Parameters	SSC (Brix)			DMC(g/g, %)			TPC (mg/g)			TFC(mg/g)		
	Total	Cal.*	Test *	Total	Cal.	Test	Total	Cal.	Test	Total	Cal.	Test
Number	113	76	37	94	63	31	91	61	30	81	54	27
Mean	10.16	10.21	10.07	11.94	11.92	11.87	3.07	3.06	3.09	2.34	2.26	2.51
Max	17.39	17.39	16.37	16.54	16.54	16.43	6.67	6.67	6.48	4.01	4.01	3.98
Min	3.80	3.80	4.00	7.87	7.17	7.87	0.97	1.22	0.97	1.07	1.07	1.09
Range	13.59	13.59	12.37	8.67	9.37	8.55	5.70	5.45	5.51	2.94	2.94	2.90
Std.	3.10	3.16	3.02	2.29	2.26	2.35	1.23	1.25	1.19	0.83	0.84	0.78
C.V.	30.51	30.96	29.94	19.20	18.94	19.81	39.93	40.95	38.50	35.28	37.38	30.97

*Cal and Test stand for calibration and test set, respectively.

FIGURE 10 | Calibration spectra pretreated by SNV **(A)**, and by SNV + 1st derivative **(B)**.
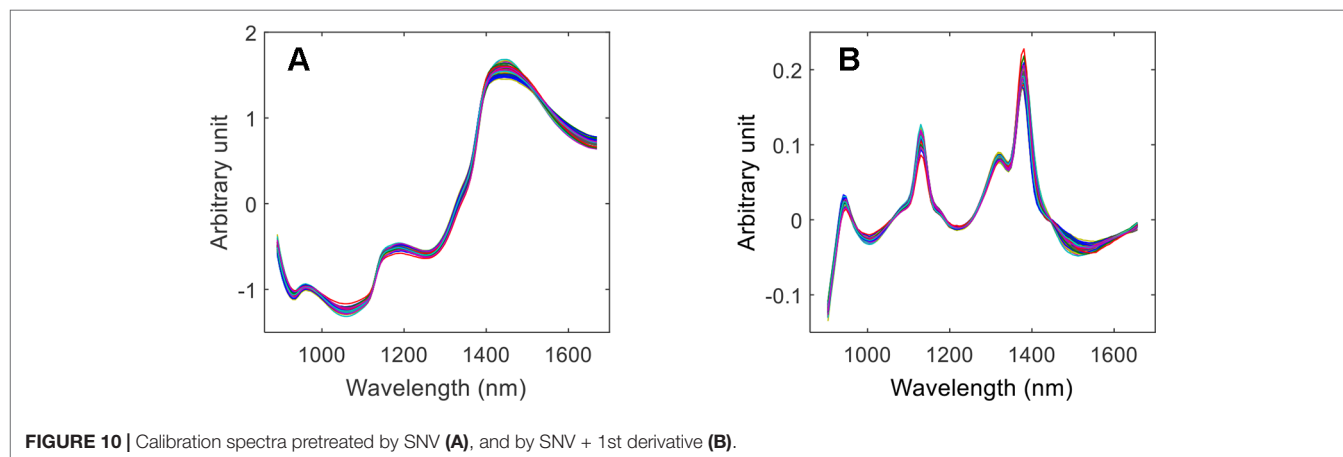
TABLE 2 | The influence of spectra pretreatment methods on the calibration performance (the best calibration results are reproduced in bold numbers).

Parameters	Pretreatment Methods	Factors	R_c^2	RMSEC	R_{cv}^2	RMSECV	R_p^2	RMSEP
SSC	None	9	0.9142	0.9198	0.872	1.1931	0.9001	0.9667
	SNV	**7**	**0.9129**	**0.9266**	**0.8883**	**1.0962**	**0.8891**	**1.0412**
	SNV + 1st	7	0.918	0.8989	0.8867	1.1041	0.8974	1.0168
	1st	7	0.9112	0.9358	0.8751	1.2263	0.8781	1.076
	1st + SNV	6	0.9113	0.9352	0.8856	1.1038	0.8879	1.059
DSC	None	7	0.9016	0.7031	0.8409	0.9695	0.8613	1.0215
	SNV	**7**	**0.9148**	**0.654**	**0.8683**	**0.8806**	**0.9164**	**0.7328**
	SNV + 1st	7	0.9324	0.5825	0.8894	0.8163	0.8962	0.8901
	1st	7	0.9119	0.665	0.8621	0.8918	0.8919	0.9753
	1st + SNV	5	0.8703	0.8072	0.8277	0.9656	0.8647	0.9776
TPC	None	7	0.8176	0.5307	0.7312	0.688	0.8288	0.5764
	SNV	7	0.8675	0.4524	0.8057	0.5917	0.8485	0.5184
	SNV + 1st	**6**	**0.8829**	**0.4253**	**0.8343**	**0.5364**	**0.8385**	**0.537**
	1st	6	0.865	0.4567	0.8165	0.5703	0.8422	0.5568
	1st + SNV	5	0.8301	0.5123	0.7558	0.65	0.8395	0.5722
TFC	None	7	0.7733	0.3977	0.6632	0.52	0.5665	0.5717
	SNV	7	0.8146	0.3596	0.7124	0.48	0.7027	0.4662
	SNV + 1st	**6**	**0.8249**	**0.3494**	**0.7203**	**0.4737**	**0.7364**	**0.4023**
	1st	6	0.8088	0.3652	0.7208	0.4751	0.5864	0.5625
	1st + SNV	6	0.8292	0.3451	0.7431	0.4567	0.7116	0.417

of low-content and complex components (such as the polyphenols and flavonoids).

Wavelength Selection

Figure 11A shows a diagram of the NIR wavelength selection screening for the SSC content that is similar to the previous application example. By the CARS selection, the most sensitive wavelength variables were obtained (see **Table 3**). For SSC, TPC, and TFC, the performance of CARS was better than that of GA. As shown in **Figure 11B** for DMC, 54 variables were selected in 200 runs of genetic algorithms and subsequently used for the development of a PLS model. The different variables selected by these two methods for the four components are shown in **Figure 12**. It is of interest that the variables at about 900 nm, 1,110 nm and in the 1,380–1,440 nm range, selected for TFC are also selected for TPC; the reason maybe that the flavonoids belong to the class of polyphenols and these variables are important for both, TPC and TFC.

Analysis of the Calibration Statistics

The number of optimal factors chosen for a calibration model has a significant impact on its prediction ability. When the number of factors is too low, the model does not entirely reflect the characteristics of the substance, which leads to lower prediction accuracy. Too many factors lead to over-fitting and yield an—apparently—high prediction accuracy. However, when the model is applied to unknown samples, the prediction effect is weak because the model is not robust. Cross-validation was applied to the calibration models with the smallest optimal number of factors. For SSC, DSC, TPC, and TFC, the optimal number of factors are 5,7,5 and 5, respectively. In **Figure 13** the graphs of the RMSEC and RMSECV versus the number of factors are shown for the SSC, DMC, TPC, and TFC. The errors mark the final choice of the optimum number of factors for the individual parameter.

The calibration parameters for the different components are summarized in **Table 3**. Although only nine wavelength
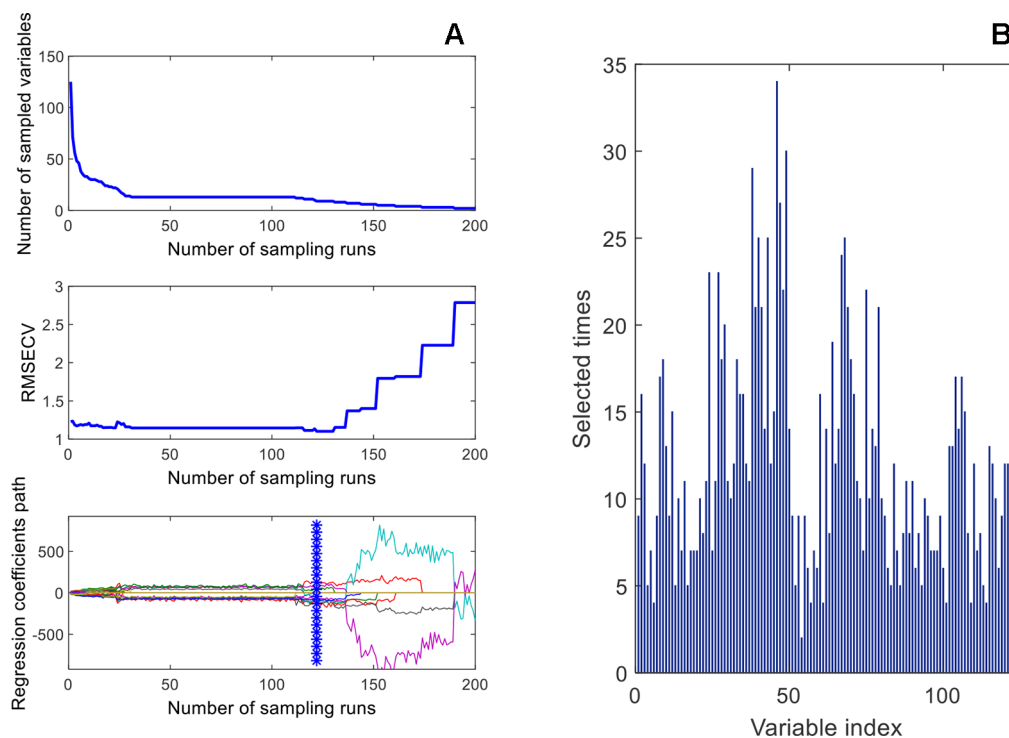
FIGURE 11 | Wavelength-variable screening by CARS **(A)**, and GA **(B)**.

TABLE 3 | Comparison of the impact of the two wavelength selection methods CARS and GA on the calibration performance of the four quality parameters of mulberry fruits (the bold numbers highlight the best calibration results).

Methods	Parameters	Variables	Factors	R_c^2	RMSEC	R_{cv}^2	RMSECV	RPD	R_p^2	RMSEP
Cars	SSC	**9**	**5**	**0.9179**	**0.8998**	**0.8979**	1.0462	3.51	**0.9313**	**0.8843**
	DMC	10	4	0.9036	0.6957	0.8842	0.7841	3.25	0.9194	0.7961
	TPC	**19**	**5**	**0.8989**	**0.3952**	**0.8643**	0.4818	3.16	**0.8651**	**0.4884**
	TFC	**11**	**5**	**0.8154**	**0.3588**	**0.7711**	0.412	2.34	**0.7177**	**0.4061**
GA	SSC	14	7	0.9287	0.8382	0.9046	1.0108	3.77	0.9043	1.0146
	DMC	**54**	**7**	**0.9295**	**0.5950**	**0.8977**	0.7608	3.80	**0.9071**	**0.7758**
	TPC	75	6	0.8942	0.4042	0.8585	0.4916	3.09	0.8642	0.4876
	TFC	27	6	0.7914	0.3815	0.7299	0.4536	2.20	0.7153	0.4097
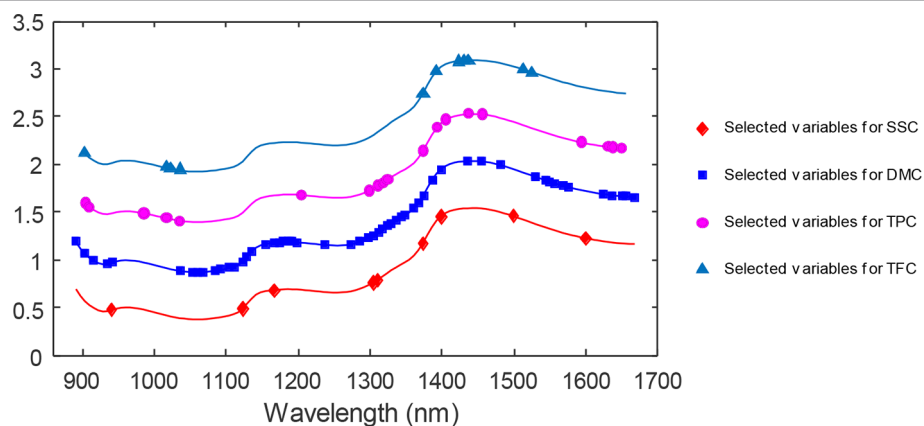
FIGURE 12 | The wavelength variables selected by CARS (◆, ●, ▲) and GA (■) for the four parameters under investigation.
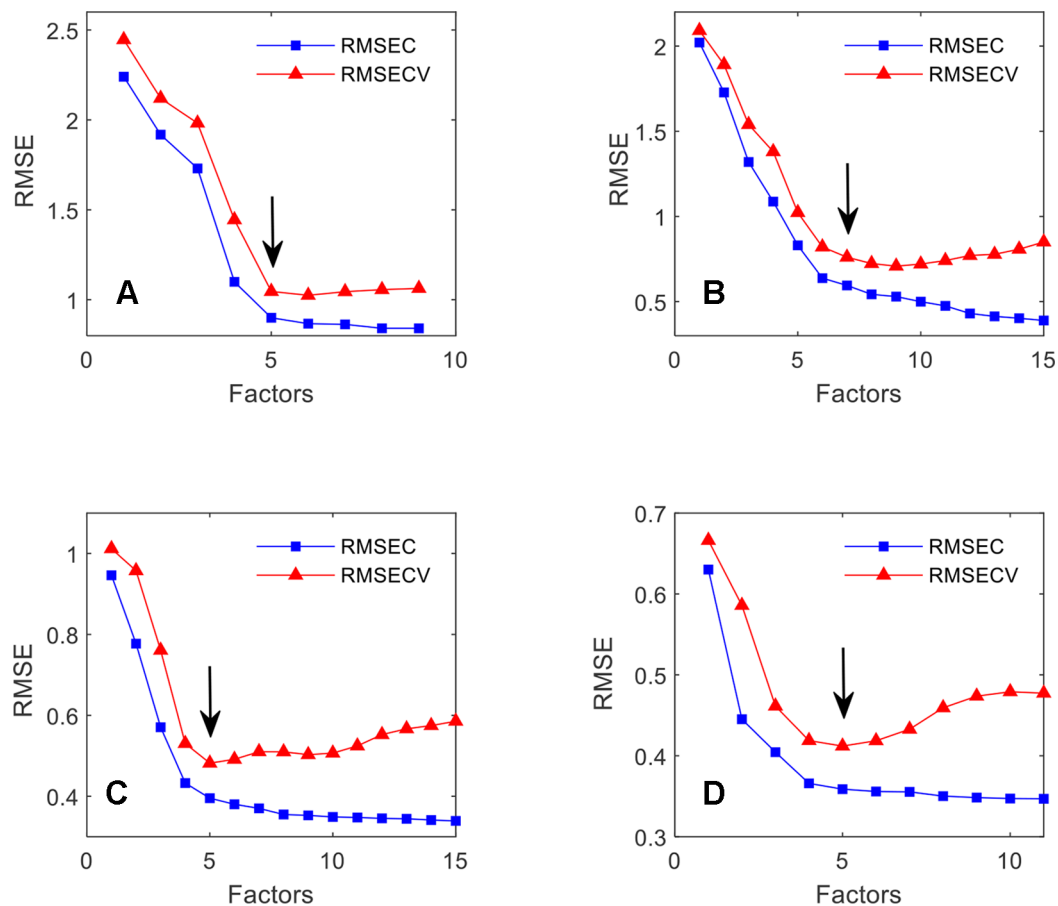
FIGURE 13 | The effect of the number of factors on the RMSEC and RMSECV for SSC **(A)**, DMC **(B)**, TPC **(C)** and TFC **(D)**.

variables were selected for SSC, the calibration performance is the highest. The R_c^2 and R_{cv}^2 are 0.9179 and 0.8979, and the corresponding RMSEC and RMSECV are 0.8998 Brix and 1.0462 Brix, respectively. The high R^2 values and the low RMSEs are characteristic of a good prediction capability. Furthermore, the R^2 and RMSE values for the calibration and cross-validation are similar, which indicates that the calibration model is robust. For DMC, the best calibration is built with the 54 wavelength variables selected by GA. The R^2 values for the calibration and cross-validation are 0.9295 and 0.8977, respectively, and the corresponding RMSEC and RMSECV are 0.5950% and 0.7608%, also suggesting a good calibration performance. However, the robustness is not as good as that of the SSC calibration, because of the larger difference between the statistical parameters of the calibration and the cross-validation. For TPC, 19 wavelength variables were selected for the calibration, and the R^2 values are not as high as that of the DMC calibration. Therefore, the calibration yields results of lower accuracy than the DMC calibration, and furthermore, its robustness is also lower. Finally, the performance of the TFC calibration with 11 wavelength variables is also not as high as that of the TPC component. The R_c^2 and R_{cv}^2 are 0.8154 and

0.7711, respectively, with the consequence of lower calibration accuracy. The RPD values are also included to estimate how well the calibration model can predict the compositional data (Williams and Sobering, 1993; Fearn, 2002). The RPDs for SSC, DMC, TPC, and TFC are 3.77, 3.80, 3.16 and 2.34, respectively, which furnish evidence that SSC, DMC, and TPC can be accurately predicted in the investigated concentration range, whereas, at best, a medium quality calibration has been achieved for TFC.

The scatter plots of the measured versus the predicted parameters are shown in **Figure 14**. In agreement with the previously discussed calibration statistics results, the scatter distances from the regression lines also reflect that proper calibrations have been developed for SSC, DMC and TPC whereas for TFC a comparatively lower calibration performance has been achieved.

Validation With Test Samples
In order to test the performance of the calibrations, a series of test samples (defined as "unknowns" despite available reference values) were used to validate the prediction accuracy. Their calibration statistics results have been summarized in **Table 3**. The R_p^2 for SSC, DMC, TPC and TFC are 0.9313,
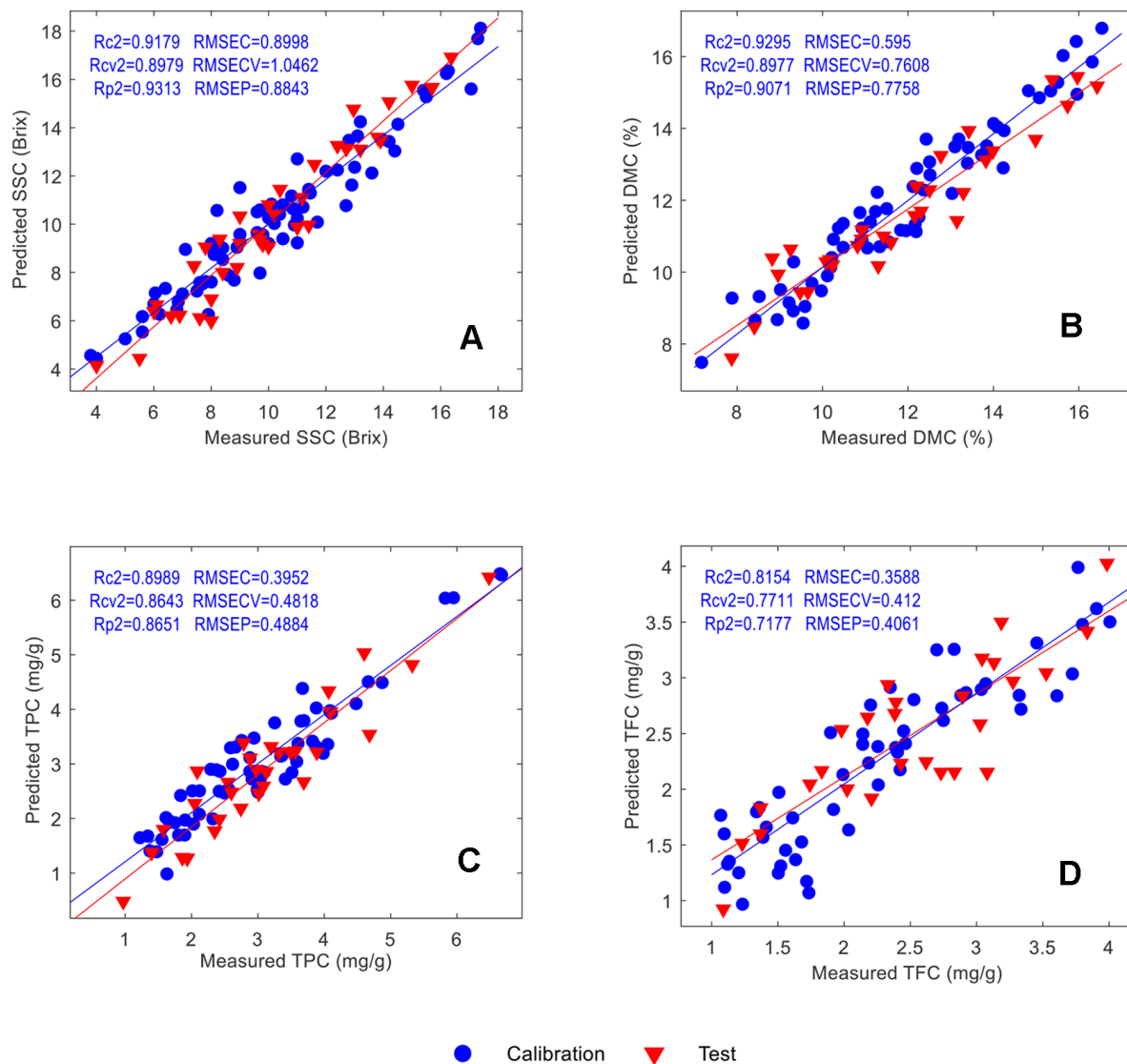
FIGURE 14 | Scatter plots of the measured and predicted parameter values of the calibration and test samples: SSC **(A)**, DMC **(B)**, TPC **(C)** and TFC **(D)**.

0.9071, 0.8651 and 0.9071, respectively, and the corresponding RMSEPs are 0.8843 Brix, 0.7758 %, 0.4884 mg/g and 0.4061 mg/g, respectively. The similar accuracy for the calibration set and cross-validation set suggests that the calibrations are robust. A detailed comparison of prediction and reference results is provided in **Table 4**. In general, for the SSC and DMC, the absolute and relative errors are small, which meets the application requirements. Large relative errors were obtained for the TPC and TFC, but because the absolute errors are small, the calibrations are suitable for screening purposes of consumers, who use a handheld NIR spectrometer to detect whether the mulberry fruits contain a high content of TPC or TFC that is beneficial for the human body.

CONCLUSIONS

Generally, hand-held NIR instruments have launched vibrational spectroscopy into a new era of in-the-field and on-site analysis. In the present communication hand-held NIR spectrometers were applied for qualitative and quantitative plant analytical case studies. In the qualitative example, it was demonstrated that high-value fengdous based on DOK plants can be successfully discriminated from lower quality fengdous of DDP plants. The quantitative application example outlined in detail the assay of the nutritional parameters SSC, DMC, TPC, and TFC of mulberry fruits by hand-held NIR spectroscopy. In both cases, the analysis of the spectroscopic

TABLE 4 | Prediction results for the "unknown" test samples.

Parameters	No.	Measured	Predicted	Absolute Error	Relative Error(%)	No.	Measured	Predicted	Absolute Error	Relative Error(%)
SSC (Brix)	S1	4.00	4.14	0.14	3.57	S20	10.00	10.80	0.80	8.00
	S2	5.50	4.44	−1.06	−19.30	S21	10.20	10.41	0.21	2.06
	S3	6.00	6.37	0.37	6.24	S22	10.40	11.44	1.04	10.00
	S4	6.10	6.65	0.55	9.02	S23	10.00	9.07	−0.93	−9.30
	S5	6.60	6.20	−0.40	−6.09	S24	13.90	13.46	−0.44	−3.17
	S6	6.90	6.23	−0.67	−9.77	S25	11.00	9.87	−1.13	−10.27
	S7	7.40	8.28	0.88	11.89	S26	11.20	11.10	−0.10	−0.89
	S8	7.60	6.12	−1.48	−19.48	S27	11.40	9.96	−1.44	−12.63
	S9	7.80	9.05	1.25	16.05	S28	11.60	12.47	0.87	7.50
	S10	8.00	6.90	−1.10	−13.78	S29	12.40	13.26	0.86	6.94
	S11	8.00	5.99	−2.01	−25.17	S30	12.70	13.14	0.44	3.46
	S12	8.30	9.38	1.08	12.99	S31	12.90	14.76	1.86	14.42
	S13	8.40	8.00	−0.40	−4.81	S32	13.20	13.10	−0.10	−0.76
	S14	8.90	8.20	−0.70	−7.85	S33	13.80	13.60	−0.20	−1.45
	S15	9.00	10.33	1.33	14.82	S34	14.20	15.07	0.87	6.13
	S16	9.00	9.20	0.20	2.21	S35	15.00	15.75	0.75	5.00
	S17	9.60	9.58	−0.07	−0.72	S36	15.70	15.66	-0.04	−0.25
	S18	9.70	9.34	−0.02	−0.20	S37	16.40	16.92	0.52	3.17
	S19	9.80	9.17	−0.63	−6.41					
DMC (%)	D1	7.87	7.61	−0.27	−3.37	D17	12.16	11.57	−0.59	−4.83
	D2	8.40	8.48	0.08	0.90	D18	12.19	12.39	0.20	1.63
	D3	8.82	10.39	1.57	17.79	D19	12.31	11.69	−0.61	−4.98
	D4	8.96	9.94	0.98	10.94	D20	12.51	12.29	−0.21	−1.72
	D5	9.25	10.64	1.39	15.00	D21	12.77	13.24	0.47	3.71
	D6	9.47	9.45	−0.03	−0.29	D22	13.14	11.43	−1.71	−13.04
	D7	9.66	9.45	−0.21	−2.14	D23	13.29	12.21	−1.08	-8.13
	D8	10.09	10.28	0.20	1.96	D24	13.42	13.94	0.51	3.83
	D9	10.17	10.34	0.18	1.74	D25	13.82	13.11	−0.71	−5.15
	D10	10.23	10.20	−0.03	−0.33	D26	13.99	13.37	−0.62	−4.40
	D11	10.82	10.74	−0.08	−0.76	D27	14.99	13.69	−1.29	−8.63
	D12	10.91	10.92	0.01	0.05	D28	15.38	15.35	−0.02	−0.15
	D13	10.93	11.18	0.25	2.28	D29	15.74	14.64	−1.10	−6.96
	D14	11.31	10.17	−1.13	−10.02	D30	15.96	15.44	−0.52	−3.28
	D15	11.42	11.00	−0.42	−3.70	D31	16.43	15.18	−1.25	−7.61
	D16	11.60	10.83	−0.77	−6.63					
TPC (mg/g)	P1	2.05	2.27	0.23	11.01	P16	3.01	2.46	−0.55	−18.29
	P2	1.58	1.80	0.22	14.04	P17	3.08	2.59	−0.49	−15.89
	P3	1.40	1.38	−0.02	−1.38	P18	3.13	2.86	−0.27	−8.61
	P4	1.86	1.28	−0.58	−31.05	P19	3.32	3.22	−0.10	−2.92
	P5	1.94	1.27	−0.66	−34.30	P20	3.49	3.22	−0.27	−7.63
	P6	2.09	2.87	0.78	37.45	P21	3.57	3.24	−0.33	−9.26
	P7	3.20	3.32	0.12	3.71	P22	3.69	2.67	−1.02	−27.65
	P8	2.35	1.77	−0.58	−24.63	P23	3.89	3.22	−0.67	−17.29
	P9	2.42	1.99	−0.43	−17.85	P24	4.07	4.34	0.27	6.66
	P10	2.55	2.66	0.11	4.41	P25	4.10	3.96	−0.14	−3.44
	P11	2.60	2.48	−0.12	−4.67	P26	4.60	5.03	0.43	9.38
	P12	2.74	2.19	−0.56	−20.28	P27	4.68	3.54	−1.14	−24.37
	P13	2.79	3.39	0.60	21.58	P28	5.33	4.82	−0.51	−9.54
	P14	2.88	3.11	0.22	7.78	P29	6.48	6.42	−0.06	−0.97
	P15	2.97	2.90	−0.07	−2.43	P30	0.97	0.48	−0.49	−50.44
TFC (mg/g)	F1	1.74	2.05	0.30	17.46	F15	2.90	2.84	−0.06	−2.00
	F2	1.98	2.53	0.55	27.98	F16	3.04	3.17	0.13	4.39
	F3	3.13	3.14	0.01	0.17	F17	3.84	3.42	−0.42	−10.92
	F4	2.18	2.64	0.47	21.52	F18	1.37	1.83	0.46	33.80
	F5	2.21	1.92	−0.29	−13.04	F19	2.02	2.00	−0.02	−1.04
	F6	2.33	2.94	0.61	26.17	F20	2.73	2.15	−0.58	−21.23
	F7	2.39	2.78	0.39	16.30	F21	3.19	3.50	0.31	9.76
	F8	2.43	2.23	−0.20	−8.08	F22	3.28	2.97	−0.31	−9.37
	F9	2.62	2.25	−0.37	−14.23	F23	1.37	1.60	0.23	16.66
	F10	2.83	2.15	−0.68	−24.00	F24	1.09	0.92	−0.16	−15.07
	F11	3.03	2.59	−0.44	−14.58	F25	2.38	2.68	0.30	12.45
	F12	3.08	2.15	−0.93	−30.16	F26	1.83	2.17	0.33	18.30
	F13	3.53	3.04	−0.48	−13.67	F27	1.23	1.51	0.28	22.99
	F14	3.98	4.03	0.04	1.06					

data was performed with chemometric evaluation routines in combination with wavelength selection methods.

Although the measurement and evaluation routines have not yet reached the convenience for public use by a non-expert user community, the integration of NIR spectrometers into mobile phones and the development of apps for specific analytical procedures in food, plant and material quality control will significantly change the every-day-life of consumers in the near future.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

HY: Investigation, Data curation, Methodology. Y-CX: Investigation, Data curation. HS: Methodology, Supervision. B-XH: Funding acquisition, Investigation. G-ZZ: Funding acquisition, Investigation.

REFERENCES

Altinpinar, S., Sorak, D., and Siesler, H. (2013). Near infrared spectroscopic analysis of hydrocarbon contaminations in soil with a hand-held spectrometer. *J. Near Infrared Spectrosc.* 21 (6), 15–16. doi: 10.1255/jnirs.1079

Chen, X., and Guo, S. (2001). Advances in the research of constituents and pharmacology of Dendrobium. *Nat. Prod. Res. Dev.* 11 (3), 70–75. doi: 10.16333/j.1001-6880.2001.01.020

Chen, Q., Zhao, J., Liu, M., Cai, J., and Liu, J. (2008). Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms. *J. Pharm. Biomed. Anal.* 46 (3), 568–573. doi: 10.1016/j.jpba.2007.10.031

Crocombe, R. A. (2018). Portable spectroscopy. *Appl. Spectrosc.* 72 (12), 1701–1751. doi: 10.1177/0003702818809719

Fan, S. X., Zhang, B. H., Li, J. B., Huang, W. Q., and Wang, C. P. (2016). Effect of spectrum measurement position variation on the robustness of NIR spectroscopy models for soluble solids content of apple. *Biosys. Eng.* 143, 9–19. doi: 10.1016/j.biosystemseng.2015.12.012

Fearn, T. (2002). Assessing calibrations: SEP, RPD, RER and R2. *NIR News* 13 (6), 12–13. doi: 10.1255/nirn.689

Grassi, S., Casiraghi, E., and Alamprese, C. (2018). Handheld NIR device: a non-targeted approach to assess authenticity of fish fillets and patties. *Food. Chem.* 243, 382–388. doi: 10.1016/j.foodchem.2017.09.145

Guillemain, A., Degardin, K., and Roggo, Y. (2017). Performance of NIR handheld spectrometers for the detection of counterfeit tablets. *Talanta* 165, 632–640. doi: 10.1016/j.talanta.2016.12.063

Huang, L., Wu, D., Jin, H., Zhang, J., He, Y., and Lou, C. (2011). Internal quality determination of fruit with bumpy surface using visible and near infrared spectroscopy and chemometrics: a case study with mulberry fruit. *Biosyst. Eng.* 109 (4), 377–384. doi: 10.1016/j.biosystemseng.2011.05.003

Jantra, C., Slaughter, D. C., Liang, P. S., and Pathaveerat, S. (2017). Nondestructive determination of dry matter and soluble solids content in dehydrator onions and garlic using a handheld visible and near infrared instrument. *Postharvest. Biol. Tec.* 133, 98–103. doi: 10.1016/j.postharvbio.2017.07.007

Kamiloglu, S., Serali, O., Unal, N., and Capanoglu, E. (2013). Antioxidant activity and polyphenol composition of black mulberry (Morus nigra L.) products. *J. Berry. Res.* 3 (1), 41–51. doi: 10.3233/JBR-130045

Karunathilaka, S. R., Yakes, B. J., He, K., Bruckner, L., and Mossoba, M. M. (2018). First use of handheld Raman spectroscopic devices and on-board chemometric analysis for the detection of milk powder adulteration. *Food Control* 92, 137–146. doi: 10.1016/j.foodcont.2018.04.046

Khalifa, I., Zhu, W., Li, K.-k., and Li, C.-m. (2018). Polyphenols of mulberry fruits as multifaceted compounds: compositions, metabolism, health benefits, and stability—A structural review. *J. Funct. Foods* 40, 28–43. doi: 10.1016/j.jff.2017.10.041

Li, H., Liang, Y., Xu, Q., and Cao, D. (2009). Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal. Chim. Acta* 648 (1), 77–84. doi: 10.1016/j.aca.2009.06.046

Li, H. D., Liang, Y. Z., Long, X. X., Yun, Y. H., and Xu, Q. S. (2013). The continuity of sample complexity and its relationship to multivariate calibration: a general perspective on first-order calibration of spectral data in analytical chemistry. *Chemom. Intell. Lab. Syst.* 122 (5), 23–30. doi: 10.1016/j.chemolab.2013.01.003

Lou, H., Hu, Y., Zhang, L., Sun, P., and Lu, H. (2012). Nondestructive evaluation of the changes of total flavonoid, total phenols, ABTS and DPPH radical scavenging activities, and sugars during mulberry (Morus alba L.) fruits development by chlorophyll fluorescence and RGB intensity values. *LWT-Food. Sci. Technol.* 47 (1), 19–24. doi: 10.1016/j.lwt.2012.01.008

Marinova, D., Ribarova, F., and Atanassova, M. (2005). Total phenolics and total flavonoids in Bulgarian fruits and vegetables. *J. Univ. Chem. Technol. metallurgy* 40 (3), 255–260.

Næs, T., Isaksson, T., Fearn, T., and Davies, T. (2002). *A user friendly guide to multivariate calibration and classification* (Chichester UK: NIR publications).

O'Brien, N., Hulse, C., Pfeifer, F., and Siesler, H. (2013). Technical Note: near infrared spectroscopic authentication of seafood. *J. Near Infrared Spectrosc.* 21 (4), 299–305. doi: 10.1255/jnirs.1063

Piao, S., Okura, T., and Irie, M. (2018). On-site evaluation of Wagyu beef carcasses based on the monounsaturated, oleic, and saturated fatty acid composition using a handheld fiber-optic near-infrared spectrometer. *Meat Sci.* 137, 258–264. doi: 10.1016/j.meatsci.2017.11.032

Pissard, A., Baeten, V., Romnee, J. M., Dupont, P., Mouteau, A., and Lateur, M. (2012). Classical and NIR measurements of the quality and nutritional parameters of apples: a methodological study of intra-fruit variability. *Biotechnol. Agron. Soc* 16 (3), 294–306.

Sánchez-Salcedo, E. M., Mena, P., García-Viguera, C., Martínez, J. J., and Hernández, F. (2015). Phytochemical evaluation of white (Morus alba L.) and black (Morus nigra L.) mulberry fruits, a starting point for the assessment of their beneficial properties. *J. Funct. Foods* 12, 399–408. doi: 10.1016/j.jff.2014.12.010

Shao, X., Wang, F., Chen, D., and Su, Q. (2004). A method for near-infrared spectral calibration of complex plant samples with wavelet transform and elimination of uninformative variables. *Analytical Bioanalytical Chem.* 378 (5), 1382–1387. doi: 10.1007/s00216-003-2397-9

Silva, D. C., Pastore, T. C. M., Soares, L. F., de Barros, F. A. S., Bergo, M. C. J., Coradin, V. T. H., et al. (2018). Determination of the country of origin of true mahogany (Swietenia macrophylla King) wood in five Latin American countries using handheld NIR devices and multivariate data analysis. *Holzforschung* 72 (7), 521–530. doi: 10.1515/hf-2017-0160

Sorak, D., Herberholz, L., Iwascek, S., Altinpinar, S., Pfeifer, F., and Siesler, H. W. (2012). New developments and applications of handheld Raman, mid-Infrared, and near-infrared spectrometers. *Appl. Spectrosc. Rev.* 47 (2), 83–115. doi: 10.1080/05704928.2011.625748

Soriano-Disla, J. M., Janik, L. J., and McLaughlin, M. J. (2018). Assessment of cyanide contamination in soils with a handheld mid-infrared spectrometer. *Talanta* 178, 400–409. doi: 10.1016/j.talanta.2017.08.106

Tino, P., Jens, K., and Heinrich, G. (2016). Near-Infrared Grating Spectrometer for Mobile Phone Applications. *Appl. Spectrosc.* 70 (5), 734–745. doi: 10.1177/0003702816638277

Vargas Jentzsch, P., Gualpa, F., Ramos, L. A., and Ciobota, V. (2018). Adulteration of clove essential oil: detection using a handheld Raman spectrometer. *Flavour. Frag. J.* 33 (2), 184–190. doi: 10.1002/ffj.3438

Williams, P. C., and Sobering, D. C. (1993). Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *J. Near Infrared Spectrosc* 1 (1), 25–32. doi: 10.1255/jnirs.3

Wolffenbuttel, R. F. (2005). MEMS-based optical mini-and microspectrometers for the visible and infrared spectral range. *J. Micromech. Microeng.* 15 (7), S145–S152. doi: 10.1088/0960-1317/15/7/021

Yan, H., and Siesler, H. W. (2018a). Identification performance of different types of handheld near-infrared (NIR) spectrometers for the recycling of polymer commodities. *Appl. Spectrosc.* 72 (9), 1362–1370. doi: 10.1177/0003702818777260

Yan, H., and Siesler, H. W. (2018b). Quantitative analysis of a pharmaceutical formulation: Performance comparison of different handheld near-infrared spectrometers. *J. Pharm. Biomed. Anal.* 160, 179–186. doi: 10.1016/j.jpba.2018.07.048

Yan, H., Han, B., Wu, Q., Jiang, M., and Gui, Z. (2011). Rapid detection of Rosa laevigata polysaccharide content by near-infrared spectroscopy. *Spectrochim. Acta Part. A* 79 (1), 179–184. doi: 10.1016/j.saa.2011.02.032

Yu, Z., and Dahlgren, R. A. (2000). Evaluation of methods for measuring polyphenols in conifer foliage. *J. Chem. Ecol.* 26 (9), 2119–2140. doi: 10.1023/A:1005568416040

Yun, Y. H., Liang, Y. Z., Xie, G. X., Li, H. D., Cao, D. S., and Xu, Q. S. (2013). A perspective demonstration on the importance of variable selection in inverse calibration for complex analytical systems. *Analyst* 138 (21), 6412–6421. doi: 10.1039/c3an00714f

Zhang, C. H., Yun, Y. H., Fan, W., Liang, Y. Z., Yu, Y., and Tang, W. X. (2015). Rapid analysis of polysaccharides contents in Glycyrrhiza by near infrared spectroscopy and chemometrics. *Int. J. Biol. Macromol.* 79, 983–987. doi: 10.1016/j.ijbiomac.2015.06.025

Combining Chemical Information From Grass Pollen in Multimodal Characterization

Sabrina Diehn[1,2], Boris Zimmermann[3], Valeria Tafintseva[3], Stephan Seifert[1,2], Murat Bağcıoğlu[3], Mikael Ohlson[4], Steffen Weidner[2], Siri Fjellheim[5], Achim Kohler[3,6] and Janina Kneipp[1,2]*

[1] Department of Chemistry, Humboldt-Universität zu Berlin, Berlin, Germany, [2] BAM Federal Institute for Materials Research and Testing, Berlin, Germany, [3] Faculty of Science and Technology, Norwegian University of Life Sciences, Ås, Norway, [4] Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway, [5] Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway, [6] Nofima AS, Ås, Norway

The analysis of pollen chemical composition is important to many fields, including agriculture, plant physiology, ecology, allergology, and climate studies. Here, the potential of a combination of different spectroscopic and spectrometric methods regarding the characterization of small biochemical differences between pollen samples was evaluated using multivariate statistical approaches. Pollen samples, collected from three populations of the grass *Poa alpina,* were analyzed using Fourier-transform infrared (FTIR) spectroscopy, Raman spectroscopy, surface enhanced Raman scattering (SERS), and matrix assisted laser desorption/ionization mass spectrometry (MALDI-TOF MS). The variation in the sample set can be described in a hierarchical framework comprising three populations of the same grass species and four different growth conditions of the parent plants for each of the populations. Therefore, the data set can work here as a model system to evaluate the classification and characterization ability of the different spectroscopic and spectrometric methods. ANOVA Simultaneous Component Analysis (ASCA) was applied to achieve a separation of different sources of variance in the complex sample set. Since the chosen methods and sample preparations probe different parts and/or molecular constituents of the pollen grains, complementary information about the chemical composition of the pollen can be obtained. By using consensus principal component analysis (CPCA), data from the different methods are linked together. This enables an investigation of the underlying global information, since complementary chemical data are combined. The molecular information from four spectroscopies was combined with phenotypical information gathered from the parent plants, thereby helping to potentially link pollen chemistry to other biotic and abiotic parameters.

Keywords: pollen, consensus principal component analysis, ANOVA simultaneous component analysis, Fourier-transform infrared spectroscopy, matrix assisted laser desorption/ionization mass spectrometry, surface-enhanced Raman scattering, Raman spectroscopy, *Poa alpina*

INTRODUCTION

The analysis of pollen samples is a crucial task that is necessary in several fields, including agriculture, plant physiology, ecology, allergology, and climate studies. Therefore, significant efforts have been undertaken to utilize analytical techniques that give insight into pollen chemical composition, to achieve a characterization that is more detailed than the morphological typing by light microscopy.

Vibrational spectroscopic methods, such as FTIR (Pappas et al., 2003; Gottardini et al., 2007; Dell'Anna et al., 2009; Julier et al., 2016; Depciuch et al., 2018; Jardine et al., 2019), Raman scattering (Ivleva et al., 2005; Schulte et al., 2008), and surface enhanced Raman scattering (SERS) (Sengupta et al., 2005; Seifert et al., 2016), as well as mass spectrometric methods (Krause et al., 2012; Lauer et al., 2018) can be applied to classify pollen according to taxonomic relationships based on molecular composition. Pollen spectra can also indicate changes in chemical composition according to genetic background and environmental influences (Zimmermann and Kohler, 2014; Zimmermann et al., 2017; Diehn et al., 2018). A vibrational or mass spectrum carries fingerprint-like information from all biomolecular species in the pollen samples that are probed with the respective spectroscopy, albeit with different selectivity and sensitivity (Bagcioglu et al., 2015; Diehn et al., 2018). For example, FTIR spectra of pollen reveal different biochemical composition for different plant species (Pappas et al., 2003; Gottardini et al., 2007; Dell'Anna et al., 2009; Zimmermann, 2010; Julier et al., 2016) and within a specific species (Zimmermann et al., 2017), mainly based on vibrations of protein and lipid molecules contained in the pollen grains. Raman microspectroscopy, since based on different selection rules, can give molecular and structural information complementary to infrared spectroscopy. Moreover, due to the different geometry in Raman micospectroscopic experiments and the penetration depth of the light used to excite the Raman scattering, different parts of the pollen grains are probed. For example, Raman spectra show many contributions by stored starch and lipid bodies and by the sporopollenin polymer that comprises the pollen exine (Schulte et al., 2008). This biopolymer consists of coniferyl aldehyde and ferulic acid blocks (Rozema et al., 2001; Blokker et al., 2006; Li et al., 2019) and provides high stability and protection to the gametes. Surface-enhanced Raman scattering (SERS), in turn, gives very strong signals from pollen constituents that must interact with metal nanoparticles, i.e., the SERS substrate, and although it enables the investigation of less abundant molecular species, it has a high selectivity for specific classes of molecules. This can be the water-soluble pollen fraction, extracted in a facile way (Seifert et al., 2016) or the sporopollenin polymer after embedding the SERS nanoparticle substrate inside the nanoscopic cavities of the pollen shell (Joseph et al., 2011).

In contrast to vibrational spectra, the molecular basis of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) mass spectra from the complex pollen samples is still much less understood (Krause et al., 2012), but it was successfully shown to serve as fingerprint-like data for the classification and identification of pollen species as well (Krause et al., 2012; Lauer et al., 2018), even at the sub-species level (Diehn et al., 2018). MALDI-TOF-MS delivers chemical identifiers that are complementary to those of FTIR spectroscopy when applied to the same set of pollen samples (Zimmermann et al., 2017; Diehn et al., 2018).

Biospectroscopic data are usually evaluated using multivariate methods, including principal component analysis (PCA) (Lasch and Naumann, 1998; Ellis and Goodacre, 2006; Qian et al., 2008). It factorizes the data matrix that contains all spectra to one score value for each spectrum and one loading vector for all spectra. The weighting of the data based on variance in a PCA enables easier identification of differences in a spectral data set and helps identification of latent structures (Pearson, 1901; Hotelling, 1933; Bro and Smilde, 2014). The outcome of a PCA can be explored easily by scores plots and interpretation of spectral features in corresponding loadings.

Since each of the four analytical approaches provides unique information about one particular fraction of the complex pollen chemistry, a combination of the data in one extensive analysis would be very promising to improve pollen characterization and classification. In particular, the combination of different chemical data is expected to reveal chemical aspects of plant/pollen phenotype in a more sensitive and more comprehensive fashion, enabling more insight into, e.g., the adaptation of plants to environmental conditions. Recent studies show the great potential of applying consensus principal component analysis (CPCA) (Wold et al., 1987; Westerhuis et al., 1998) as a multi-block method to the data from different analytical techniques. The combination of very different data blocks can be used in the investigation of biological samples (Perisic et al., 2013), including pollen (Bagcioglu et al., 2015). CPCA is an extension of the PCA concept and aims for the maximization of common variation patterns in the different data blocks. In CPCA, the data blocks are deflated with respect to the variation that is expressed in the so-called global scores. A difference between PCA on every single block and a CPCA analysis is that the same variation appears in the same components in every data block. Thereby, we can compare results directly between the different types of spectroscopic information. A potential co-variation in the blocks is specified in the explained variance of the respective block scores. Furthermore, a correlation loadings plot can be generated as result of a CPCA. CPCA not only joins the information from different data blocks in one analysis, it also enables the evaluation of interactions between the different blocks (Hassani et al., 2010; Hassani et al., 2013).

Here, we apply CPCA to the pollen data of the four complementary methods FTIR spectroscopy, Raman microspectroscopy, SERS, and MALDI-TOF MS and compare the results to those of PCA of each of the single data blocks. The data are measured from pollen samples obtained in a large-scale greenhouse experiment that was aiming for a diverse range of investigations connecting to pollen research

(Zimmermann et al., 2017; Diehn et al., 2018), making also other phenotypic data on the parent plants available to be included in the analysis. The sample set discussed comprises pollen from one grass species, *Poa alpina*. The parent plants originate from three different populations, within which four different growth conditions were applied to individuals of identical genetic constitution (**Figure 1**). The design of this experiment generates two initially separate questions. The first is regarding the different chemical composition of pollen from different populations in the same species. The second question relates to the differences in pollen composition as a result of different growth conditions of genetically identical plants within one population. Therefore, the results of both CPCA and the PCA are compared for the different spectroscopic methods and separately for the different design factors, that is, population and growth condition. One of the aims is to assess the sensitivity of the multimodal characterization towards an influence of population and environmental conditions, respectively, on pollen chemistry, regardless of the hierarchical structure of the variation introduced in the specific sample set. To address this, we have used an ANOVA Simultaneous Component Analysis (ASCA) in order to investigate the possibility of separating between different sources of variation in the complex sample set.
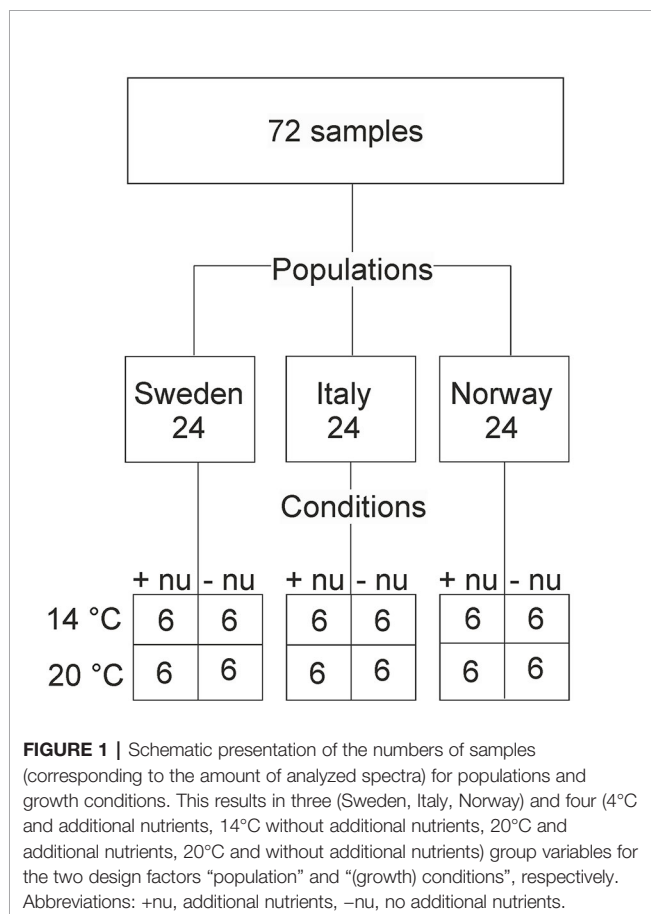
EXPERIMENTAL

Pollen Samples

In a greenhouse experiment, plants of the grass species *Poa alpina* were grown under different environmental conditions using seeds acquired from the Nordic Gene Bank. The seeds belonged to three different populations of origin, Sweden, Italy, and Norway, that were chosen to cover geographic and climatic variation (**Figure 1**). Details of the growth experiment can be found in Zimmermann et al. (2017). Briefly, for each population, six individuals were grown from seeds in the spring, and, after the summer, each individual was divided into four clones. The plants were subsequently vernalized for 12 weeks at 4°C with a day length of 8 h. After vernalization, the plants were grown under long day conditions (20 h), and the respective clones of the individuals were subjected to four different environmental conditions: at 14°C and additional nutrients in the irrigation water (+nu), at 14°C without additional nutrients in the irrigation water (-nu), at 20°C +nu, and at 20°C –nu, respectively. Pollen samples were collected from the pollinating plants. Thereby, the sample set contained 24 different pollen samples for each of the three populations, and the whole sample set consisted of 72 pollen samples (**Figure 1**). The pollen grains were stored at -20°C after collection until further preparation.

FTIR Spectroscopy

Bulk samples of pollen were prepared as homogenous suspensions and measured by using high-throughput FTIR accessory. Approximately 1 mg of a pollen sample was transferred into 1.5 ml microcentrifuge tube containing 500 µl of distilled water. The sample was sonicated in ice bath, by a 2 mm probe coupled to a Q55 Sonicator ultrasonic processor (QSonica, LLC, USA) under 100% power. The sonication period was 2 min in total, with 30 s intermission after the first minute of sonication to minimize the increase in temperature. Following the sonication, the sample suspension was centrifuged with 13,000 rpm for 10 min, and the suspension was concentrated by removing 400 µl of supernatant. Of the remaining suspension, three aliquots (technical replicates), each containing 8 µl, were transferred onto an IR-transparent silicon 384-well microtiter plate (Bruker Optik GmbH, Germany). The microtiter plate was dried at room temperature for 1 h to create adequate films for FTIR measurements.

FTIR measurements were obtained using a HTS-XT extension unit coupled to a TENSOR 27 spectrometer (both Bruker Optik GmbH, Germany). The system is equipped with a globar mid-IR source and a DTGS detector. The spectra were recorded in transmission mode, with a spectral resolution of 4 cm^{-1} and digital spacing of 0.964 cm^{-1}. Background (reference) spectra of an empty well on a microtiter plate were recorded before each sample well measurement. The spectra were measured in the 4,000–500 cm^{-1} spectral range, with 32 scans for both background and sample spectra, and using an aperture of 5.0 mm. Data acquisition and instrument control were carried out using the OPUS/LAB software (Bruker Optik GmbH, Germany). Spectra were pre-processed, first by taking the second derivative employing the Savitzky–Golay algorithm

FIGURE 1 | Schematic presentation of the numbers of samples (corresponding to the amount of analyzed spectra) for populations and growth conditions. This results in three (Sweden, Italy, Norway) and four (4°C and additional nutrients, 14°C without additional nutrients, 20°C and additional nutrients, 20°C and without additional nutrients) group variables for the two design factors "population" and "(growth) conditions", respectively. Abbreviations: +nu, additional nutrients, –nu, no additional nutrients.

(Savitzky and Golay, 1964) with a polynomial of degree two and a window size of 7 points, and second by using extended multiplicative signal correction (EMSC) with linear and quadratic components (Martens and Stark, 1991) (Savitzky and Golay, 1964; Zimmermann and Kohler, 2013). The spectral range from 800–1,800 cm^{-1} was used for multivariate analysis. An average spectrum was calculated from the spectra of the three technical replicates (aliquots) per sample, resulting in a set of 72 average spectra that were used for further analysis.

Raman Microspectroscopy

Single pollen grains of each pollen sample were measured using a Raman microspectrometer (Horiba, Bensheim, Germany) with a 50x microscope objective (Olympus, Hamburg, Germany) and a diode laser operating at a wavelength of 785 nm and an intensity of $7 \cdot 10^5$ W/cm^2. For each sample, ten spectra from ten different single pollen grains were collected, using an accumulation time of 10 s per spectrum. In total, 720 individual spectra were obtained. Spectral resolution was 1.3-1.6 cm^{-1}, considering the full spectral range. For frequency calibration, six bands in the spectrum of 4-acetaminophenol (1648.4, 1323.9, 1168.5, 857.9, 651.6, 390.9 cm^{-1}) were used. After spike removal, the raw spectra were interpolated in the range from 400 to 1,750 cm^{-1} to achieve an equal distribution of data points across the whole spectral range. A distance of 1.45 cm^{-1}, corresponding to the average spectral resolution in the experiment was chosen as distance between variables. Subsequently, a baseline for each spectrum was estimated by asymmetric least square smoothing (Eilers, 2003) and subtracted from the respective spectrum, followed by vector normalization of the baseline corrected spectrum. An average spectrum was calculated for each sample from the 10 respective spectra, resulting in a set of 72 average spectra that were used for further analysis.

Surface-Enhanced Raman Scattering (SERS)

In the SERS experiments, the water-soluble components of the pollen grains were extracted and mixed with an aqueous solution of citrate-stabilized gold nanoparticles as described previously in reference (Seifert et al., 2016). For this purpose, 100 μl Millipore water were added to 0.2 mg of the pollen sample. After 5 min, the samples were centrifuged and the supernatant was pipetted off. 2 μl of this aqueous pollen extract were mixed with 20 μl citrate-stabilized gold nanoparticles obtained based on the protocol described in ref. (Lee and Meisel, 1982) and 2 μl of a 0.1 M sodium chloride solution were added. Subsequently, 20 μl of this mixture were transferred to a calcium fluoride slide for the SERS measurement. The SERS experiments were performed on a Raman microscope (Horiba, Bensheim, Germany) in the focal volume of a 60x water immersion objective (Olympus, Hamburg) with a laser operating at a wavelength of 785 nm and an intensity of $2.9 \cdot 10^5$ W/cm^2. Two extracts for each sample (technical replicates) were prepared and analyzed. For each extract, 1,000 spectra with an accumulation time of 1 s per

spectrum were collected. This procedure yielded SERS data sets containing 144,000 individual spectra in total (2,000 spectra per pollen sample). The spectra were frequency calibrated using a spectrum of 4-acetamidophenol. Further pre-processing included spike removal, interpolation, baseline correction, and vector normalization as described in the previous section. The 2,000 spectra for each sample (obtained from different extracts) were averaged so that in total 72 average SERS spectra were analyzed.

MALDI-TOF MS

For the MALDI-TOF MS experiments, each pollen sample was deposited on a MALDI stainless steel target. 1 μl of formic acid (90%) was added, and after drying at room temperature, 1 μl of matrix solution (10 mg of α-cyano-4-hydroxycinnamic acid in 1 ml 1:1 acetonitrile/water and 0.1% trifluoroacetic acid) was applied. (Seifert et al., 2015) MALDI-spectra were obtained in the mass range from m/z 1,000 to 15,000 using an Autoflex III MALDI-TOF mass spectrometer (Bruker Daltonik, Bremen Germany) equipped with a 355 nm Smartbeam laser (200 Hz) and operating in positive linear mode at an acceleration voltage of 19.13 kV. Two technical replicates for each sample were prepared, resulting in 144 spectra in total. To obtain equal distances between the variables, the spectra were interpolated with a distance of m/z 2 between data points in the mass range from m/z 5,000 to 9,000, and a 6-degree polynomial baseline correction was applied before the spectra were vector-normalized. The two technical replicates were averaged to yield one spectrum for each sample in order to obtain the same amount of spectra as in each of the other data blocks.

Morphological and Dry Weight Measurements of Parent Plants

During the pollination stage, the height of the flowering shoots of the parent plants was determined, using the average value for three highest flowering shoots per individual plant. Furthermore, the number of flowering shoots for each individual was determined. Plant dry mass was determined at the end of seed production life stage by cutting the parent plants at ground level and drying them at 60°C for 24 h.

Chlorophyll Content of Parent Plants

Chlorophyll a and b concentrations were measured by ultraviolet-visible absorption measurements (Solhaug, 1991) Leaf samples from each individual were collected during the pollination life stage. Approximately 4–8 mg of a sample were transferred to microcentrifuge tubes containing 1.5 ml N, N-dimethylformamide, and kept at +4°C for 24 h. The extracts were measured on a UV-Vis spectrophotometer (Shimadzu 1800, Japan) using cuvettes with 1 cm path length. Chlorophyll a and b concentrations were calculated by using absorbance values at 647 nm and 664 nm according to the equations by Porra et al. (Porra et al., 1989). The chlorophyll content, morphological and dry weight data were combined in a separate, fifth data block, termed here additional plant data. The data were normalized based on data dispersion (autoscaling) before further analysis.

Data Analysis

All pre-processing steps of the spectra and all analyses were performed using the statistics and machine learning toolbox in Matlab R2016a (The Mathworks, Inc., Natick, MA, USA). First, the individual data sets were analyzed by principal component analysis (PCA) as described in previous work (Seifert et al., 2015; Seifert et al., 2016; Joester et al., 2017; Zimmermann et al., 2017; Diehn et al., 2018). Consensus principal component analysis (CPCA) was used to combine the five different kinds of data (FTIR, Raman, SERS, and MALDI-TOF MS spectra, as well as a block with the additional data obtained from the parent plants), each one pre-processed as described before. CPCA enables the analysis of the sample variances within the data blocks and between different data blocks (Hassani et al., 2010). In order to apply CPCA, all the data sets need to have the same sample dimension, and the order of the samples should be identical for all data sets included in the analysis. Therefore, all technical replicates were averaged, such that 72 spectra for each data block were obtained. Outliers were not removed from the averaged data sets, since each method probes different parts of the samples, and the data acquisition differs greatly. For the common representation of the loadings of all the different kinds of data in one correlation loadings plot as result, thresholds were defined for the respective data types and those positive/negative peaks above/below the respective threshold are represented in the unified plot. For clarity, all other spectral variables are not shown in these plots. Only the variables belonging to the additional plant data are displayed as a whole.

To evaluate the ability to discriminate each of the three groups in the design factor "population" as well as the four individual groups of the design factor "growth condition" in the PCA, Kruskal-Wallis H-test and MANOVA were used. The populations Sweden, Italy and Norway, as well as the conditions 14°C and additional nutrients (14+nu), 14°C without additional nutrients (14-nu), 20°C and additional nutrients (20+nu) and 20°C and without additional nutrients (20-nu) are defined as group variables of the design factors "population" and "growth condition", respectively (**Figure 1**).

Kruskal-Wallis H-test and MANOVA were also applied to the groups of score values after CPCA on the global and block scores, respectively. As result, the two tests give one p-value and one d-value for each PCA or CPCA. The Kruskal-Wallis H-test as a non-parametric statistical test was chosen after assessment of the data sets regarding their normal distribution. As known from previous work, the data obtained from SERS experiments do often not show normal distribution (Seifert et al., 2016). Here, the test is used to prove the null hypothesis that the distribution of the data within each respective group, that is, three groups for the three different populations in the whole data set, and four groups for the four different growth conditions within each of the populations, is equal. A p-value below 0.05 indicates a significant difference in this distribution for at least one of the groups. The Kruskal-Wallis H-test was applied to the score values of each PCA and CPCA that was conducted using the *kruskalwallis* function in Matlab. Each of the first ten components was investigated, and the p-value was reported using always the

first PC. However, in case of a high p-value when using the first PC, the lowest p-value with any of the other first ten PCs is discussed (see **Tables 1** and **2**). The distribution of the score values of the first PC is also visualized using box plots.

MANOVA, comparing the multivariate means for a specific dimensionality, was executed using the *manova1* function in Matlab. The first ten PCs (covering at least ~90% of the variance, with over ~96% of the variance in the FTIR and Raman data sets) were used for MANOVA, since a balance had to be found between the requirement to have as much variance as possible covered, an equal treatment of all data sets, and a reasonable time for computation. MANOVA was used to estimate the non-random variation of the group mean of each population and each growth condition, respectively. In this case, the dimensionality was either three, corresponding to the three different populations, or four, due to formation of four groups corresponding to the four different growth conditions. If the group means were equal, that is, when no discrimination was found, the d-value would be 0. A d-value of 1 would indicate that the group means show a linear dependence on each other, so that two groups are separated. For the data sets here, the d-value could reach two in the case of the three different populations and three in the case of the four different growth conditions.

To calculate the variation in the data induced by the different design factors, such as population, nutrients, temperature, as well as their interaction, we used an approach underlying ANOVA-

TABLE 1 | Results of the PCA (p- and d-values) for the discrimination of pollen samples from all populations and from the individual populations grown under different environmental conditions.

Method	Population	p-values for the separation of the pollen samples based on environmental conditions	d-values for grouping based on environmental conditions (max. 3)
FTIR	Sweden	<0.01	3
	Italy	0.035	3
	Norway	0.072 (0.011, PC4)	2
	All	0.64 (<0.01, PC6)	3
Raman	Sweden	0.51 (<0.01, PC4)	1
	Italy	<0.01	3
	Norway	0.36 (<0.01, PC3)	2
	All	<0.01	2
SERS	Sweden	0.73 (<0.01, PC2)	2
	Italy	0.37 (<0.01, PC4)	2
	Norway	0.64 *	0
	All	0.78 (0.014, PC7)	0
MALDI	Sweden	0.62 (0.046, PC3)	2
	Italy	0.012	1
	Norway	0.018	1
	All	0.98 (<0.01, PC5)	1
Additional plant data	Sweden	<0.01	1
	Italy	<0.01	2
	Norway	<0.01	2
	All	<0.01	2

*no p-value below 0.05 can be found for the first 10 PCs. The p-values are obtained for the score values of PC1. In case of p-values above 0.05 in PC1, the lowest p-value with any of the other first 10 PCs and respective PC are shown in parentheses. For the calculation of d-values, the score values of the first 10 PCs were used.

TABLE 2 | Results of the CPCA (p- and d-values) for the discrimination of pollen samples from all populations and from the individual populations grown under different environmental conditions.

Method	Population	p-values for the separation of the pollen samples based on environmental conditions	d-values for grouping based on environmental conditions (max. 3)
Global	Sweden	0.036	2
	Italy	<0.01	3
	Norway	<0.01	2
	All	0.95 (<0.01, CPC3)	3
FTIR	Sweden	0.032	2
	Italy	0.013	3
	Norway	0.084 (0.021, CPC4)	3
	All	0.70 (0.011, CPC3)	3
Raman	Sweden	0.47 (0.013, CPC6)	2
	Italy	<0.01	3
	Norway	0.19 (<0.01, CPC5)	3
	All	0.64 (<0.01, CPC3)	3
SERS	Sweden	0.093 (0.032, CPC5)	2
	Italy	0.089 (<0.01, CPC3)	3
	Norway	0.40 *	2
	All	0.98 (<0.01, CPC3)	3
MALDI	Sweden	0.60 (0.027, CPC5)	2
	Italy	<0.01	3
	Norway	<0.01	3
	All	0.98 (<0.01, CPC3)	2
Additional plant data	Sweden	<0.01	2
	Italy	<0.01	3
	Norway	<0.01	2
	All	<0.01	3

*no p-value below 0.05 can be found for the first 10 PCs. The p-values are obtained for the score values of PC 1. In case of p-values above 0.05 in PC 1, the lowest p-value with any of the other first ten PCs and respective PC are shown in parentheses. For the calculation of d-values, the score values of the first ten PCs were used.

PCA and ASCA (Harrington et al., 2005; Smilde et al., 2005), which are widely used for this purpose (Jansen et al., 2005; de Haan et al., 2007). In both of these methods an ANOVA model is established. It represents the original data as a sum of matrices, each of which corresponds to one design factor. Each of these matrices consists of the means of the spectra that correspond to different levels of the factor. As an example, if one design factor has two levels, its respective matrix will have repeated means of the two levels in the corresponding rows.

In the ANOVA model used in this study, the design factor "temperature" has two levels (14°C and 20°C), the design factor "nutrients" consists of two levels (+nu and -nu), the design factor "interaction" is the interaction of 'temperature' and 'nutrients' and has four levels, the design factor "population" has three levels (Italy, Norway, and Sweden), the factor 'individual' has 72 levels that correspond to each individual plant in the set of samples. The residual variance is summarized in the factor "residuals".

ASCA uses the ANOVA model to study the effects of the design factors on the variation in the data and runs PCA on each of the matrices to interpret this variation. In ANOVA-PCA the ANOVA model is analyzed further by PCA to find if the differences in the levels for each factor are significant. In this study, we use the ANOVA model purely to calculate each design factor contribution into variation in the data. Since it was of

interest to learn about the variation contributions in each block of data representing each measurement (FTIR, MALDI, Raman, SERS and other parental plant data, respectively), the same analysis was done separately on each data block.

RESULTS AND DISCUSSION

Analysis of the Separate Data Blocks

The well-defined sample set (**Figure 1**) was measured by the four different methods FTIR, Raman, and SERS spectroscopy and MALDI-TOF mass spectrometry, after different pre-processing of the samples according to the requirement of each spectroscopy (see *Experimental* section), leading to the probing of complementary constituents of the pollen. The four different types of spectra were obtained from the 72 pollen samples, constituting four separate data blocks. Furthermore, phenotypic information from the respective parent plants was combined in a fifth data block.

Figure 2 shows the four types of spectra obtained for the three different populations, with averaging information in each population over pollen samples obtained from all parent plants grown at the four different environmental conditions. The signals in the FTIR spectra (**Figure 2A**) can mainly be assigned to proteins, represented, e.g., by the amide I and amide II bands at 1669 and 1540 cm^{-1}, respectively, to lipids, exemplified by vibrations at 1156, 1467, and 1744 cm^{-1} and to sporopollenin, e.g., at 835, 1512, and 1624 cm^{-1}, in agreement with spectra reported previously (Bagcioglu et al., 2015; Zimmermann et al., 2017).

The average Raman spectra in **Figure 2B** are very similar to each other as well and display similar signals, albeit at slightly varying positions, suggesting small differences in the chemical composition of pollen from different populations. The bands at 1008, 1161, and 1528 cm^{-1} can be assigned to carotenoids (Schulte et al., 2009), while the signals at 526, 549, 725, 855, 1271, 1457, and 1662 cm^{-1} are due to vibrations of proteins (Schulte et al., 2008; Joester et al., 2017). The bands at 483, 1082, and 1322 cm^{-1} are assigned to carbohydrates (Schulte et al., 2008; Pigorsch, 2009) that can occur at high local concentrations in the pollen grains as starch deposits. Due to superposition of several molecular vibrations, some bands in the Raman spectra of pollen can be assigned to other origins as well. As examples, the bands at 1161, 1271, 1313, and 1608 cm^{-1} could also be assigned to the ferulic acid and coumaric acid building blocks in sporopollenin (Blokker et al., 2006; Bagcioglu et al., 2015). Furthermore, the band at 1608 cm^{-1} has also been associated with mitochondrial activity (Huang et al., 2004; Pully and Otto, 2009).

Due to the sample preparation as aqueous extract and the use of aqueous nanoparticle solutions, the SERS experiments probe the water-soluble fraction of the pollen grains. Because of the high variation in the SERS spectra caused by the specifics of the SERS experiment, high numbers of spectra are needed for a reliable statistical analysis. (Seifert et al., 2016) Therefore, 2,000 spectra were measured from each sample, resulting in reproducible average spectra that are based on 24,000 individual spectra per population. They are shown in **Figure 2C**. The average spectra
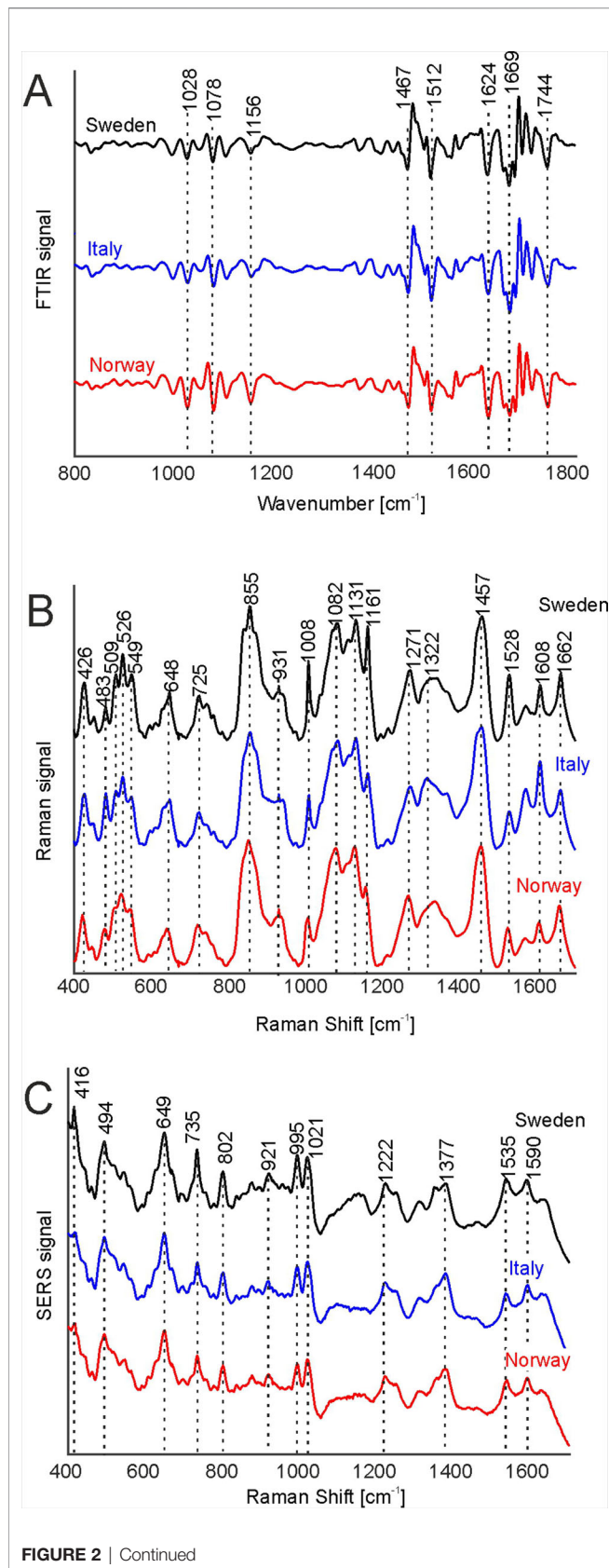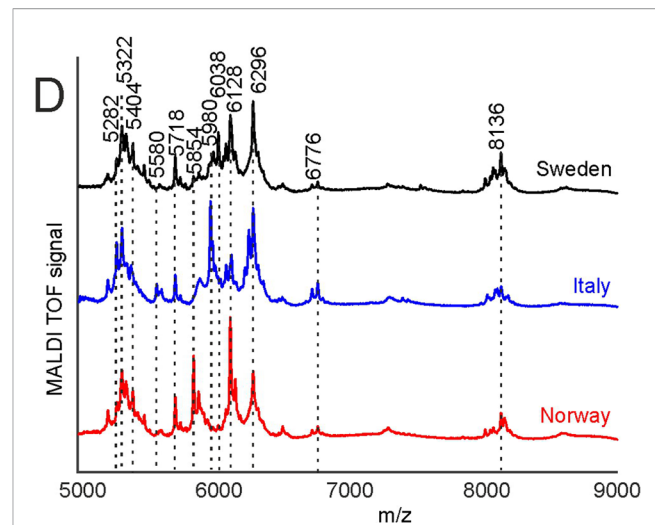
FIGURE 2 | FTIR **(A)**, Raman **(B)**, SERS **(C)**, and MALDI-TOF MS **(D)** spectra of pollen from the populations Sweden (black), Italy (blue), and Norway (red). All spectra were pre-processed according the requirements of the respective spectroscopic method and are averages from the respective population, including samples obtained for all growth conditions. The spectra are stacked for clarity. FTIR, Fourier-transform infrared spectroscopy; Raman, Raman spectroscopy; SERS, surface enhanced Raman scattering; MALDI-TOF MS, matrix assisted laser desorption/ionization mass spectrometry.

are very similar, and their characteristic bands can mainly be assigned to vibrational modes of nucleobases, e.g., at 494, 649, 735, 802, 921 cm^{-1} (Seifert et al., 2016) and amino acids, at 995, 1021, 1221 cm^{-1} (Kyu et al., 1987; Stewart and Fredericks, 1999), in agreement with the probing of water-soluble biomolecules extracted from the pollen.

MALDI TOF mass spectrometry was utilized to detect large molecules with a mass over 5 k Dalton. **Figure 2D** displays average MALDI-TOF mass spectra showing peaks at m/z 5282, 5322, 5404, 5580, 5718, 5853, 5980, 6128, 6296, 6776, and 8136. The differences in the population averages are obvious and indicate that the pollen samples differ in their composition in each population. From earlier attempts to interpret the bands we infer that they include oligosaccharides (Krause et al., 2012; Seifert et al., 2015; Diehn et al., 2018) and larger peptides.

By PCA of the respective type of spectra/data, the pollen samples of the three different populations can be discriminated using each of the individual data blocks. To visualize the distribution of the score values of PC1 for each population, **Figure 3** shows the corresponding box plots with minimum and maximum score values. Outliers are mainly observed for the SERS data (**Figure 3C**) due to high variation in this type of data owing to the specific measurement approach (Seifert et al., 2016). P-values below 0.05 for all five data sets indicate a separation of at least one group for all these data sets.

The box plots of **Figure 3** show that an unequivocal separation of all three populations based on PC1 is only possible when the MALDI-TOF MS scores (**Figure 3D**) are used. The scores of FTIR (**Figure 3A**) and Raman data (**Figure 3B**) for example
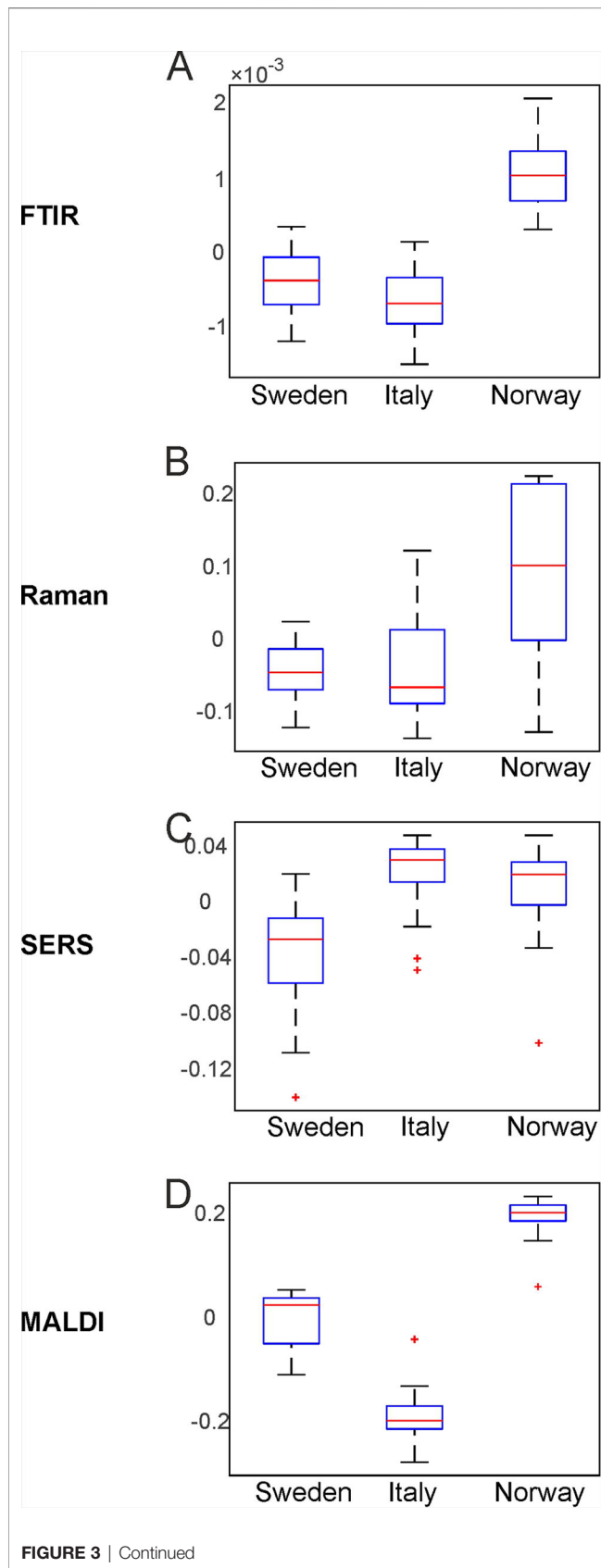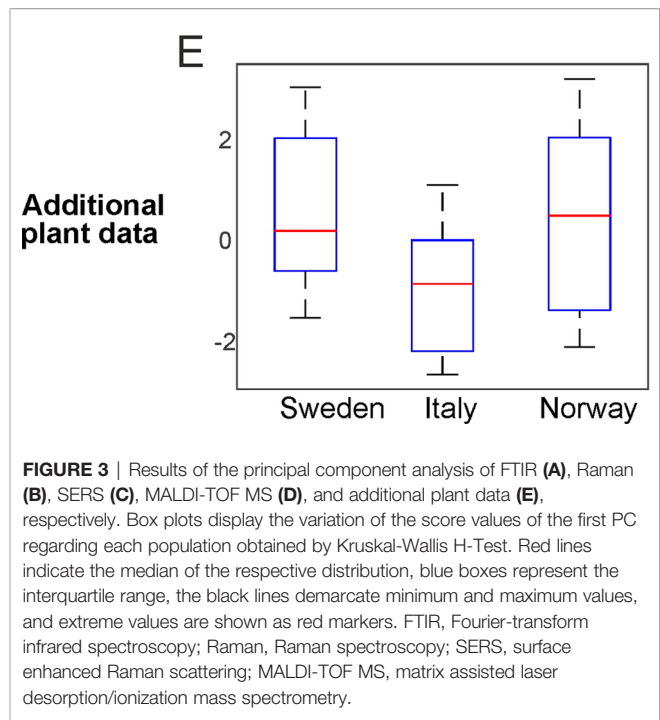
FIGURE 3 | Results of the principal component analysis of FTIR **(A)**, Raman **(B)**, SERS **(C)**, MALDI-TOF MS **(D)**, and additional plant data **(E)**, respectively. Box plots display the variation of the score values of the first PC regarding each population obtained by Kruskal-Wallis H-Test. Red lines indicate the median of the respective distribution, blue boxes represent the interquartile range, the black lines demarcate minimum and maximum values, and extreme values are shown as red markers. FTIR, Fourier-transform infrared spectroscopy; Raman, Raman spectroscopy; SERS, surface enhanced Raman scattering; MALDI-TOF MS, matrix assisted laser desorption/ionization mass spectrometry.

FIGURE 3 | Continued

show very similar distributions for the two populations Sweden and Italy. In order to include more than one principal component when evaluating separation of the three populations by PCA, d-values were determined by MANOVA of the scores of the first ten principal components of each PCA/data block. For all data blocks a d-value of 2 is obtained. This indicates the separation between three groups, corresponding here to the three populations. Therefore, we conclude that a separation of the three different populations is possible with any of the data sets.

The parent plants in each population were grown under four different conditions. Discrimination regarding potential effects of additional nutrients and temperature as design factors on pollen chemistry was studied using each of the five data sets separately as well. This was done for each population individually, as well as for all populations together. **Table 1** summarizes for each data block the PCA results. The p-values were determined using one PC (**Table 1**, left column). In case of a high p-value when using the first PC, the lowest p-value with any of the other first ten PCs is shown in the table. The d-values were determined using the first ten principal components (**Table 1**, right column).

The first section of **Table 1** displays the outcome of the PCAs obtained from the FTIR data sets. The separation based on FTIR spectra receives a p-value below 0.05 and a d-value of 3 for the populations Sweden and Italy, indicating that FTIR data alone enable differentiation of the applied growth conditions for these two populations (compare also the box plots in **Figures S2** and **S3**). The FTIR data set of the population Norway with p-value larger than 0.05 and a d-value of 2 comprises less variance between growth conditions. When all populations are analyzed together, a high p-value for the first PC is obtained, which means that none of the four different growth conditions is separated

using the variance explained by the first PC. Nevertheless, using the 1st to 10th PC, the d-value of 3 indicates a possible separation of all four growth conditions by FTIR alone.

Using the Raman data sets, the p and d-values of the PCAs from the data of the populations Sweden and Norway indicate a less sufficient discrimination ability (**Table 1**, second section). Only for the population Italy, a low p-value and a d-value of 3 can be interpreted as a separation of the four groups of the different growth conditions. In addition, the analysis of all populations together leads to a small p-value, showing the separation of at least one group based on Raman spectral information. We attribute the smaller discrimination ability compared to FTIR to this different selectivity of Raman spectroscopy. The high variances according to the growth conditions in the population Italy explained by the first PC are remarkable, and in good agreement with previous studies on phenotypic plasticity in pollen (Zimmermann et al., 2017). The higher the phenotypic plasticity, the more the chemical composition in pollen varies when environmental conditions change. The high phenotypic plasticity of the population Italy has been inferred from MALDI TOF MS and FTIR spectra of the same *Poa alpina* population previously (Zimmermann et al., 2017; Diehn et al., 2018), where a lower inner-group variance regarding different genotypes of the plants was found.

Investigation of the SERS spectra from aqueous pollen extract by PCA results in p-values above 0.05 for each individual population as well as the whole data set (**Table 1**, third section), clearly showing that an analysis of the samples by SERS alone will not be sufficient for the discrimination of pollen from parent plants that were grown under different environmental conditions. Nevertheless, according to the p-values found in PC2 in population Sweden and PC4 in population Italy (p-values in parentheses in **Table 1**), the variances from the effect of the growth conditions can also be detected in the aqueous extract for these two data sets and therefore add complementary information in the multi-block analysis discussed below.

The MALDI TOF MS data from population Sweden lead to a p-value above 0.05, whereas the p-values for the other two populations stay slightly below 0.05 (**Table 1**, fourth section). In contrast, based on the outcome for the whole population, we cannot conclude chemical variance as result of different growth conditions using only one PC. The d-value of 1 (obtained using the first ten PCs), found for the whole data set as well as for population Italy and population Norway, can be interpreted as the formation of two distinct groups of MALDI spectra. This is in agreement with our previous results (Diehn et al., 2018), where we found a high ability to discriminate between pollen from plants growing with additional nutrients and pollen from plants without additional nutrients using MALDI data from the same samples of *Poa*. Since the discrimination takes place in the range m/z 5000-9000, we infer that the detected signals belong to proteins and their derivatives from pollen nutrient storage.

The last section in **Table 1** contains the p- and d-values for the analysis of additional plant phenotype data, namely height and number of flowering shoots, plant dry mass, and chlorophyll content. The p-values for each population and for the data set

with all three populations combined are below 0.05, and we conclude that the variances regarding the separation of at least one specific group of scores from the other growth conditions are high. The d-values for the analyses of the data sets, however, are 2 or smaller, indicating that discrimination regarding all four growth conditions is not obtained. This is also be illustrated in the box plots in **Figures S1–S4** (last rows).

The variation contributions of the different design factors, such as population, nutrients, temperature and their interaction as well as the contributions from individual variation, were calculated with an approach underlying ANOVA-PCA and ASCA (Harrington et al., 2005; Jansen et al., 2005; Smilde et al., 2005; de Haan et al., 2007). **Figure 4A** shows the contribution of all possible design factors, that is, each type of variation for the whole data set of 72 spectra for each method. The variation contribution of the populations (**Figure 4A**, cyan bars) is very large in the four spectroscopic/spectrometric data sets, larger than the variation contribution due to the different growth conditions (**Figure 4A**, blue, orange and yellow bars). Interestingly, and in agreement with previous work (Zimmermann et al., 2017), the contributions of variation of the individual samples (**Figure 4A** first column, purple bars) is of similar magnitude as that introduced by changes in growth condition of the parent plant, and in the data sets from SERS and MALDI (**Figure 4A** second and third column, purple bars), this contribution by individual variation is even larger.

Considering the data gathered from the parent plants, the largest variation contribution is the effect of the nutrient addition (**Figure 4A**, rightmost bar, orange coloring), obviously having more consequences for the constitution of the plant itself than for the chemical make-up of the pollen. In addition, differences between phenotypic features of the plants in the different populations are of a similar magnitude as variation due to individual differences. (**Figure 4A**, rightmost bar, cyan coloring). The contribution of the residual variation (**Figure 4A**, green bars) is relatively high for all data sets. In some, such as Raman and SERS (**Figure 4A**, second and third column, respectively), the residual variation contributes the most. We think that this must be due to the type of experiment, which are in these cases much more prone to spectrum-to-spectrum fluctuation. Moreover, the Raman and SERS data sets were collected over a course of several weeks, whereas MALDI and FTIR were high-throughput measurements obtained in one-preparation procedures. So, the big residual variation in SERS and Raman can be explained by the experimental variations.

In **Figure 4B**, relative variation contributions of the growth conditions, namely temperature, nutrients and the interaction of both factors are presented. Variations by these factors are emphasized by omitting population variation, individual variation, and residual variation. To calculate these, a variation of each factor was normalized by the sum of the variations for the three factors of interest. While the variation contribution of both, the temperature and nutrients is high for the three spectroscopic methods FTIR, Raman and SERS, for MALDI the variation of the nutrient factor is higher than the variation contribution of the temperature.
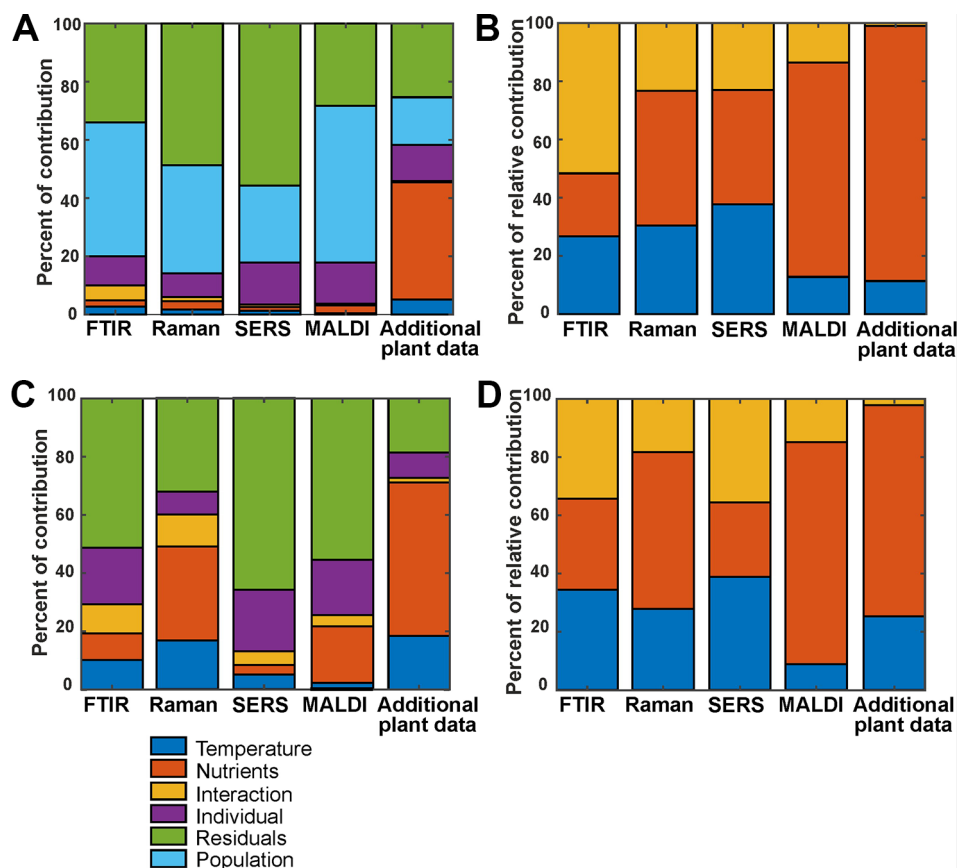
FIGURE 4 | (A) Variation contribution of the design factors temperature (blue), nutrients (orange), the interaction of temperature and nutrients (yellow), different individuals (purple), populations (cyan), and residual variance (green) for the 72 spectra from the whole data set. **(B)** Relative contribution of temperature (blue), nutrients (orange), and the interaction of both (yellow) for the 72 spectra from the whole data set **(C)** and variation contribution of the design factors temperature (blue), nutrients (orange), the interaction of temperature and nutrients (yellow), individuals (purple) and the residuals (green) for the 24 spectra from the population Italy. **(D)** Relative contribution of temperature (blue), nutrients (orange), and the interaction of both (yellow) for the 24 spectra from the population Italy. In **(B, D)**, the contribution to the variance by specific population and the residual variance were left out, and the variation of each factor was normalized by the sum of the variations for the three other factors of interest.

The contribution of the different design factors to the total variation were also analyzed for each population separately (**Figures 4C, D**, **Figure S5**). As an example, **Figures 4C, D** show the outcome of the analysis for the population Italy. For the population Sweden (**Figures S5A, B**), the overall variation contribution of the individuals (**Figure S5A**, purple), is higher compared to the other populations, and contribution of variation due to the growth conditions is rather small.

This type of analysis helps understanding the underlying variation in the data introduced by different design factors and by other unwanted factors. PCA analysis and other multivariate data analysis techniques, if successfully working on the data, ensure that the amount of relevant variation available in the data is sufficient to discriminate between groups. As an example, although the different growth conditions contribute to only 10% of the variation in the FTIR data from all populations (**Figure 4A**, first column), we observed a good discrimination of growth conditions using the first ten PCs, yielding a d-value of 3 (**Table 1**, first section). This shows that the methods are powerful

enough to focus on the relevant information in the data, and that the residual variation is not systematic. Regarding the hierarchical nature of the variance, the results of the ASCA approach are in good agreement with the results obtained by PCA. In data sets that show large contributions by different sources of variation, separation in a PCA is not unequivocal (see **Table 1**).

In conclusion, the different analytical methods vary greatly in their potential to discriminate the pollen from the sample set based on population and environmental influences. For FTIR spectroscopy (Zimmermann et al., 2017) and MALDI-MS (Diehn et al., 2018) this has been discussed previously. Due to the different selectivity in MALDI compared to FTIR, there is a superposition by the variation between the different genotypes (that is, individual variation) that impairs the discrimination ability for different growth conditions within one population (Diehn et al., 2018). While both Raman micro-spectroscopy of single pollen grains and SERS enable classification of the pollen samples with respect to the corresponding population, no strong

variation is found when these data sets are used to assess separation according to the varied environmental conditions of the parental plants. Nevertheless, the variation due to varied growth conditions is highly dependent on the considered population.

CPCA for the Classification of Pollen Samples According to Plant Populations

With consensus principal component analysis (CPCA), the five individual data blocks can be combined, and the impact of each method on a global analysis can be evaluated. **Figure 5** shows the results of the CPCA for the classification of the different populations of *Poa alpina*, consisting of five block score plots (**Figures 5B–F**) that correspond to the different analyses, and of a global scores plot (**Figure 5A**). The global score values of the first and second PC (**Figure 5A**) show a clear discrimination of the three different populations. In particular, based on the variance represented by CPC1, data from the population Norway and data from the population Italy are separated. As revealed by the block scores plots, the first component is mostly influenced by the FTIR block, comprising 41.7% explained variance (**Figure 5B**) and the MALDI block, explaining 39.62% of the variance (**Figure 5E**). The second PC is influenced in particular by the SERS data, explaining 37.55% of the variance (**Figure 5D**) and the block with the data on the parent plants, explaining 21.84% of the variance (**Figure 5F**). In all of the scores plots, the data sets of the population Sweden have positive score values, while the data sets of the populations Italy and Norway have mostly negative values regarding CPC2 (**Figure 5**),

particularly for the Raman (**Figure 5C**), SERS (**Figure 5D**), and MALDI (**Figure 5E**) block. A CPCA containing FTIR, Raman, SERS and MALDI without the additional plant information leads to very similar scores plots, where also all three populations would be discriminated (**Figure S6**).

In **Figure 6**, the results of the separation of the respective first CPC are summarized in box plots for each block as well as for the global scores (**Figure 6**). Furthermore, we calculated a d-value of 2 based on the CPCA scores of CPC1 to CPC10 for the global scores as well as for all block scores. The data indicate that separation of the three populations is readily achieved based on the global scores (**Figure 6A**), and that the FTIR (**Figure 6B**) and the MALDI data sets (**Figure 6E**) have the greatest influence on the separation in the global scores.

In order to analyze which variables of the respective methods cause the separation in the global analysis and to investigate the correlations between them, a correlation loadings plot was generated (**Figure 7**). It shows the correlation between the global scores of the populations Sweden (red cross), Italy (red circle) and Norway (red triangle) and the relevant variables of the different blocks. For the clarity only the extrema of the loadings of the first and second component from the spectroscopic and MALDI blocks are shown, as well as all five variables from the additional plant data. Therefore, there are no variables visible close to the origin of the plot. The different populations are characterized by variables that are located close to the global scores of the populations. The separation of the data from the population Sweden is caused by a high amount of spikes in the
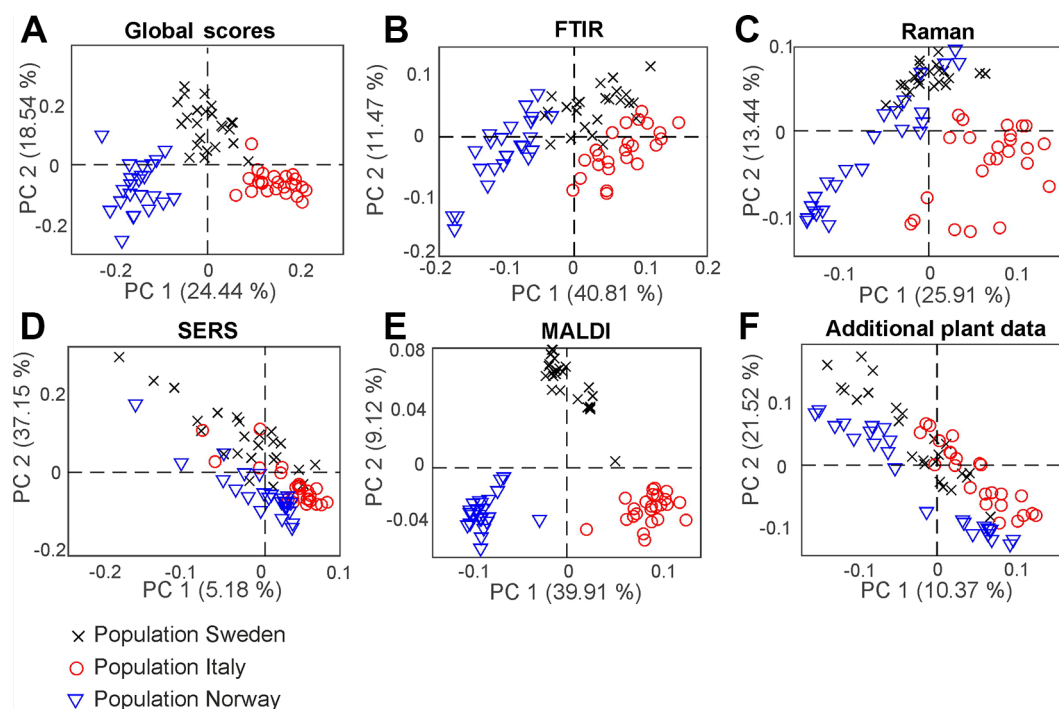
FIGURE 5 | Score values of the CPCA analysis for the classification of samples from the populations Sweden (black crosses), Italy (red circles), and Norway (blue triangles). **(A)** Scores for the global scores **(B–F)** individual data blocks. CPCA, consensus principal component analysis.

FIGURE 6 | Continued

FIGURE 6 | Results of the consensus principal component analysis of the five data sets visualized by box plots of the global scores **(A)** and the blocks scores for FTIR **(B)**, Raman **(C)**, SERS **(D)**, MALDI-TOF MS **(E)**, and additional plant data **(F)**. The box plots display the variation of the score values of the first CPC regarding each population obtained by Kruskal-Wallis H-Test. Red lines indicate the median of the respective distribution, blue boxes represent the interquartile range, the black lines demarcate minimum and maximum values, and extreme values are shown as red markers. FTIR, Fourier-transform infrared spectroscopy; Raman, Raman spectroscopy; SERS, surface enhanced Raman scattering; MALDI-TOF MS, matrix assisted laser desorption/ionization mass spectrometry.

respective progenitor plants and their high dry mass. In addition, this population is characterized by Raman bands at 1007, 1161, and 1529 cm^{-1} that can be assigned to carotenoids (Schulte et al.,
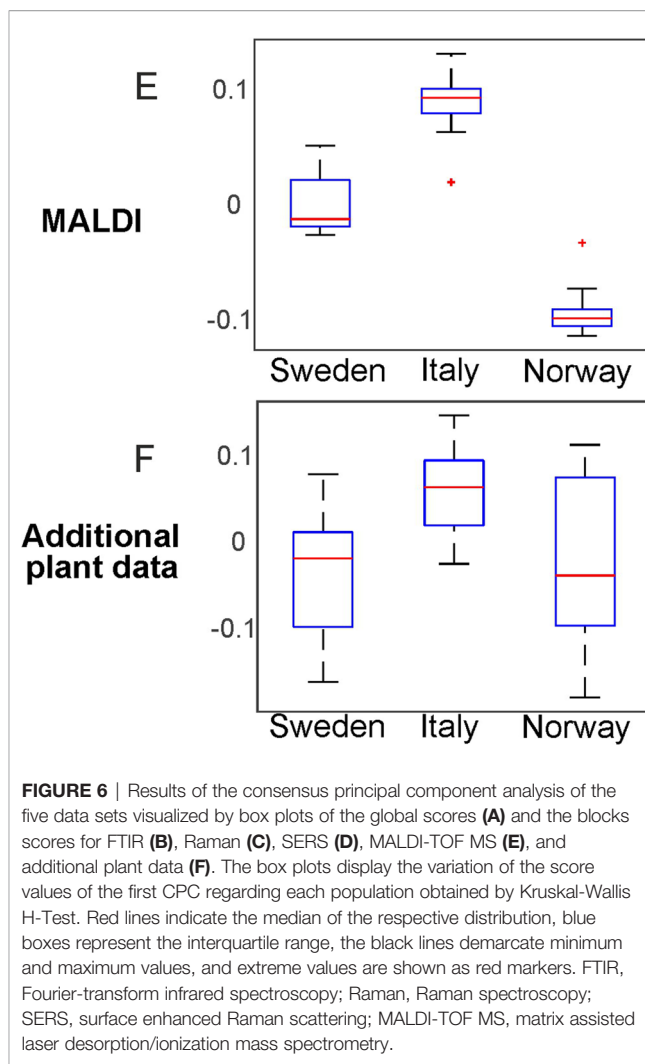
2009) as well as by bands at 555 cm^{-1} that can be assigned to proteins (Schulte et al., 2008), and a MALDI peak at m/z 6038. The great influence that the SERS data block has on CPC2, separating population Sweden (see **Figure 5C**), reflects in a correlation with SERS signals at 416, 733, 994, 1154 and 1545 cm^{-1} that are particularly important to discriminate the pollen data from the population Sweden (**Figure 7**, magenta markers). In the two other populations, SERS signals at 581, 774, 1051, 1379, and 1424 cm^{-1} are observed. They might be attributed to the water-soluble part of proteins or carbohydrates.

The differentiation between the populations from Norway and Italy is achieved utilizing CPC1. The population from Italy is mainly separated by chemical information contained in the FTIR bands (**Figure 7**, blue markers) at 1026, 1079, 1151, 1472, 1525, 1649, and 1688 cm^{-1}, Raman bands (**Figure 7**, green markers) at 484, 649, 948, and 1609 cm^{-1}, and MALDI TOF MS peaks (**Figure 7**, yellow markers) at m/z 5282, 5968, 5980, and 6264. The FTIR and Raman bands can be assigned to starch, protein and sporopollenin vibrations (Schulte et al., 2008; Zimmermann, 2010; Bagcioglu et al., 2015). Although an assignment of the MALDI peaks is more challenging, their positive correlation with
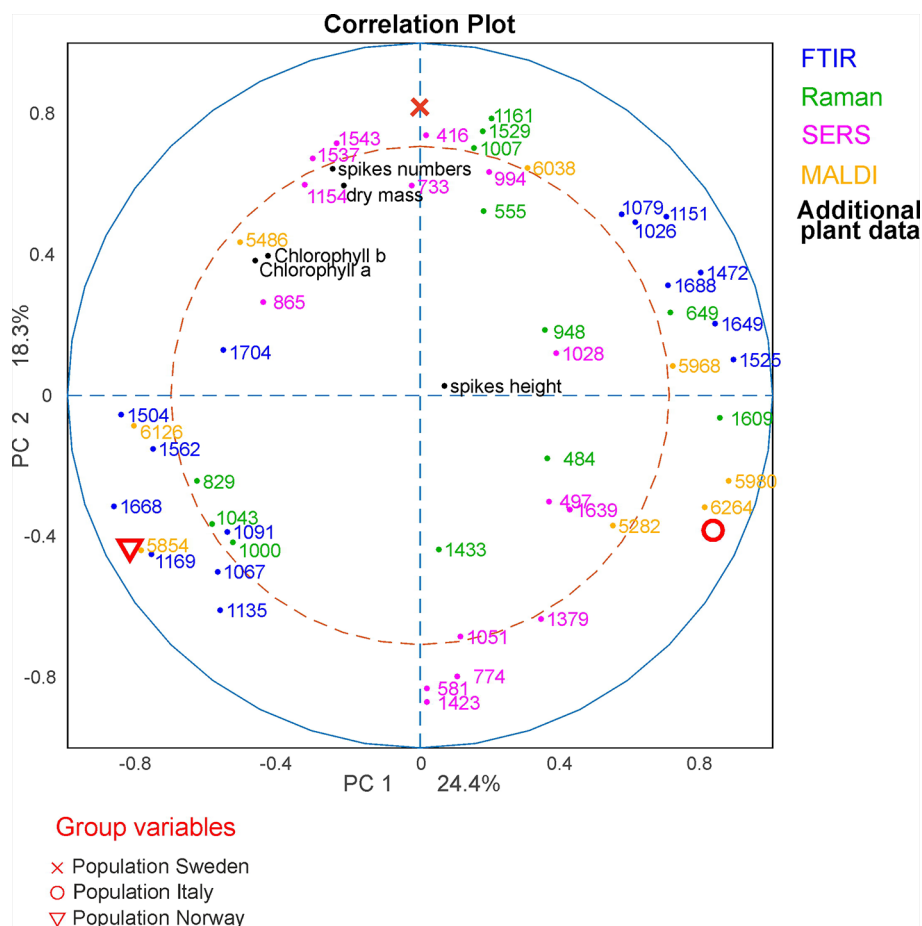
FIGURE 7 | CPCA correlation loadings plot for the 1st and 2nd CPCA component. Displayed are the global scores of the three populations Sweden, (red cross), Italy (red circle), and Norway (red triangle), as well as the loadings of the blocks of FTIR, Raman SERS, MALDI-TOF and additional plant data. For clarity only extrema of the loadings were shown for the spectroscopic/spectrometric data. CPCA, consensus principal component analysis; FTIR, Fourier-transform infrared spectroscopy; Raman, Raman spectroscopy; SERS, surface enhanced Raman scattering; MALDI-TOF MS, matrix assisted laser desorption/ionization mass spectrometry.

these bands suggests that some of them are connected to nutrients, in agreement with previous discussions suggesting their assignment to oligosaccharides (Krause et al., 2012; Seifert et al., 2015; Diehn et al., 2018).

The data sets of the population Norway show a positive correlation to the FTIR vibrational bands at 1089, 1166, 1503, 1666, and 1746 cm^{-1} as well as to the MALDI peaks at m/z 5880 and 6296 (**Figure 7**, bottom left section). Variances in Raman bands at 829 and 1043 cm^{-1} are positively correlated to the population Norway. Most of the Raman bands can be assigned as protein vibrations (De Gelder et al., 2007), whereas the FTIR bands could be mainly assigned to carbohydrates (Bagcioglu et al., 2015; Zimmermann et al., 2017).

As illustrated by the band assignments, in addition to a redundancy in information (e.g., in some bands in FTIR and Raman spectra) each data block contains some exclusive molecular information, leading to their complementarity. The different contribution of the five data blocks in the discrimination of the three populations shown in the correlation plot (**Figure 7**)

indicates that particular parts of the pollen chemistry are responsible for the differences between populations, and that very different molecular/compositional parameters are responsible in the biochemical variation between two populations. The MALDI-TOF MS data have great influence on the analysis and can be exploited for a precise discrimination of all three populations. This is in accordance with the results of the PCA of the isolated data block above (**Figure 3D**) in this paper and supports previous results that indicate that MALDI-TOF MS and the biochemical fingerprint of glycoproteins and other macromolecules are specific for the pollen of a particular grass population (Diehn et al., 2018).

CPCA for the Classification of Pollen Samples According to Different Environmental Influences

CPCA was applied as well to discriminate between pollen samples within each population that were collected from progenitor plants grown under four different environmental

conditions: 14°C and additional nutrients, 14°C without additional nutrients, 20°C and additional nutrients, 20°C without additional nutrients. **Table 2** shows the resulting p and d-values analyzing the whole data set from all populations and the data from each of the three different populations individually for the global scores (**Table 2**, first section) and all the block scores (second to sixth section, respectively). The p-values for the global scores are below 0.05 for each population, indicating the separation of the different groups in the first CPCA component (**Table 2**, first section). However, considering all three populations together, separation is based on the third CPCA component. MANOVA utilizing the first ten CPCA components shows the highest possible d-value of 3, proving successful classification of all four groups of samples for population Italy, as well as the for the whole sample set of all three populations. The lower d-value for the global scores in the population Sweden and Norway may be explained by a lower phenotypic plasticity of these populations compared to the population Italy. This is in good agreement with previous analyses of other data of some of the samples discussed here (Zimmermann et al., 2017; Diehn et al., 2018).

Comparison of the results for the block scores (**Table 2**, second to sixth section) will help to identify those data blocks that are responsible for a separation based on the global scores. Based on the d-values, a separation of the samples into four groups—corresponding to four environmental conditions—is observed when all populations are analyzed together (last line in each of the sections of **Table 2**). The separation into four groups is possible for each of the five block scores except those of the MALDI block (last line in section 5 of **Table 2**). By interpreting the corresponding dendrogram shown in **Figure S7**, these three groups correspond to the condition 14°C without additional nutrients, 20°C without additional nutrients, and plants that obtained additional nutrients (regardless of growth temperature). The Raman block scores indicate separation of the four groups in the two populations Norway and Italy (**Table 2**, third section). For the other block scores (FTIR, SERS, and MALDI), the separate analysis of each of the populations gives very different results, with the samples from the population Italy showing separation according to the four growth conditions (**Figure 1**) in most of them, but less than four distinct groups in the populations Sweden and Norway. The block scores for the data gathered from the parent plants show very similar behavior and result in clear classification of all four conditions only in population Italy (**Table 2**, sixth section).

The weighting of each block in the CPCA can be interpreted and allows more insight into the influence the data blocks on each other. As an example, **Figure 8** shows the results for the CPCA applied to the data of the pollen samples from the population Italy. The first component of the global scores plot (**Figure 8A**) separates between positive scores values of the data of samples from progenitor plants that were grown with the addition of nutrients (black crosses and blue triangles) and negative score values for samples from plants that were grown without the addition of nutrients (red circles and green diamonds). In **Figures 8C, F**, the great influence of the Raman and the additional plant

data block, respectively, are revealed. Both blocks display similar group formation in the scores plots, with high variances explained by the first CPC of 35.52% and 54.50%, respectively. The corresponding p-values in **Table 2** are very low.

The scores of the second CPCA component separate pollen samples grown at 14 °C, as well as at 20 °C without additional nutrients (black crosses, red circles and green diamonds) with positive values from negative values of those pollen samples grown at 20°C without additional nutrients (blue triangles) (**Figure 8A**). The CPC2 is mainly influenced by the SERS data (**Figure 8D**), explaining 33.77% of the variance. In the plot of the block score values (**Figure 8D**), no separation of the groups that could correspond to growth conditions of the plants can be found. This suggests that other sources of variance, in this experiment resulting from individual genotypes, superimpose the influence of the growth conditions as discussed for other data previously (Diehn et al., 2018). It is also in agreement with the calculated p- and d-values for the SERS block (**Table 2**, section 4). Furthermore, the Raman block scores plot, as well as the scores from the additional plant data, show great potential regarding the discrimination of different growth conditions in the population Italy. Since the additional plant data block explains most of the variance in the first CPC, CPCA was also performed without it, by using only the spectroscopic/ spectrometric data blocks, in order to confirm that the obtained global pattern is also driven only by the pollen chemical composition (compare with **Figure S8**), not by phenotypic features of the parent plant. Nevertheless, the additional plant data lead to a more complete view in this study and show correlation to the spectroscopic data blocks (compare **Figure 9**).

The molecular differences that cause the separation of the data reveal themselves in the correlation loadings plot for the data from population Italy (**Figure 9**). Again, only those loadings with the highest impact are shown for clarity and only the variables of the additional plant data were presented in full. As expected after the discussion of the block scores (**Figure 8**), the first CPCA component that separates samples from plants grown with additional nutrients (crosses and triangles) from samples without additional nutrients (circles and diamonds, also compare **Figure 8**) is mainly influenced by the Raman block and the additional plant data.

Raman bands that represent pollen samples with nutrient addition are 474, 830, 1003, 1435, and 1602 cm^{-1}. The bands at 1435 cm^{-1} and 1602 cm^{-1} can be assigned to lipids (Ivleva et al., 2005; Schulte et al., 2008) and a high mitochondrial activity, respectively (Huang et al., 2004; Pully and Otto, 2009). The other bands are associated with proteins (Schulte et al., 2008; Bagcioglu et al., 2017). The negative scores of the first CPC and the data of the pollen samples without additional nutrients (**Figure 9**, diamonds and circles) are mainly influenced by Raman bands at 485, 949, 1010, 1138, and 1471 cm^{-1}. These bands are associated with carbohydrates, such as starch (Schulte et al., 2008; Pigorsch, 2009; Bagcioglu et al., 2017). Pollen are storing their nutrients in lipid bodies as well as in starch bodies, which are occupying most of the space in pollen grains (Wang et al., 2015). Our results confirm that plants growing under different
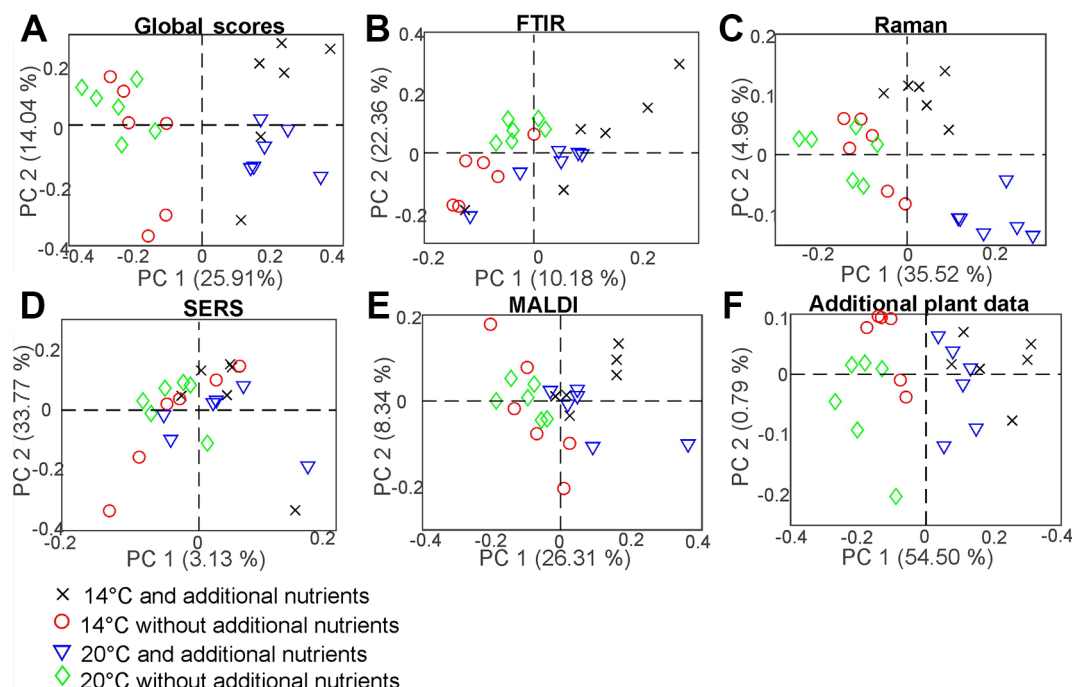
FIGURE 8 | Scores of the CPCA analysis for the classification of samples from pollen of the population Italy regarding the four different growth conditions 14°C and additional nutrients (black crosses), 14°C without additional nutrients (red circles), 20°C and additional nutrients (blue triangles), and 20°C without additional nutrients (green diamonds) for **(A)** the global scores and **(B–F)** the five block scores.

nutrient conditions vary in their quality and/ or amount of such storage bodies inside the pollen.

The second component can be used to separate between rather positive scores values corresponding to samples that were grown at low temperatures (crosses and circles) and negative scores values corresponding to samples that were grown at higher temperatures (diamonds and triangles). As discussed before (compare **Figure 8**), this separation is mainly influenced by SERS and FTIR bands. In particular, samples from plants grown at lower temperatures are characterized by a set of SERS bands that include 445 cm^{-1} and the FTIR bands at 1721 and 1475 cm^{-1}. The FTIR signals can be assigned to lipids (Bagcioglu et al., 2015). Samples grown under higher temperatures are characterized by SERS bands at 419, 929, 957, and 1,564 cm^{-1}, and FTIR bands at 1,666 and 1,503 cm^{-1}. The bands could be assigned to nucleobases and proteins (Seifert et al., 2016). Based on the influence the combination of SERS and FTIR data, we can assume that the discrimination regarding the different growth conditions is probably mostly influenced by the chemical composition of the pollen interior, although –in the preparations for SERS experiments- also water soluble compounds from the pollen outer shell may be found in the aqueous extract.

To summarize the results from the correlations loadings plot, discrimination of different nutrient conditions is mainly influenced by Raman bands that can be assigned to pollen outer shell and nutrient storage, as well as by plant parameters that are present in the additional plant data block. From our results, we infer on differences in amount and quality in lipid and starch bodies inside the pollen grains to be responsible for a distinction of samples from plants grown at different nutrient conditions. This is in good agreement with previous studies on *Poa alpina* using only FTIR spectroscopy (Zimmermann et al., 2017). The temperature conditions at which parent plants are grown mainly affects the SERS and FTIR data blocks and, probably, mainly the chemical composition of the interior of the pollen grains. It has to be pointed out that this conclusion is only made based on the data of the pollen from population Italy, were the samples are showing the highest phenotypic plasticity of the three investigated populations. Within the other populations, the correlation of the signals can differ greatly, indicating higher phenotypic rigidity, as discussed above.

CONCLUSION

A well described sample set of pollen from *Poa alpina* was analyzed by FTIR spectroscopy, Raman microspectroscopy, surface enhanced Raman scattering (SERS) and MALDI TOF mass spectrometry, as well as by collected additional data from the parental plants. The chosen methods are complementary regarding sample preparations, selectivity, and sensitivity of the analytical technique. Our results show the ability to detect and describe variances within the pollen composition related either to the place of origin of parent plants (i.e., populations) or the growth conditions. However, suitable data analysis is needed in order to discuss the relatively small chemical differences in these complex biological samples.

FIGURE 9 | CPCA Correlation loadings plot for the 1st and 2nd CPCA component. Displayed are the global scores of population Italy regarding the four growth conditions 14°C and additional nutrients, (black crosses); 14°C without additional nutrients, (red circles); 20°C and additional nutrients (blue triangles); 20°C without additional nutrients (green diamonds), as well as the loadings of the blocks of FTIR, Raman SERS, MALDI-TOF and additional plant data. For clarity only extrema of the loadings are shown for the spectroscopic/spectrometric data. CPCA, consensus principal component analysis; FTIR, Fourier-transform infrared spectroscopy; Raman, Raman spectroscopy; SERS, surface enhanced Raman scattering; MALDI-TOF MS, matrix assisted laser desorption/ionization mass spectrometry.

The sample set is designed using plants from three populations that were grown under different nutrient conditions and temperatures. Therefore, different levels of classification and influence on pollen composition could be analyzed. As expected, the separation of groups in the sample set according to populations and growth conditions, respectively, is not achieved equally well by each of the methods. As shown by an analysis of different sources of variance using ASCA, different analytical techniques are emphasizing different parameters of pollen chemistry related either to the genetic background or the environmental influence. This has been suggested in previous work where data from FTIR (Zimmermann et al., 2017) and MALDI-TOF-MS (Diehn et al., 2018) on a similar sample set were analyzed but has been shown here using three more, very different types of data. By combination of the different data blocks in a CPCA, a complete set of many differences, observed

with the complementary methods can be used to describe the variation with respect to the different groups. We have also compared the individual classification ability of the different methods and the different levels using PCA in combination with simple statistical tests. The different populations can be easily distinguished using MALDI-TOF MS, whereas the three spectroscopic methods are more suitable to separate between different growth conditions. Moreover, as discussed, the same data blocks can have a different influence on the distinction between different growth conditions in the three populations. This implies that, due to the different fraction of the pollen chemistry that is represented by each data block (or analyzed by each of the methods), the biochemical effect of the growth conditions on pollen chemistry can vary for different populations. This would be in agreement with variation in phenotypic plasticity between the populations, in particular

regarding different metabolic and molecular pathways used in environmental adaptation.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

AK, BZ, JK, MO, and SF conceived the research idea. SF designed the growth experiments. AK, BZ, and MB designed the FTIR experiments. JK, SD, and SS designed the Raman and SERS experiments. JK, SD, SS, and SW designed the mass spectrometry experiments. BZ and MB performed sampling, FTIR experiments and the measurement of the additional plant data. SD performed the mass spectrometry, Raman and SERS experiments. BZ, SD, and VT analyzed the data. JK and SD wrote the article. AK, BZ, MB, MO, SF, SS, SW, and VT discussed and revised the article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01788/full#supplementary-material

REFERENCES

Bagcioglu, M., Zimmermann, B., and Kohler, A. (2015). A multiscale vibrational spectroscopic approach for identification and biochemical characterization of pollen. *PloS One* 10, e0137899. doi: 10.1371/journal.pone.0137899

Bagcioglu, M., Kohler, A., Seifert, S., Kneipp, J., and Zimmermann, B. (2017). Monitoring of plant-environment interactions by high-throughput FTIR spectroscopy of pollen. *Methods Ecol. Evol.* 8, 870–880. doi: 10.1111/2041-210X.12697

Blokker, P., Boelen, P., Broekman, R., and Rozema, J. (2006). The occurrence of p-coumaric acid and ferulic acid in fossil plant materials and their use as UV-proxy. *Plant Ecol.* 182, 197. doi: 10.1007/978-1-4020-4443-4_13

Bro, R., and Smilde, A. K. (2014). Principal component analysis. *Anal. Methods* 6, 2812–2831. doi: 10.1039/C3AY41907J

De Gelder, J., De Gussem, K., Vandenabeele, P., and Moens, L. (2007). Reference database of Raman spectra of biological molecules. *J. Raman Spectrosc.* 38, 1133–1147. doi: 10.1002/jrs.1734

De Haan, J. R., Wehrens, R., Bauerschmidt, S., Piek, E., Van Schaik, R. C., and Buydens, L. M. C. (2007). Interpretation of ANOVA models for microarray data using PCA. *Bioinformatics* 23, 184–190. doi: 10.1093/bioinformatics/btl572

Dell'Anna, R., Lazzeri, P., Frisanco, M., Monti, F., Campeggi, F. M., Gottardini, E., et al. (2009). Pollen discrimination and classification by Fourier transform infrared (FT-IR) microspectroscopy and machine learning. *Anal. Bioanal. Chem.* 394, 1443–1452. doi: 10.1007/s00216-009-2794-9

Depciuch, J., Kasprzyk, I., Drzymala, E., and Parlinska-Wojtan, M. (2018). Identification of birch pollen species using FTIR spectroscopy. *Aerobiologia (Bologna)* 34, 525–538. doi: 10.1007/s10453-018-9528-4

Diehn, S., Zimmermann, B., Bağcıoğlu, M., Seifert, S., Kohler, A., Ohlson, M., et al. (2018). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) shows adaptation of grass pollen composition. *Sci. Rep.* 8, 16591. doi: 10.1038/s41598-018-34800-1

Eilers, P. H. C. (2003). A perfect smoother. *Anal. Chem.* 75, 3631–3636. doi: 10.1021/ac034173t

Ellis, D. I., and Goodacre, R. (2006). Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. *Analyst* 131, 875–885. doi: 10.1039/b602376m

Gottardini, E., Rossi, S., Cristofolini, F., and Benedetti, L. (2007). Use of Fourier transform infrared (FT-IR) spectroscopy as a tool for pollen identification. *Aerobiologia* 23, 211–219. doi: 10.1007/s10453-007-9065-z

Harrington, P. D. B., Vieira, N. E., Espinoza, J., Nien, J. K., Romero, R., and Yergey, A. L. (2005). Analysis of variance-principal component analysis: a soft tool for proteomic discovery. *Anal. Chim. Acta* 544, 118–127. doi: 10.1016/j.aca.2005.02.042

Hassani, S., Martens, H., Qannari, E. M., Hanafi, M., Borge, G. I., and Kohler, A. (2010). Analysis of -omics data: graphical interpretation- and validation tools in multi-block methods. *Chemom. Intell. Lab. Syst.* 104, 140–153. doi: 10.1016/j.chemolab.2010.08.008

Hassani, S., Hanafi, M., Qannari, E., and Kohler, A. (2013). Deflation strategies for multi-block principal component analysis revisited. *Chemometrics Intell. Lab. Syst.* 120, 154–168. doi: 10.1016/j.chemolab.2012.08.011

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24, 417–441. doi: 10.1037/h0071325

Huang, Y. S., Karashima, T., Yamamoto, M., Ogura, T., and Hamaguchi, H. O. (2004). Raman spectroscopic signature of life in a living yeast cell. *J. Raman Spectrosc.* 35, 525–526. doi: 10.1002/jrs.1219

Ivleva, N., Niessner, R., and Panne, U. (2005). Characterization and discrimination of pollen by Raman microscopy. *Anal. Bioanal. Chem.* 381, 261–267. doi: 10.1007/s00216-004-2942-1

Jansen, J. J., Hoefsloot, H. C. J., Van Der Greef, J., Timmerman, M. E., Westerhuis, J. A., and Smilde, A. K. (2005). ASCA: analysis of multivariate data obtained from an experimental design. *J. Chemometr.* 19, 469–481. doi: 10.1002/cem.952

Jardine, P. E., Gosling, W. D., Lomax, B. H., Julier, A. C. M., and Fraser, W. T. (2019). Chemotaxonomy of domesticated grasses: a pathway to understanding the origins of agriculture. *J. Micropalaeontol.* 38, 83–95. doi: 10.5194/jm-38-83-2019

Joester, M., Seifert, S., Emmerling, F., and Kneipp, J. (2017). Physiological influence of silica on germinating pollen as shown by Raman spectroscopy. *J. Biophotonics* 10, 542–552. doi: 10.1002/jbio.201600011

Joseph, V., Schulte, F., Rooch, H., Feldmann, I., Dorfel, I., Osterle, W., et al. (2011). Surface-enhanced Raman scattering with silver nanostructures generated *in situ* in a sporopollenin biopolymer matrix. *Chem. Commun.* 47, 3236–3238. doi: 10.1039/c0cc05326k

Julier, A. C. M., Jardine, P. E., Coe, A. L., Gosling, W. D., Lomax, B. H., and Fraser, W. T. (2016). Chemotaxonomy as a tool for interpreting the cryptic diversity of Poaceae pollen. *Rev. Palaeobot. Palynol.* 235, 140–147. doi: 10.1016/j.revpalbo.2016.08.004

Krause, B., Seifert, S., Panne, U., Kneipp, J., and Weidner, S. M. (2012). Matrix-assisted laser desorption/ionization mass spectrometric investigation of pollen and their classification by multivariate statistics. *Rapid Commun. Mass Spectrom.* 26, 1032–1038. doi: 10.1002/rcm.6202

Kyu, K. S., Soo, K. M., and Won, S. S. (1987). Surface-enhanced Raman scattering (SERS) of aromatic amino acids and their glycyl dipeptides in silver sol. *J. Raman Spectrosc.* 18, 171–175. doi: 10.1002/jrs.1250180305

Lasch, P., and Naumann, D. (1998). FT-IR microspectroscopic imaging of human carcinoma thin sections based on pattern recognition techniques. *Cell. Mol. Biol. (Noisy-le-grand)* 44, 189–202. doi: 10.1063/1.55833

Lauer, F., Diehn, S., Seifert, S., Kneipp, J., Sauerland, V., Barahona, C., et al. (2018). Multivariate analysis of MALDI imaging mass spectrometry data of mixtures of single pollen grains. *J. Am. Soc. Mass Spectrom.* 29, 2237–2247. doi: 10.1007/s13361-018-2036-5

Lee, P. C., and Meisel, D. (1982). Adsorption and surface-enhanced Raman of dyes on silver and gold sols. *J. Phys. Chem.* 86, 3391–3395. doi: 10.1021/j100214a025

Li, F.-S., Phyo, P., Jacobowitz, J., Hong, M., and Weng, J.-K. (2019). The molecular structure of plant sporopollenin. *Nat. Plants* 5, 41–46. doi: 10.1038/s41477-018-0330-7

Martens, H., and Stark, E. (1991). Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. *J. Pharm. Biomed. Anal.* 9, 625–635. doi: 10.1016/0731-7085(91)80188-F

Pappas, C. S., Tarantilis, P. A., Harizanis, P. C., and Polissiou, M. G. (2003). New method for pollen identification by FT-IR spectroscopy. *Appl. Spectrosc.* 57, 23–27. doi: 10.1366/000370203321165160

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *London Edinburgh Dublin Philos. Mag. J. Sci.* 2, 559–572. doi: 10.1080/14786440109462720

Perisic, N., Afseth, N. K., Ofstad, R., Hassani, S., and Kohler, A. (2013). Characterising protein, salt and water interactions with combined vibrational spectroscopic techniques. *Food Chem.* 138, 679–686. doi: 10.1016/j.foodchem.2012.10.117

Pigorsch, E. (2009). Spectroscopic characterisation of cationic quaternary ammonium starches. *Starch - Stärke* 61, 129–138. doi: 10.1002/star.200800090

Porra, R. J., Thompson, W. A., and Kriedemann, P. E. (1989). Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochimica Biophysica Acta (BBA) - Bioenergetics* 975, 384–394. doi: 10.1016/S0005-2728(89)80347-0

Pully, V. V., and Otto, C. (2009). The intensity of the 1602 cm−1 band in human cells is related to mitochondrial activity. *J. Raman Spectrosc.* 40, 473–475. doi: 10.1002/jrs.2269

Qian, J., Cutler, J. E., Cole, R. B., and Cai, Y. (2008). MALDI-TOF mass signatures for differentiation of yeast species, strain grouping and monitoring of morphogenesis markers. *Anal. Bioanal. Chem.* 392, 439–449. doi: 10.1007/s00216-008-2288-1

Rozema, J., Broekman, R. A., Blokker, P., Meijkamp, B. B., De Bakker, N., Van De Staaij, J., et al. (2001). UV-B absorbance and UV-B absorbing compounds (para-coumaric acid) in pollen and sporopollenin: the perspective to track historic UV-B levels. *J. Photochem. Photobiol. B: Biol.* 62, 108–117. doi: 10.1016/S1011-1344(01)00155-5

Savitzky, A., and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36, 1627–1639. doi: 10.1021/ac60214a047

Schulte, F., Lingott, J., Panne, U., and Kneipp, J. (2008). Chemical characterization and classification of pollen. *Anal. Chem.* 80, 9551–9556. doi: 10.1021/ac801791a

Schulte, F., Mader, J., Kroh, L. W., Panne, U., and Kneipp, J. (2009). Characterization of pollen carotenoids with *in situ* and high-performance thin-layer chromatography supported resonant raman spectroscopy. *Anal. Chem.* 81, 8426–8433. doi: 10.1021/ac901389p

Seifert, S., Weidner, S. M., Panne, U., and Kneipp, J. (2015). Taxonomic relationships of pollens from matrix-assisted laser desorption/ionization time-of-flight mass spectrometry data using multivariate statistics. *Rapid Commun. Mass Spectrom.* 29, 1145–1154. doi: 10.1002/rcm.7207

Seifert, S., Merk, V., and Kneipp, J. (2016). Identification of aqueous pollen extracts using surface enhanced Raman scattering (SERS) and pattern recognition methods. *J. Biophotonics* 9, 181–189. doi: 10.1002/jbio.201500176

Sengupta, A., Laucks, M. L., and Davis, E. J. (2005). Surface-enhanced Raman spectroscopy of bacteria and pollen. *Appl. Spectrosc.* 59, 1016–1023. doi: 10.1366/0003702054615124

Smilde, A. K., Jansen, J. J., Hoefsloot, H. C. J., Lamers, R. J., Van Der Greef, J., and Timmerman, M. E. (2005). ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics* 21, 3043–3048. doi: 10.1093/bioinformatics/bti476

Solhaug, K. A. (1991). Influence of photoperiod and temperature on dry matter production and chlorophyll content in temperate grasses [also incl. net assimilation rate, NAR, long days, short days]. *Norwegian J. Agric. Sci.* 5, 365–384.

Stewart, S., and Fredericks, P. M. (1999). Surface-enhanced Raman spectroscopy of amino acids adsorbed on an electrochemically prepared silver surface. *Spectroc. Acta Pt. A-Molec. Biomolec. Spectr.* 55, 1641–1660. doi: 10.1016/S1386-1425(98)00294-7

Wang, S., Wang, D., Wu, Q., Gao, K., Wang, Z., and Wu, Z. (2015). 3D imaging of a rice pollen grain using transmission X-ray microscopy. *J. Synchrotron Radiat.* 22, 1091–1095. doi: 10.1107/S1600577515009716

Westerhuis, J. A., Kourti, T., and Macgregor, J. F. (1998). Analysis of multiblock and hierarchical PCA and PLS models. *J. Chemometr.* 12, 301–321. doi: 10.1002/(SICI)1099-128X(199809/10)12:5<301::AID-CEM515>3.0.CO;2-S

Wold, S., Hellberg, S., Lundstedt, T., Sjostrom, M., and Wold, H. (1987). Proc. symp. on pls model building: theory and application. *Frankfurt am Main*.

Zimmermann, B., and Kohler, A. (2013). Optimizing Savitzky-Golay parameters for improving spectral resolution and quantification in infrared spectroscopy. *Appl. Spectrosc.* 67, 892–902. doi: 10.1366/12-06723

Zimmermann, B., and Kohler, A. (2014). Infrared spectroscopy of pollen identifies plant species and genus as well as environmental conditions. *PLoS One* 9, e95417. doi: 10.1371/journal.pone.0095417

Zimmermann, B., Bağcıoğlu, M., Tafintseva, V., Kohler, A., Ohlson, M., and Fjellheim, S. (2017). A high-throughput FTIR spectroscopy approach to assess adaptive variation in the chemical composition of pollen. *Ecol. Evol.* 7, 10839–10849. doi: 10.1002/ece3.3619

Zimmermann, B. (2010). Characterization of pollen by vibrational spectroscopy. *Appl. Spectrosc.* 64, 1364–1373. doi: 10.1366/000370210793561664

Understanding the Formation of Heartwood in Larch Using Synchrotron Infrared Imaging Combined With Multivariate Analysis and Atomic Force Microscope Infrared Spectroscopy

Sara Piqueras[1]*, Sophie Füchtner[1], Rodrigo Rocha de Oliveira[2], Adrián Gómez-Sánchez[2], Stanislav Jelavić[3,4], Tobias Keplinger[5,6], Anna de Juan[2] and Lisbeth Garbrecht Thygesen[1]

[1] Biomass Science and Technology Group, Department of Geosciences and Natural Resource Management, University of Copenhagen, Frederiksberg, Denmark, [2] Chemometrics Group, Department of Analytical Chemistry, University of Barcelona, Barcelona, Spain, [3] Nano-Science Center, Department of Chemistry, Faculty of Science, University of Copenhagen, Copenhagen, Denmark, [4] Section for GeoGenetics, Faculty of Health and Medical Sciences, Globe Institute, University of Copenhagen, Copenhagen, Denmark, [5] Wood Material Science Group, Department of Construction, Environment and Geomatics, Institute for Building Materials (IfB), ETH Zürich, Zürich, Switzerland, [6] WoodTec Group, Cellulose & Wood Materials, EMPA, Dübendorf, Switzerland

Formation of extractive-rich heartwood is a process in live trees that make them and the wood obtained from them more resistant to fungal degradation. Despite the importance of this natural mechanism, little is known about the deposition pathways and cellular level distribution of extractives. Here we follow heartwood formation in *Larix gmelinii* var. *Japonica* by use of synchrotron infrared images analyzed by the unmixing method Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS). A subset of the specimens was also analyzed using atomic force microscopy infrared spectroscopy. The main spectral changes observed in the transition zone when going from sapwood to heartwood was a decrease in the intensity of a peak at approximately 1660 cm^{-1} and an increase in a peak at approximately 1640 cm^{-1}. There are several possible interpretations of this observation. One possibility that is supported by the MCR-ALS unmixing is that heartwood formation in larch is a type II or *Juglans*-type of heartwood formation, where phenolic precursors to extractives accumulate in the sapwood rays. They are then oxidized and/or condensed in the transition zone and spread to the neighboring cells in the heartwood.

Keywords: heartwood formation, larch, extractives, synchrotron infrared imaging, Atomic Force Microscope Infrared Spectroscopy, Multivariate Curve Resolution – Alternating Least Squares

INTRODUCTION

Heartwood (HW) formation is the final step in the life cycle of ray cells. Before cell death, ray cells undergo metabolic changes in the transition zone between sapwood (SW) and HW, resulting in increased synthesis of secondary metabolic compounds called extractives. The extractives have a significant effect on the properties of wood, most notably regarding its resistance to fungal decay and other forms of biological attack (Hillis, 1987; Hinterstoisser et al., 2000; Schultz and Nicholas, 2000; Taylor et al., 2002). In order to understand how extractives contribute to wood durability, many studies have focused on the chemical interplay between extractives and decay agents (Valette et al., 2017). However, there is an increasing understanding that the distribution of extractives within cell walls also plays an important role, though hitherto under-investigated (Kampe and Magel, 2013).

The process of HW formation has been studied for decades and is known to be associated with parenchyma cell death, disappearance of storage material, and increase in extractives content (Hillis, 1985; Hillis, 1987). Kampe and Magel (2013) described two possible HW formation mechanisms: Type I or *Robinia-Type* proposes the accumulation of the phenolic extractives in the transition zone without any indication of phenolic precursors in the aging SW (Nair et al., 1981; Magel et al., 1994; Bergström et al., 1999); Type II or *Juglans*-Type of HW formation suggests a gradual accumulation of phenolic precursors in the aging SW tissues. In type II, HW extractives are formed in the transition zone by primary and secondary reactions, such as oxidation and hydrolysis of precursor substances (Dellus et al., 1997; Burtin et al., 1998; Mayer et al., 2006). Once the extractives are formed, they are released into the lumina of neighboring cells and cell walls (Déjardin et al., 2010; Kampe and Magel, 2013). When inside the cell walls, a few studies suggest that at least some extractives are covalently bound to the structural cell wall polymers through enzymatic activity (Monties, 1991).

The extractive compounds typically associated with wood durability fall into one of several polyphenolic classes such as flavonoids, stilbenes, lignans, and polymers thereof. The types and quantities of these extractives are species dependent, genetically determined, and under environmental control (Hillis, 1987; Kwon et al., 2001; Taylor et al., 2002; Bito et al., 2011; Bush et al., 2011). In some species, extractives are present in lower amounts (e.g. Spruce 0.9-1.5%) (Willför and Holmbom, 2004) as compared to other species (e.g. larch up to 30%) (Gierlinger et al., 2004).

The genus Larix species (larch) are an important European resource for durable wood (Hillis, 1987). The extractives in larch belong to the molecular families of terpenoids, flavonoids, lignans, fatty acids, and galactans (Zule et al., 2015; Zule et al., 2017). Like all conifers, larch trees contain high amounts of oleoresin, produced by specialized epithelial cells surrounding resin canals and ray parenchyma cells. The resin is composed of fatty acids and esters thereof, as well as various subgroups of terpenoids, and is distributed throughout the HW and SW through the network of resin canals and ray cells (Hillis, 1987).

It was shown for *Pinus sylvestris* that in HW, the composition of resin is enriched with phenolic compounds, presumably produced by ray parenchyma cells during HW formation (Felhofer et al., 2018).

Within the large family of terpenoids, diterpenoid acids (called resin acids) constitute the largest part of the oleoresin (Higuchi, 1997), and they have been repeatedly shown to have fungicidal properties (Keeling and Bohlmann, 2006), which is also the case for triterpenes, known as sterols (Burčová et al., 2018). Triglycerides and fatty acids may have a role in moisture regulation (Tomppo et al., 2011), which is important in the context of degradation by microorganisms. The principal phenolic compounds detectable in larch HW are flavonoids, the main compounds being taxifolin ($C_{15}H_{12}O_7$) and dihydrokaempferol ($C_{15}H_{12}O_6$). Minor amounts of lignans can also be found. Flavonoids have been attributed with fungicidal properties, but their main potential seems to be their ability to scavenge different types of radicals, as well as to reduce and chelate metals (Giwa and Swan, 1975; Cao et al., 1997; Babkin et al., 2001; Ivanova et al., 2012; Zule et al., 2017).

Larch HW is appreciated for its good mechanical properties, its color, and specially for its natural durability (Gierlinger et al., 2004). A strong relationship between extractives content and brown-rot decay resistance has been shown (Gierlinger et al., 2002; Windeisen et al., 2002). Nevertheless, very little is known about the formation and distribution of larch extractives within the xylem tissue at the cell and cell wall level. Their micro and nano-scale distribution is of importance (Taylor et al., 2002) since extractives are more effective against wood degradation within cell walls than in extracellular voids (Hillis, 1987). To investigate extractives in context with the microstructure, TOF-SIMS imaging has been applied in *Cryptomeria japonica* trees and showed that the extractives tend to accumulate near radial rays (Imai et al., 2005; Saito et al., 2008). Recently, the potential of Confocal Raman Microscopy to follow the extractive distribution in sapwood (SW) and heartwood (HW) of Scots pine (*Pinus sylvestris*, a moderately durable species) was shown (Belt et al., 2017; Felhofer et al., 2018). On the micro-level, pinosylvins were reported in the lumen, as well as in the compound middle lamella (CML), cell corner (CC), and pits of tracheid cells.

Synchrotron Radiation Fourier Transform Infrared (SR-FTIR) imaging is the ideal technique to study the extractive deposition patterns at the microscale during HW formation in larch because of the high brightness and high collimation of the beam and avoidance of the fluorescent problems experienced with Raman microspectroscopy. SR-FTIR imaging provides spatial and spectral information about the samples and, therefore, informs on the composition and location of the different sample constituents. Despite the relevant information contained in SR-FTIR images, the analysis of this kind of measurement is not straightforward because of the often large image sizes and the mixed signal components present in the spectra collected. To help in the signal-unmixing task, multivariate analysis tools like Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) are used. Indeed, MCR-ALS has already been proven to adapt particularly well to

hyperspectral image analysis because of the easy introduction of external spectral and spatial information about the image in the analysis and the ability to work with both single and multiset (several) image structures (Tauler et al., 1995; de Juan et al., 2004; Piqueras et al., 2014). This approach is the main tool used in the current study to obtain distribution maps and spectral signatures of the wood sample (Felten et al., 2015; Piqueras et al., 2015).

To supplement the SR-FTIR images, atomic force microscopy infrared spectroscopy (AFM-IR) was used. AFM-IR combines atomic force microscopy (AFM) with pulsed IR laser (**Figure S1**) to obtain localized mid-IR spectra (3600-900 cm^{-1}) of regions as small as tens of nm in the horizontal plane and with a vertical resolution of ~0.1 nm (Dazzi and Prater, 2017). Such resolution surpasses the resolution of optical IR instruments making AFM-IR a suitable technique to study nanoscale properties of wood materials. Within the study of plants, AFM-IR has been used to analyze the composition of thylakoids (Janik et al., 2013) and of epicuticular wax (Farber et al., 2019), and to understand how the structure and composition of the *Populus nigra* cell wall affects water transport within the xylem (Pereira et al., 2018). Further, it has been used to identify the products of the reaction between the cell wall of *Pinus taeda* and a phenol-formaldehyde resin (Wang et al., 2005). However, to our knowledge, no one has yet studied the nanoscale compositional variations between the cell wall, the middle lamella, and the rays of SW and HW.

The objective of this work was to obtain a detailed overview of the HW formation process in larch (*Larch gmelinii* i.e. var *Japonica*) at the micro and nanoscale by combining SR-FTIR, AFM-IR, and advanced chemometric tools.

MATERIALS AND METHODS

Sample Preparation

For this imaging study, a sample of *Larix gmelinii var japonica* (Kurile larch) was taken at 1.3 m stem height. The tree was felled in October 2017 near Hørsholm, north of Copenhagen, Denmark. The wood was freeze dried to avoid possible artifacts from air drying. From an area including nine annual rings, nine tangential (D_{T1}-D_{T9}) and nine cross-sections (D_{X1}-D_{X9}) were cut without any embedding of the samples, in order to preserve the major content and distribution of extractives. Samples of nominally 10 µm thickness were obtained using a Leica microtome (Leica RM2255) (**Figure 1**). For transportation, the samples were placed on a glass slide and covered with a glass coverslip.

Synchrotron Infrared Imaging

All the tangential and cross-sections were imaged at the IR beamline MIRAS of the ALBA synchrotron (Cerdanyola del Vallés, Spain, proposal 2018022761). Before IR imaging, samples were inspected by light microscopy in order to select regions of interest (ROI) that included ray and tracheid cells and appeared to have a sample transparency that would allow SR-FTIR measurements in transmission mode. After area selection, the sample was carefully transferred onto a ZnSe disc of 1 mm thickness. The sample edges were fixed to the disc with tape to avoid sample movement during imaging. The IR measurements were acquired with a Bruker system (Hyperion 3000 microscope coupled to a Vertex 70 spectrometer) equipped with a liquid-nitrogen cooled mercury cadmium telluride (MCT) detector.

FIGURE 1 | Representation of the sampling procedure. Left plot: Scheme of a tree trunk. Middle plots: Transverse and tangential micro-sections of wood (Photo: A. Musson/Royal Botanic Garden, Kew). The marked rectangles show the regions of interest for the heartwood formation study. Right plots: Collection of IR Synchrotron images (40 µm x 40 µm) of tracheid and ray cells of larch. The numbered rectangles mark the chosen annual rings (1-9) from sapwood to heartwood including the transition zone in between.

Preliminary tests were carried out prior to the IR imaging. During these tests, possible effects of the high-energy IR beam on the wood material was studied by collecting punctual IR spectra at different exposure times. No adverse effects on the tissue or the spectra were observed for any of the exposure times tested. Therefore, we are confident that the energy of the IR beamline did not adversely influence our SR-FTIR measurements.

The IR images were acquired in transmission mode, using a 36x objective. The images were collected with a 3 x 3 μm^2 spatial resolution. All spectra were obtained in the infrared region (4000 −800 cm^{-1}) with 64 co-added scans. Absorbance representation was used throughout. A total of nine SR-FTIR images were acquired in both tangential and cross directions.

Atomic Force Microscopy Infrared Spectroscopy

AFM-IR exploits the photothermally induced resonance effect to detect the absorption of IR radiation with the AFM tip (Dazzi et al., 2005). In short, the sample is irradiated with the IR source and mechanically expands to dissipate the absorbed energy (**Figure S1**). The strongest expansion happens when the sample is irradiated with the IR wavelength that corresponds to the maximum absorption by the sample. Thus, by placing the AFM tip directly above the irradiated area, it is possible to detect the expansion of a sample by monitoring the deflection and oscillation of the AFM cantilever. This thermal expansion is directly proportional to the absorption coefficient of the excited area (Dazzi et al., 2012). By analyzing the many frequencies and amplitudes of the resonant oscillations of the AFM cantilever with Fourier transform, we extracted the useful information and reconstructed the IR spectra of various regions on the sample with a spatial resolution that is close to the size of an AFM tip.

We used a nanoIR from Analysis Instruments, Inc., to obtain mid-IR spectra of the ray, the middle lamella, and the secondary cell wall. Only two of the samples of the cross-sections (D_{X1} and D_{X8}) were used for the AFM-IR investigation. We fixed the edges of a sample with adhesive tape to a glass slide and acquired AFM images in contact mode. First, we found a suitable region where we could see both the middle lamella and the tracheid cell wall, or the ray and the adjacent tracheid cell wall and imaged it. Then, we collected and averaged three-background IR spectra with the resolution of 4 cm^{-1} to account for the variations in power of the IR source. For the IR spectral acquisition, we chose the second mode of the cantilever vibration to record the signal. This mode had a frequency of about 190 kHz and we chose the frequency window to be ±25 kHz to account for the variations in thermoelastic properties between the cell walls and the middle lamella or the ray. The second mode was chosen to improve the signal to noise ratio of the cantilever amplitude. For AFM-IR, it is crucial to position the IR laser directly at the tip-sample contact to make sure that the recorded spectrum originates directly from the area at the sample above where the tip is positioned. To do so, we scanned around the tip-sample contact area with the IR laser to find the highest cantilever amplitude, which corresponds to the position where the IR laser has maximum power. Once the tip, sample, and laser positions were optimized, we collected the spectra at various positions of the sample with the energy resolution of 4 cm^{-1} and by co-averaging 256 scans. After IR spectral acquisition, the same area was inspected by using the build-in camera. No laser damage was observed on any of the samples. The AFM images were flattened to remove the tilt and the AFM-IR spectra were smoothed by a Savitzky-Golay filter (2nd polynomial degree, 15 points window size) (Savitzky and Golay, 1964).

SR-FTIR DATA TREATMENT

The data treatment of SR-FTIR images consisted of two consecutive steps: (1) preprocessing of the image spectra to correct for scattering effects and (2) analysis by Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) (Tauler et al., 1995; de Juan and Tauler, 2006) to obtain pure spectra of the image constituents and their related distribution maps. The next subsections describe these steps.

Data Preprocessing

Due to the thickness and density of the samples, the infrared spectra of ray and tracheid cells were oversaturated in both the low and the high spectral wavelength range. Therefore, these spectral areas had to be excluded, and only the 1200-1750 cm^{-1} range was included in the analysis.

Infrared spectra are prone to artifacts because of Mie scattering associated with surface irregularities. Such artifacts may produce a broad oscillation in the baseline spectrum and can lead to distortions in both the position and intensity of absorption bands (Romeo et al., 2006). The raw data were corrected by the algorithm Asymmetric Least Squares (AsLS) (Eilers, 2004), which has been demonstrated to cope well with this type of scattering (Piqueras et al., 2013).

Hyperspectral Image Resolution

The goal of hyperspectral image resolution and, consequently, of the multivariate curve resolution alternating least squares (MCR-ALS) algorithm, is the decomposition of the original raw image data into distribution maps and pure spectra of the constituents present in the imaged sample (Tauler et al., 1995; Jaumot et al., 2005; de Juan and Tauler, 2006; Tauler et al., 2009). In the matrix form, hyperspectral images can be well described by a bilinear model based on the Beer-Lambert law [**Eq.(1)**], where the **D** matrix contains the original raw spectra, which are decomposed into a set of concentration profiles (**C** matrix) and corresponding pure spectra (**S^T** matrix) of the constituents present in the image. Every row of the **S^T** matrix corresponds to the pure spectrum of an image constituent, while every column of the **C** matrix of concentration profiles corresponds to the related pixel-to-pixel variation of its chemical concentration. It should be pointed out that each column of the **C** matrix can be refolded appropriately in order to recover the original two-dimensional spatial image structure and then pure distribution maps are obtained.

$$\mathbf{D} = \mathbf{C S}^T + \mathbf{E} \qquad (1)$$

MCR-ALS is a flexible method that allows analyzing a single image individually or several images simultaneously. To obtain a complete and reliable description of the HW formation, the simultaneous analysis of the recorded images along the nine-annual rings obtained from larch sample was performed. Thus, two multisets were built and analyzed separately—formed by nine and eight images collected in the cross-sectional (D_X) and tangential directions (D_T), respectively. See **Figure 2** for a visualization of the image multiset structure. In these multisets, the spectral dimension of all images is the same, while the image dimensions may differ between images because the images are unfolded before being merged into a single matrix (**Figure 2**). Due to extreme cases of over-saturation, the first image of the cross-sections (D_{T1}) and the last image of the tangential sections (D_{X9}) had to be excluded from the multiset structures D_T and D_X, i.e:

$$D_X = [D_{X1}; D_{X2}; D_{X3}; D_{X4}; D_{X5}; D_{X61}; D_{X62} D_{X7}; D_{X8}]$$

$$D_T = [D_{T2}; D_{T3}; D_{T4}; D_{T5}; D_{T6}; D_{T7}; D_{T8}; D_{T9}]$$

A multiset structure also follows the bilinear model based on Beer–Lambert law [see **Eq. (1)**]. In this example, image multiset analysis by MCR-ALS provides a single matrix S^T of pure spectra, identical for all the images analyzed, and a C matrix formed by as many submatrices as the number of images included in the data set. Every column of each C submatrix can be refolded conveniently to recover the distribution map of each constituent present in the different images of the data set (**Figure 2**).

MCR-ALS multiset analysis was performed on both multiset structures D_X and D_T following the MCR-ALS steps described in the literature (Jaumot et al., 2005). The first step consisted of determining the number of components involved during HW formation by singular value decomposition of the whole preprocessed D_X and D_T matrices (Golub and Reinsch, 1970). Five contributions were needed to describe the variation in both multisets. Then, initial estimates of pure spectra were obtained with a method based on SIMPLISMA (Windig and Guilment, 1991). The spectral estimates and the original multiset are used to perform an iterative alternating least squares optimization of matrices C and S^T under constraints. To obtain unmixed resolved profiles that are chemically meaningful, the constraints used in the resolution of both multiset structures were non-negativity in both, the concentration and the spectral profiles (Bro and de Jong, 1997), and normalization of pure spectra in the S^T matrix (using 2-norm, i.e., the Euclidean norm).

After a preliminary MCR-ALS analysis of both multisets (D_X and D_T) under non-negativity constraints, we realized by inspection of the distribution maps obtained that there were components absent in some sub-images of both multisets. As a consequence, a new MCR-ALS analysis was performed to obtain more accurate solutions by imposing the additional constraint of correspondence of species, which encodes information on the presence/absence of constituents in the concerned images of the multisets (Tauler et al., 2009), i.e. when a certain constituent is absent in one image, the related concentration profile is null. The lack of fit was 6.37% and 7.44% for D_X and D_T, respectively, which is satisfactory for FTIR measurements (Felten et al., 2015).

It is important to note that, when resolving images of biological samples, each resolved contribution (component) may refer to a mixture of chemical compounds of defined composition (polysaccharides, sugars, polymers, fatty acids, flavonoids…) that are present in a particular location of the sample and often represent a distinct biological element, e.g., a tissue or a cell part. This means that an MCR contribution is not necessarily a pure chemical compound. In imaging, the component discriminating ability of MCR will also depend on

FIGURE 2 | MCR-ALS analysis for a multiset structure of images, where x and y are spatial pixels and λ represents the wavenumbers of the spectra. Dx_n is the augmented data matrix; C is the concentration profiles; and S^T is the pure spectra matrix.
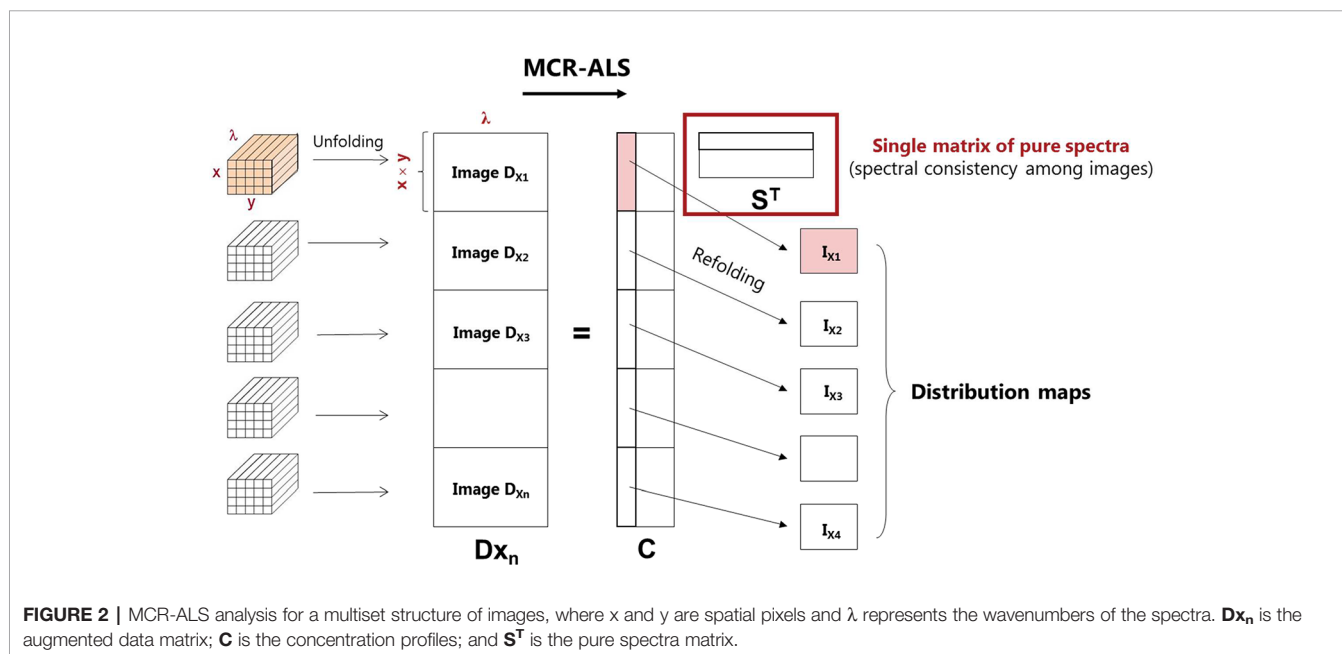
how fine the spatial resolution of the imaging technique used is, e.g., if two components are differently distributed at nanoscale level, MCR will not resolve them if imaging is performed at microscale level.

RESULTS

Exploratory Analysis of the SR-FTIR Images

In order to identify the main spectral variations during the HW formation of Kurile larch, an exploratory analysis was done. A small area of the ray and tracheid cell wall was selected for each of the images collected in the cross-sectional direction (**Figure 3A**). **Figure 3B** shows the average spectra of the ray area selected for each of the collected images (the average spectra of the tracheid cell wall area can be found in the supplementary information (**Figure S2**). The main spectral features when going from the SW to the HW of the ring area are the emergence of a new band at 1640 cm^{-1} and the decrease in intensity of the band at 1660 cm^{-1}. These spectral features appear in the sample D_{X5}, which we therefore determined to be the transition zone between SW and HW (**Figure 3C**). According to the literature, the band at 1640 cm^{-1} could correspond to adsorbed water (Popescu et al., 2007) or to a carbonyl stretching (C = O) as a consequence of the presence of para substituted ketone or aryl aldehydes (Lawther et al., 1996; Shi et al., 2012). Since the OH band at 3360 cm^{-1} appears not to co-vary with the band at 1640 cm^{-1}, we find it unlikely that the latter is related to water adsorption.

Resolution of Image Multiset Structures. MCR-ALS on Complete SR-FTIR Images

MCR-ALS multiset analysis on complete images provides the biological spectral signatures and distribution maps needed to integrally describe the ray and tracheid cells across the transition zone during the HW formation of larch. Although, as mentioned before, there is not necessarily a one-to-one correspondence between the MCR-ALS contribution and the individual chemical

FIGURE 3 | (A) Representation of the area selection of the lumen, cell wall, and ray. **(B)** Average spectra of the ray area selected for each of the cross-section images collected across the heartwood formation zone. **(C)** Zoom of the spectral range from 1200 cm^{-1} to 1800 cm^{-1} of the average spectra of the ray area selected for each of the cross-section images. The spectra gradually change from blue (sapwood) to red (heartwood) color.

FIGURE 4 | MCR-ALS image multiset results of the multiset structure formed by complete cross-section images. **(A)** Distribution maps of components involved in the heartwood formation of Kurile larch. Each line of maps represents the resolved maps of all constituents for a particular sample. Each column of maps represents the distribution map of a particular constituent in all samples analyzed. Distribution maps use a gradual color scale where yellow color refers to high concentration values and the blue color to low values. **(B)** Related pure spectra.

compounds, each contribution can be associated with some particular kind(s) of plant cell region, because of the morphology of the distribution maps and the spectroscopic features found in the resolved spectra. The resolution results for the multiset of the cross sections (D_X) are shown in **Figure 4**. The multiset analysis of the tangential section images (D_T) gave similar MCR-ALS results as D_X and are shown in **Figure S3**. The light microscopy images, corresponding to the imaged ray and the two surrounding tracheid cells areas are shown in the left of **Figure 4A**. The resolved distribution maps of each image are displayed in the right side of **Figure 4A** and the related resolved spectra in **Figure 4B**.
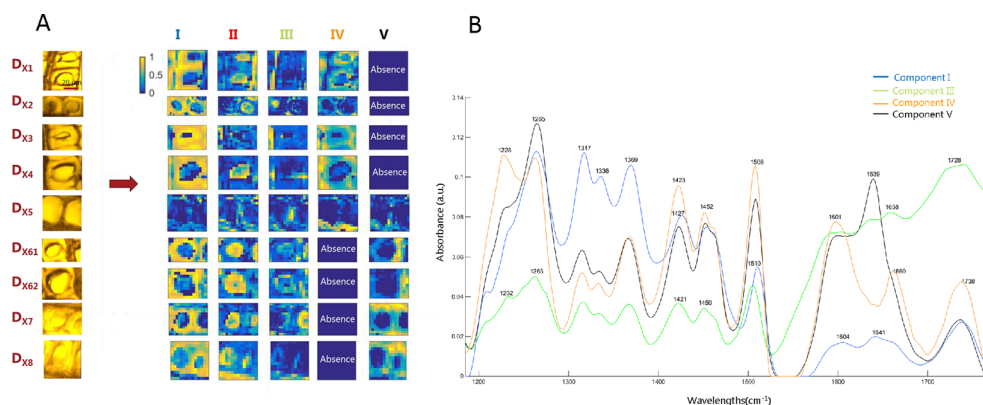
MCR-ALS was able to resolve the main plant cell constituents of the wooden tissue and to give further details on chemical changes occurring in the transition from SW to HW. Component I is characterized by signals in the region between 1317-1370 cm^{-1}, which are mainly associated with cellulose (Colom et al., 2003; Popescu et al., 2010) (see **Table 1**). The corresponding distribution maps show higher intensity in what we identify as the secondary cell wall (S2) of the tracheids, known to be thick in latewood and rich in cellulose (Fengel and Wegener, 1989). This component was distributed evenly across all the growth rings, as is also seen clearly in **Figure S3A** of the tangential sections.

The most prominent bands of lignin at 1505 and 1610 cm^{-1} are associated with C = C stretching of the aromatic ring modes (Colom et al., 2003; Weiland and Guyonnet, 2003; Popescu et al., 2010; Gorsás et al., 2011). They appear in components I, III, IV, and V, but show higher intensity and thus higher lignin contribution in components IV and V. Components IV and V appear to be situated in the same anatomical segments, i.e. in the cell corners (CC) and in the compound middle lamella (CML; middle lamella + adjacent primary walls), which is consistent with previous studies showing high lignin concentration in these

locations (Fengel and Wegener, 1989). The main spectral difference between component IV and V is the same spectral variation that was observed during the exploratory analysis of SR-FTIR images: Component IV shows the characteristic band at 1660 cm^{-1}, which is attributed to the ethylenic C = C (in coniferyl alcohol/sinapyl alcohol units) and C = O (in coniferaldehyde/ sinapaldehyde) bond stretches of lignin (Umesh and Rajai, 2010; Umesh et al., 2011; Bock and Gierlinger, 2019). This component was prevalent in the CC and CML of SW tracheid cells but disappears in the HW. The other feature observed in the preliminary analysis, the band at 1640 cm^{-1}, only appears in component V, which is distributed in the CC, CML, and part of the ray area of HW tracheids, as is also observed in **Figure S3A**. According to the literature, the IR band at 1640 cm^{-1} is assigned to a carbonyl stretching due to para substituted ketone or aryl aldehydes (Lawther et al., 1996; Shi et al., 2012). This band was also assigned to hydrogen bonding to the carbonyl group, as reported elsewhere (Pomar et al., 2002; Agarwal and Reiner, 2009; Bock and Gierlinger, 2019). Because component IV is present from D_{X1} to D_{X5} (SW region), and component V appears from D_{X5} to D_{X8} (HW region), it can be deduced that this fifth annual ring represents the transition zone, where the process of HW formation starts.

The resolved IR spectrum for component III shows a distinct, broad band at 1700-1736 cm^{-1} and is assumed to be formed by the overlapping of the C = O stretch vibration of acetyl or carboxylic acid (COOH) groups (Faix, 1991; Schwanninger et al., 2004) and the C = O stretch vibration of unconjugated ketones, carbonyls, and esters groups. This broader band centered at 1728 cm^{-1} could suggest the presence of resin acids in the rays since C = O stretching belonging to the -COOH group absorb around 1700 cm^{-1} (Traoré et al., 2018). Besides, bands are observed around 1600, 1637, and 1660 cm^{-1}; the same ones as found in components IV and V, but weaker. Candidates for para

TABLE 1 | The characteristic bands in FT-IR spectra of the studied samples and their assignments according to the literature data.

Wavenumber (cm⁻¹); literature	Band origin (assignment) with comments	Wavenumber (cm⁻¹) from SR-FTIR & AFM-IR; this work
1060–1015	C-O valence vibration mainly from C3-O3H (Cellulose)[a]; 1060cm⁻¹ polysaccharide[b]	1056
1072	C-O deformation in secondary alcohol of galactosyl subunits[c]	1072–1076
1095	C-C and C-O stretching motions (cellulose)[a]	1096
1108	COH in plane deformation (celluloses and hemicelluloses)[d]; Aromatic C-H in plane deformation (typical syringyl units)[e]; C = O stretch[e]	1108
1162-1125	C-O-C valence vibration (polysaccharide)[b]; 1165cm⁻¹ (characteristic for 5,7-dihydroxysubstituted flavonoids)[f]	1156–1164
1230-1221	C-C plus C-O plus C = O stretch[a]; lignin, Guaiacyl condensed > Guaiacyl etherified[g]	1220–1228
1235-1225	OH plane deformation, also COOH[a]	1232
1270-1266	Guaiacyl ring plus C = O stretch[a]	1263–1265
1280-1277	CH-deformation in cellulose[d]	1272–1284
1317-1315	CH₂ rocking vibration (cellulose)[b]	1315–1317
1335-1320	CH in plane bending in cellulose[d]	1320–1332
1365-1335	OH plane deformation vibration (cellulose)[a]	1336
1375-1365	CH bending in cellulose[d]	1368–1372
1430-1422	CH₂ scissoring in lignin and cellulose[d] CH bending; aromatic skeletal vibrations with C-H plane deformation in lignin[a]	1423–1432
1460	Asymmetric C-H bending from methoxy groups in lignins[a,h]; asymmetric C-H bending in CH₃ and CH₂ in pyran for hemicelluloses[h]	1452–1456
1520-1505	Aromatic skeletal vibrations in lignin[a]	1520–1505
1605-1593	Aromatic skeletal vibrations plus C = O stretching in lignin[a]	1595–1608
1640-1635	C = O stretching of para substituted ketone or aryl aldehydes[i]; C = O stretch of taxifolin[j]	1640–1635
1650-1640	H-bonded C = O stretching in coniferyl/sinapyl aldehyde[k]	1640–1648
1655-1660	Ring conj. C = C stretching of Coniferyl/sinapyl alcohol; C = O stretch of coniferyl/sinapyl aldehyde[k]	1660
1690-1670	C = O stretching in conjugated ketones[h]	1672–1692
1700-1690	C = O vibration in carboxylic group in resin acid[h]	1700–1690
1730-1725	C = O vibration of acetyl-or COOH-groups[a]	1728
1738-1709	C = O stretching in unconjugated ketones, carbonyls and ester groups (frequently of carbohydrate origin)[a,e]	1724–1740
1770-1760	C = O stretching in conjugated ketones[e]	1752–1764

[a]Schwanninger et al., 2004; [b]Fackler et al., 2010; [c]Ghosh et al., 2015; [d]Popescu et al., 2010; [e]Popescu et al., 2007; [f]Zu et al., 2012; [g]Faix, 1991; [h]Traoré et al., 2018; [i]Lawther et al., 1996; [j]Kocábová et al., 2016; [k]Bock and Gierlinger, 2019.

substituted ketones are flavonoids such as taxifolin and dihydrokaempferol (Ruddick and Xie, 1994; Kocábová et al., 2016), known to be abundant in larch (Nisula, 2018). This component is represented in all the images in the ray, as well as tracheid lumina and S3 layer.

As mentioned before, resolved IR spectra reflect a mixture of different kinds of biomolecules. For example, the IR spectrum related to component IV consists of a mixture of mainly lignin and hemicelluloses, known by the presence of the band at 1738 cm⁻¹, which is attributed to ester carbonyl groups prominent in hemicelluloses (Stewart et al., 1995; Gorsás et al., 2011) (see **Table 1** for the different IR bands of wood assembled from the literature). Finally, since the samples were measured in the dry state for the SR-FTIR experiment, the IR spectrum of component II associated with part of the lumen is not shown, since it does not contain any biologically relevant information. The spectrum corresponds to the IR absorbance of the ZnSe slide used in the measurements.

AFM-IR Spectra Analysis

The AFM-IR spectra collected in the tracheid cell in the cross-sectional sections of SW (D_{X1}) and HW (D_{X8}) are shown in **Figure 5**, together with the AFM images. Where possible, we acquired the AFM image of the middle lamella and cell wall in a single image (**Figure 5A**). However, in some places, the middle lamella (and the ray) contain topographical features that exceed a few micrometers. Such complex topography is difficult to image in contact mode because of the limited vertical range of the AFM scanner. In addition, the imaging of such complex topographical features blunts the AFM tip rapidly. Hence, we took images of various dimensions, depending on the topography of the area, in order to minimize the mechanical strain exerted on the AFM tip, but to still be able to get a view of both the cell wall and the middle lamella in the same image (**Figure 5B**). Numbered crosses on the AFM images indicate the location of AFM-IR spectra. In contrast to the SR-FTIR spectra, the spectral region between 1000-1200 cm⁻¹ is not saturated in the AFM-IR measurements, so we were able to obtain information from that region as well. Although the IR spectra of CML (point nr. 2) and S2 layer (point nr. 1) are very similar to each other in the SW (**Figure 5A**), we can observe different intensities around 1504-1510 cm⁻¹, associated with C = C stretching of the aromatic ring modes of lignin. This band is more intense in the CML, which was also seen in the SR-FTIR data and is consistent with literature (Fengel and Wegener, 1989). The C = O stretching mode is slightly shifted to lower wavelengths (1724 cm⁻¹) in point nr. 3, which is characteristic of pectins and/or hemicelluloses (Gorsás et al., 2011). Cellulose bands dominate the IR spectra in the S2 layer (see **Table 1**). In the case of the HW tracheids (**Figure 5B**), no spectral differences between the CML (point nr. 2) and S2 layer (point nr. 1) were observed. By comparing AFM-IR spectra of **Figures 5A** (SW) and **B** (HW), we see the presence of the band at 1648 cm⁻¹, while the band at 1660 cm⁻¹ is missing in the HW tracheid cell wall; the opposite is the case for SW. These are the same main spectroscopic features that were found in the analysis of SR-FTIR images. Finally, a new band at 1108 cm⁻¹ appears in HW (**Figure 5B**), linked with COH in plane deformation of celluloses and hemicelluloses (Schwanninger et al., 2004) and/or with aromatic C-H in plane deformation of lignin.

Figure 6 shows the AFM images and AFM-IR spectra of the ray region and the S2 cell wall of an adjacent tracheid in SW (**Figure 6A**) and HW (**Figure 6B**). Imaging of the ray and the cell wall in contact mode was particularly difficult because of their different material properties. This is why the image is blurry and
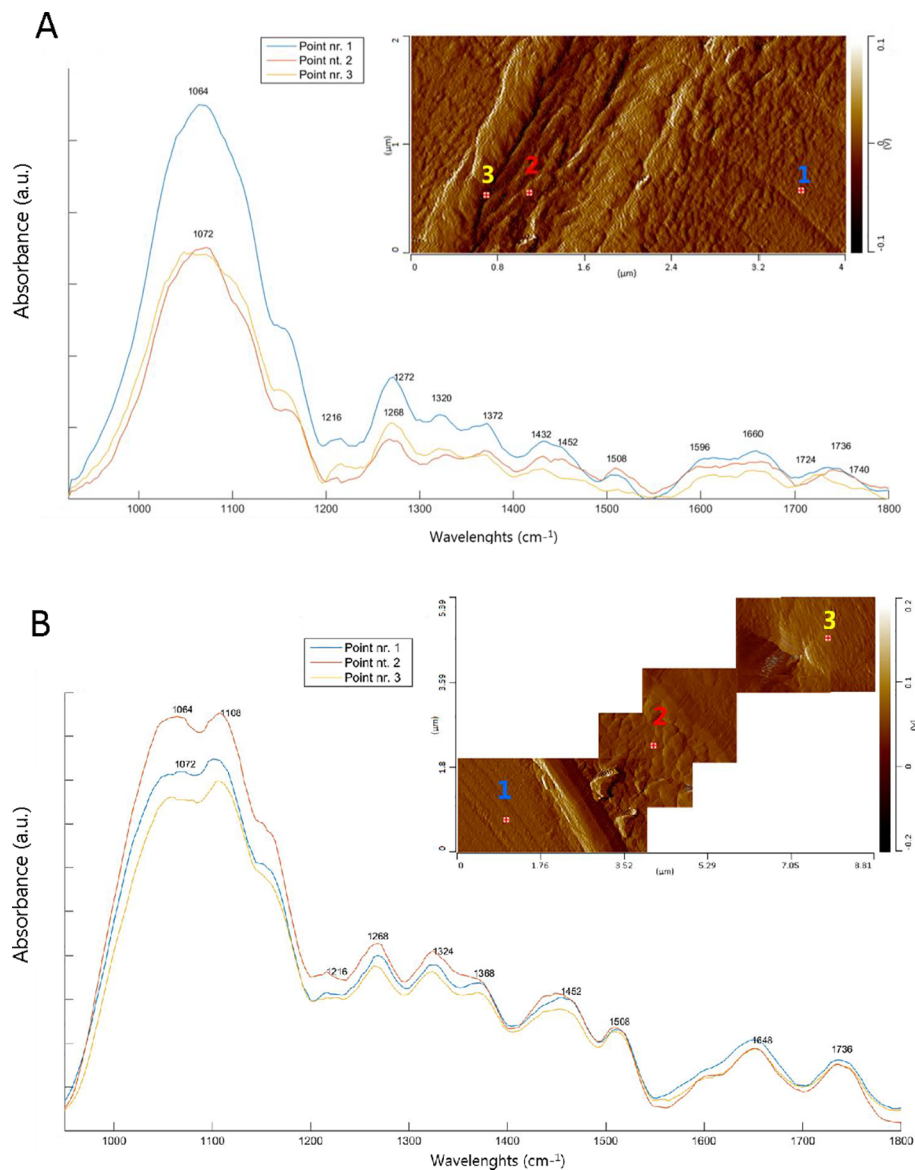
FIGURE 5 | AFM deflection images and AFM-IR spectra of the middle lamella and secondary cell wall. The numbered crosses on the AFM images indicate the location of the AFM-IR spectra. Point nr.1: Secondary cell wall (S2); point nr.2: Compound middle lamella (CML) and point nr.3: Assumable primary cell wall. **(A)** sapwood tracheid cell (D_{x1}) and **(B)** heartwood tracheid cell (D_{x8}).

it appears as if the ray is smeared over the cell wall (**Figure 6**). Such topography and relationship between the ray and the cell wall are unrealistic and simply an artefact of imaging in contact mode. This artefact does not affect the acquisition of AFM-IR spectra or its spectral features because the spectra are acquired after imaging was finished and at the frequency characteristic for the tip-substrate system. Lignin and cellulose bands are more intense in the S2 cell wall (points nr. 3, 8, and 9) of SW compared to the ray area (points nr. 1, 2, 4, 5, 6, and 7). A characteristic peak occurs at 1076 cm^{-1} inside the ray, assigned to C-O bands in

primary and secondary alcoholic groups (Compounds and Hergert, 1960). A low intensity of the band at 1728 cm^{-1} is seen in the S2 cell wall. The ray region spectrum of HW (points nr.1 and 2) (**Figure 6B**) reveals characteristics peaks at 1692, 1756, and 1764 cm^{-1}, which represent the C = O stretching in conjugated ketones (Popescu et al., 2007) and carboxylic acid groups (Schwanninger et al., 2004), alkyl esters (including the methyl ester of fatty acids), and in γ-lactone (Lievens et al., 2011), respectively. The band at 1164 cm^{-1} is typical for 5,7- dihydroxy-substituted flavonoids (Zu et al., 2012) and indicates the presence
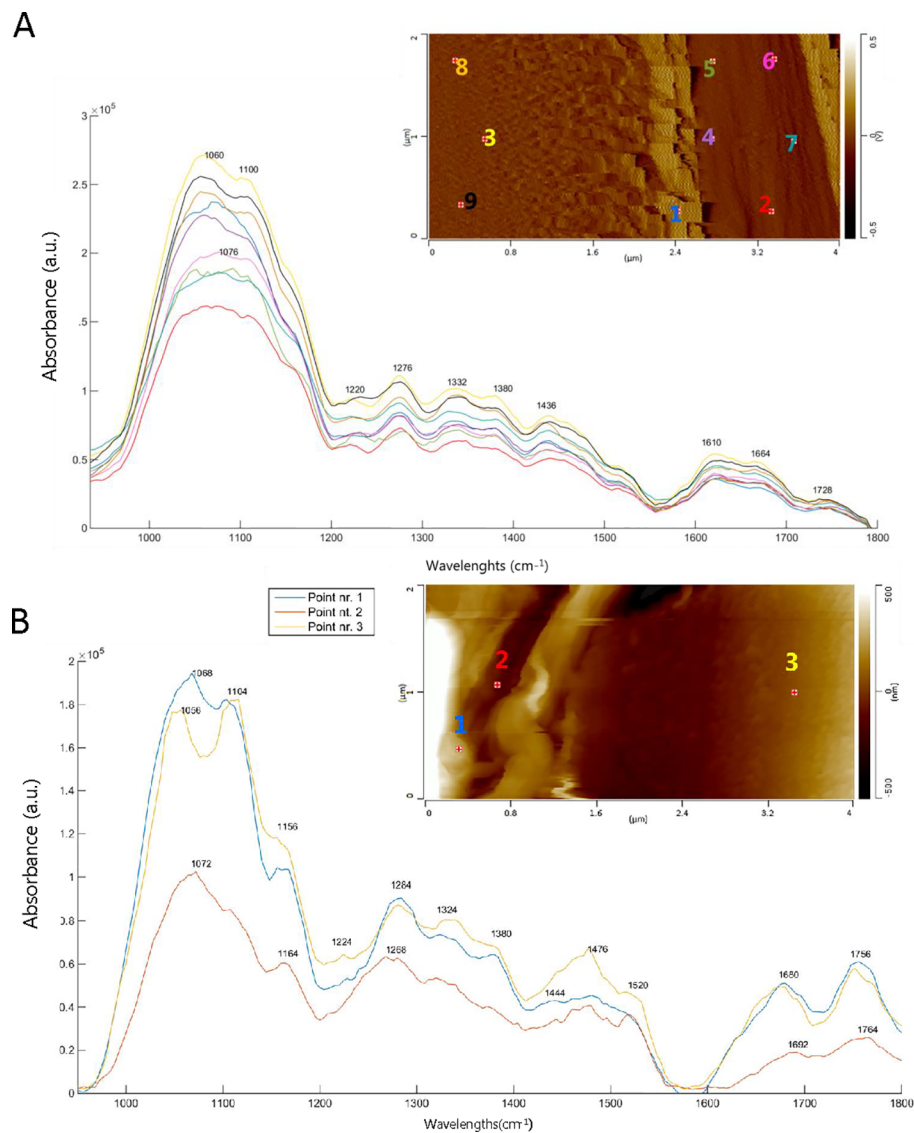
FIGURE 6 | AFM deflection (a) and height (b) images and AFM-IR spectra of the ray cell and secondary cell wall. The numbers on the AFM images indicate the location of the AFM-IR spectra. **(A)** sapwood tracheid and ray region (D_{x1}) where points nr.: 1, 2, 4, 5, 6, 7 correspond to the ray and points nr.: 3, 8, 9 to the secondary layer (S2) of tracheid cell wall. **(B)** heartwood tracheid and ray region (D_{x8}) where points nr.: 1, 2 are the analyses of the ray and point nr. 3 of the secondary layer (S2) of tracheid cell wall.

of taxifolin. It is also important to highlight the presence of the band at 1072 cm^{-1}, associated with C-O deformation in primary and secondary alcohol groups of galactosyl subunits (Ghosh et al., 2015). In the S2 layer (point nr. 3), the AFM-IR spectrum shows more intense cellulose/hemicellulose bands, whereas the band at 1660 cm^{-1} is shifted to higher wavenumbers compared to the S2 layer in SW, likely due to the formation of C = O conjugated ketones. Nevertheless, the lignin band is again more intense and a shoulder at 1640-1660 cm^{-1} appears.

DISCUSSION

The formation of HW is linked to the occurrence of non-structural substances called extractives, which play an important role in the resistance of wood to fungal decay (Hinterstoisser et al., 2000; Schultz and Nicholas, 2000). By combining high resolution SR-FTIR with the powerful unmixing algorithm MCR-ALS, we were able to identify a component associated mainly with phenolic compounds and likely with deposition of resin acids (component III from D_X and

D_T, **Figures 4** and **S2**). Component III shows prominent IR bands at 1637, 1658, and 1728 cm^{-1} in the ray, tracheid lumen, and S3 layer. The band around 1637 cm^{-1} was assigned to the ketone bond in taxifolin (Ruddick and Xie, 1994; Miklečić et al., 2012; Kocábová et al., 2016; Liu et al., 2018), one of the most abundant phenolic compounds in larch wood (Giwa and Swan, 1975; Babkin et al., 2001; Ivanova et al., 2012; Zule et al., 2017) (see reference spectrum of taxifolin in **Figure S4** for comparison). The other peaks seen in the taxifolin spectrum are not seen due to overlap with the spectra of the structural wood cell wall biopolymers. The distinct, broad band 1700-1736 cm^{-1} centered at 1728 cm^{-1}, is assigned to the C = O vibration of carboxylic acid groups in resin acids (Faix, 1991; Schwanninger et al., 2004). It is important to emphasize that component III appears to be more common in the ray than in tracheid cells in the SW and vice versa in the HW (see **Figures 4B** and **S3**). As described by Hillis (1987), parenchyma cells die when HW forms and the polyphenols diffuse into cell walls. The relatively higher concentration of taxifolin and resin acid mixtures in the HW cell walls likely explains its natural durability.

With the MCR-ALS multiset analysis of tangential and cross-sections, two MCR-ALS contributions were found to be directly linked to the process of HW formation in Kurile larch (components IV and V, **Figures 4** and **S3**). Both of them mainly showed lignin bands but had different spatio-temporal distributions, as well as different spectral features. Component IV was located in the CC and CML of tracheid cells in the SW, although in the transition zone it was almost exclusively present in the ray area. The other lignin contribution (Component V) was distributed in CC, CML, and in the ray cells of HW. The main spectroscopic difference between component IV and V was the appearance of the band at 1640 cm^{-1} and the simultaneous disappearance of the band at 1660 cm^{-1}. The disappearance of the 1660 cm^{-1} band might be explained by a decrease in the intensity or as a drastic shift in the frequency to 1640 cm^{-1}.

The decrease of the intensity of band at 1655-1660 cm^{-1} could be explained in terms of condensation reactions (Yamauchi et al., 2005) of coniferyl alcohol in lignin. By condensation, coniferyl alcohol loses the ethylenic bond which then no longer contributes to the vibration at 1660 cm^{-1}. However, this process cannot explain the appearance of the 1640 cm^{-1} band. Possible coupling of the condensation reaction with the integration of new coniferyl aldehyde moieties into the lignin structure and their H-bonding to unreacted alcohols would decrease the frequency of the 1660 cm^{-1} band to 1640 cm^{-1}, as described by Bock and Gierlinger (2019) and Agarwal and Reiner (2009). Another possibility for the band shift is the oxidation of coniferyl alcohol to its aldehyde and subsequent H-bonding to the carbonyl groups during the aging of the wood cells. Lastly, the appearance of taxifolin and other flavonoids may explain the appearance of the 1640 cm^{-1} band, as they contain carbonyl vibrations at 1640 cm^{-1}. Since flavonoids are rich in OH-groups, they are likely to interact with lignin and cause the shift of the 1660 cm^{-1} band frequency.

From the MCR-ALS distribution maps, interpretation of the deposition pathways of the components involved during HW formation of larch was achieved. It is interesting to observe that distribution of extractives (component III from D_X and D_T, see **Figures 4A** and **S3A**) is represented in all the images and can be found in the ray region, lumen, and S3 layer of tracheid cells. It may suggest that precursor molecules are present before and after the actual transition from SW to HW. This pattern is observed in Juglans-Type II HW formation. Not much is known about this mechanism, but it has been described for deciduous trees, as well as conifers, such as *Prunus*, *Platycarya*, *Eucalyptus*, and *Pseudotsuga* (Kampe and Magel, 2013). If we follow their deposition in the distribution maps of **Figures 4A and S3A**, it seems that extractives are accumulated in the ray in SW and after ray cell death, they spread to the surrounding wood tissues. We can also see how component V emerges and component IV diminishes during the HW formation process. The similar IR spectra and spatial deposition indicate that component IV corresponds to a set of precursor molecules of component V. Thus, Type II seems the reasonable mechanism of larch HW formation.

AFM-IR spectroscopy revealed variations in the composition between the tracheid and the ray regions in SW and HW (**Figures 5A, B**). The main spectral differences were found at 1108, 1456, 1648, and 1660 cm^{-1}. In the HW tracheid cell walls, the band at 1108 cm^{-1}, assigned to COH in plane deformation of celluloses and hemicelluloses (Schwanninger et al., 2004) and/or to aromatic C-H in plane deformation of lignin, increases. At the same time, the band at 1456 cm^{-1}, related to C-H bending of methoxyl groups (Schwanninger et al., 2004) becomes broader. According to the literature, the IR band at 1108 cm^{-1} could appear because of cross-linking reactions of –OH groups of cellulose/hemicellulose with phenolic compounds at the cell wall level (Wang et al., 2016), resulting in an increase in the hardness of the wood cell walls. The most important difference is again the appearance of the band at 1648 cm^{-1} and the absence of the band at 1660 cm^{-1} in the HW tracheid cell walls. As mentioned earlier, the band at 1655 cm^{-1} could be mainly present because of C = O and C = C groups in coniferyl aldehyde and coniferyl alcohol structures of lignin. Hence, it is likely that the reduction of band intensity may be attributable mainly to condensation reactions of lignin molecules or oxidative alteration of lignin.

It is well known that lignin condensation makes lignocellulosic biomass more recalcitrant, mainly due to limiting the accessibility to the polysaccharides in the cell wall (Li et al., 2016). Additionally, generation of new carbonyl groups by oxidation, as discussed above, may increase non-productive binding of cellulases or enzyme inhibition *via* chelation of metal co-factors (Li et al., 2016), thereby starving the fungus. The attachment of other, potentially fungitoxic or antioxidant phenolics, i.e. flavonoids, to these newly formed reaction sites are other possibilities that lignin modification would allow for. These steric or chemical inhibitory effects are important for understanding the durability of heartwood (Valette et al., 2017).

The AFM-IR spectra of the ray also show differences between SW and HW (**Figures 6A, B**).

For example, the appearance of the band around 1680-1692 cm^{-1} suggests the presence of conjugated ketones and carboxylic acids,

such as resin acids, in the HW ray and its surrounding cell wall tracheid. The presence of taxifolin inside the ray is supported by the presence of the band at 1164 cm^{-1} (a characteristic band for 5,7-dihydroxysubstituted flavonoids) (Zu et al., 2012).

We can also observe an intense C = O stretching vibration at 1764 cm^{-1} and around 1750 cm^{-1} (see **Figure 6B**), indicating the presence of γ-lactones and alkyl esters, including the methyl esters of fatty acids (Lievens et al., 2011). Prominent bands that appear at 1072 and 1728 cm^{-1} are assigned to C-O deformation in primary and secondary alcohol groups of galactosyl- and carbonyl of carboxylic groups.

The interpretation of the spectra is complex since mixtures of chemical compounds are present throughout the plant tissue, i.e spectra of pure chemical constituents are rarely possible to obtain, neither by use of high spatial resolution (as in AFM-IR), nor by use of MCR-ALS modelling. Consequently, comparing to a spectral data base with the most common components present in plant cell walls might help to further identify the chemical compounds, but it would most likely not lead to conclusive results. AFM-IR is presented in this study as a potent technique to further characterize plant cell wall components because of its higher spatial resolution than SR-IR imaging, and because spectra could be obtained for a broader range of wavenumbers. However, measurements of a more comprehensive sample set would be necessary in order to not simply illustrate the technique but obtain representative results.

CONCLUSIONS

MCR-ALS multiset analysis on sets of SR-FTIR images collected across the HW formation zone of Kurile larch provided a cellular level description of the components involved in HW formation. In particular, the IR resolved spectral signatures and comparison with IR reference spectra from the literature allowed us to identify taxifolin, one of the most abundant extractive in larch, in rays as well as in the lumen and S3 cell wall layer of adjacent tracheids. Moreover, refolding of the concentration profiles to the original image formats allowed us to see that one initial phenolic lignan contribution (component IV) was present in the SW, while a second somewhat similar contribution (component V) emerged in the transition zone and continued in the HW. Our interpretation of this result is that component IV is a set of precursor molecules for component V. Such a pattern is characteristic for Type II heartwood formation, also called Juglans-type. The main spectroscopic difference between component IV and V was the appearance of the band at 1640 cm^{-1} and the simultaneous disappearance of the band at 1660 cm^{-1}. We hypothesize that the disappearance of the 1660 cm^{-1} band may be attributable mainly to condensation reactions of lignin/lignan molecules or oxidative alteration of lignin. Lignin condensation reactions are known to make lignin more recalcitrant. Generation of new carbonyl groups by oxidation of coniferyl alcohol to coniferyl aldehyde could also help explain both the peak shift and the resistance against fungal attack of Kurile larch HW.

AFM-IR has been proven to be a powerful technique to study the nanoscale compositional variations between the cell wall, CML, and the ray of SW and HW of larch. The AFM-IR results confirmed the trends observed in the SR-FTIR image analysis and provided more detail about the plant cell wall composition as spectra were obtained for a broader spectral range. Conjugated ketones and carboxylic acids accompanied with the presence of γ-lactone and alkyl ester were also found in the HW rays. Finally, AFM-IR spectra proposed the existence of cross-linked reactions of cellulose/hemicelluloses with phenol compounds at the cell wall level in HW.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

All the authors discussed the results and contributed to the final manuscript. SP, SF, RO, AG-S, TK, AJ and LT carried out the SR-FTIR experiment and SJ performed the AFM-IR measurements.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019. 01701/full#supplementary-material

FIGURE S1 | Schematic diagram of AFM-IR setup. IR laser irradiates the sample which then expands. The expansion of the sample is detected by the AFM tip which deflects the whole AFM cantilever. The deflection is then monitored by tracking the movements of the AFM laser shone on the cantilever.

FIGURE S2 | (A) Representation of the area selection of the lumen, cell wall, and ray. **(B)** Average spectra of the cell wall selected for each of the cross section images collected across the heartwood formation zone. **(C)** Zoom of the spectral range 1200 cm^{-1} to 1800 cm^{-1} of the average spectra of the cell wall selected for

each of the cross section. The spectra gradually change from blue (sapwood) to red (heartwood) color.

FIGURE S3 | MCR-ALS results of the multiset structure formed by a series of tangential section images. **(A)** distribution maps of components involved in the heartwood formation of Kurile larch. Each line of maps represents the resolved

maps of all constituents for a particular sample. Each column of maps represents the distribution map of a particular chemical constituent in all samples analyzed. Distribution maps use a gradual color scale where yellow color refers to large concentration values and blue color to small values. **(B)** Related pure spectra.

FIGURE S4 | FTIR spectrum of taxifolin crystal according to (Liu et al., 2018).

REFERENCES

Agarwal, U. P., and Reiner, R. S. (2009). Near-IR Surface-Enhanced Raman Spectrum of Lignin † A. *J. Raman Spectrosc.* 40, 1527–1534. doi: 10.1002/Jrs.2294

Babkin, V. A., Ostroukhova, L. A., Malkov, Y. A., Babkin, D. V., Onuchina, N. A., and Ivanova, S. Z. (2001). Isolation of biologically active compounds from larch wood. *Int. Symp. Wood Pulping Chem.* 2, 119–122.

Belt, T., Keplinger, T., Hänninen, T., and Rautkari, L. (2017). Cellular level distributions of Scots pine heartwood and knot heartwood extractives revealed by Raman spectroscopy imaging. Nice, France. *Ind. Crops Prod.* 108, 327–335. doi: 10.1016/j.indcrop.2017.06.056

Bergström, B., Gustafsson, G., Gref, R., and Ericsson, A. (1999). Seasonal changes of pinosylvin distribution in the sapwood/heartwood boundary of Pinus sylvestris. *Trees - Struct. Funct.* 14, 65–71. doi: 10.1007/s004680050210

Bito, N., Nakada, R., Fukatsu, E., Matsushita, Y., Fukushima, K., and Imai, T. (2011). Clonal variation in heartwood norlignans of Cryptomeria japonica: Evidence for independent control of agatharesinol and sequirin C biosynthesis. *Ann. For. Sci.* 68, 1049–1056. doi: 10.1007/s13595-011-0118-7

Bock, P., and Gierlinger, N. (2019). Infrared and Raman spectra of lignin substructures: Coniferyl alcohol, abietin, and coniferyl aldehyde. *Raman Spectrosc.* 50, 778–792. doi: 10.1002/jrs.5588

Bro, R., and de Jong, S. (1997). A fast non-negativity-constrained least squares algorithm. *J. Chemom.* 11, 393–401.

Burčová, Z., Kreps, F., Greifová, M., Jablonský, M., Ház, A., Schmidt, Š., et al. (2018). Antibacterial and antifungal activity of phytosterols and methyl dehydroabietate of Norway spruce bark extracts. *J. Biotechnol.* 282, 18–24. doi: 10.1016/J.JBIOTEC.2018.06.340

Burtin, P., Jay-Allemand, C., Charpentier, J. P., and Janin, G. (1998). Natural wood colouring process in Juglans sp. (J. nigra, J. regia and hybrid J. nigra 23 x J. regia) depends on native phenolic compounds accumulated in the transition zone between sapwood and heartwood. *Trees - Struct. Funct.* 12, 258–264. doi: 10.1007/PL00009716

Bush, D., McCarthy, K., and Meder, R. (2011). Genetic variation of natural durability traits in Eucalyptus cladocalyx (sugar gum). *Ann. For. Sci.* 68, 1057–1066. doi: 10.1007/s13595-011-0121-z

Cao, G., Sofic, E., and Prior, R. L. (1997). Antioxidant and Prooxidant Behavior of Flavonoids: Structure-Activity Relationships. *Free Radic. Biol. Med.* 22, 749–760. doi: 10.1016/S0891-5849(96)00351-6

Colom, X., Carrillo, F., Nogue, F., and Garriga, P. (2003). Structural analysis of photodegraded wood by means of FTIR spectroscopy. *Polym. Degrad. Stab.* 80, 543–549. doi: 10.1016/S0141-3910(03)00051-X

Compounds, M., and Hergert, H. L. (1960). Infrared Spectra of Lignin and Related Compounds. II. Conifer Lignin and. *J. Org. Chem.* 25, 405–413. doi: 10.1021/jo01073a026

Déjardin, A., Laurans, F., Arnaud, D., Breton, C., Pilate, G., and Leplé, J.-C. (2010). Wood formation in Angiosperms. *C. R. Biol.* 333, 325–334. doi: 10.1016/J.CRVI.2010.01.010

Dazzi, A., and Prater, C. B. (2017). AFM-IR: Technology and applications in nanoscale infrared spectroscopy and chemical imaging. *Chem. Rev.* 117, 5146–5173. doi: 10.1021/acs.chemrev.6b00448

Dazzi, A., Prazeres, R., and Ortega, J. M. (2005). Local infrared microspectroscopy with subwavelength spatial resolution with an atomic force microscope tip used as a photothermal sensor. *Opt. Lett.* 30, 2388–2390. doi: 10.1364/OL.30.002388

Dazzi, A., Prater, C. B., Hu, Q., Chase, D. B., Rabolt, J. F., and Marcott, C. (2012). AFM-IR: Combining atomic force microscopy and infrared spectroscopy for nanoscale chemical characterization. *Appl. Spectrosc.* 66, 1365–1384. doi: 10.1366/12-06804

de Juan, A., and Tauler, R. (2006). Multivariate Curve Resolution (MCR) from 2000: progress in concepts and applications. *Crit. Rev. Anal. Chem.* 36, 163–176. doi: 10.1080/10408340600970005

de Juan, A., Tauler, R., Dyson, R., Marcolli, C., Rault, M., and Maeder, M. (2004). Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis. *TrAC Trends Anal. Chem.* 23, 70–79. doi: 10.1016/S0165-9936(04)00101-3

Dellus, V., Scalbert, A., and Janin, G. (1997). Polyphenols and colour of Douglas fir heartwood. *Holzforschung* 51, 291–295. doi: 10.1515/hfsg.1997.51.4.291

Eilers, P. H. C. (2004). Parametric time warping. *Anal. Chem.* 76, 404–411. doi: 10.1021/ac034800e

Fackler, K., Stevanic, S. J., Ters, T., Hinterstoisser, B., Schwanninger, M., Salmén, L., et al. (2010). Localisation and characterisation of incipient brown-rot decay within spruce wood cell walls using FT-IR imaging microscopy. *Enzyme Microb. Technol.* 47 (6), 257–267. doi: 10.1016/j.enzmictec.2010.07.009

Faix, O. (1991). Classification of Lignins from different botanical origins by FT-IR spectroscopy. *Holzforschung* 45, 21–28. doi: 10.1515/hfsg.1991.45.s1.21

Farber, C., Wang, R., Chemelewski, R., Mullet, J., and Kurouski, D. (2019). Nanoscale structural organization of plant epicuticular wax probed by atomic force microscope infrared spectroscopy. *Anal. Chem.* 91, 2472–2479. doi: 10.1021/acs.analchem.8b05294

Felhofer, M., Prats-Mateu, B., Bock, P., and Gierlinger, N. (2018). Antifungal stilbene impregnation: transport and distribution on the micron-level. *Tree Physiol.* 38, 1526–1537. doi: 10.1093/treephys/tpy073

Felten, J., Hall, H., Jaumot, J., Tauler, R., de Juan, A., and Gorzsás, A. (2015). Vibrational spectroscopic image analysis of biological material using multivariate curve resolution-alternating least squares (MCR-ALS). *Nat. Protoc.* 10, 217–240. doi: 10.1038/nprot.2015.008

Fengel, D., and Wegener, G. (1989). *Wood: Chemistry, Ultrastructure,Reactions* (Berlin-New York: Walter de Gryter).

Ghosh, T., Basu, A., Adhikari, D., and Roy, D. (2015). Antioxidant activity and structural features of Cinnamomum zeylanicum. *3 Biotech* 5, 939–947. doi: 10.1007/s13205-015-0296-3

Gierlinger, N., Schwanninger, M., Hinterstoisser, B., and Wimmer, R. (2002). Rapid determination of heartwood extractives in Larix sp. by means of Fourier transform near infrared spectroscopy. *J. Near Infrared Spectrosc.* 10, 203–214. doi: 10.1255/jnirs.336

Gierlinger, N., Jacques, D., Grabner, M., Wimmer, R., Schwanninger, M., Rozenberg, P., et al. (2004). Colour of larch heartwood and relationships to extractives and brown-rot decay resistance. *Trees - Struct. Funct.* 18, 102–108. doi: 10.1007/s00468-003-0290-y

Giwa, S., and Swan, E. (1975). Heartwood extractives of a western larch tree (Larix occidentalis Nutt.). *Wood Fiber* 7, 216–221.

Golub, G. H., and Reinsch, C. (1970). Singular value decomposition and least squares solutions. *Numer. Math.* 14, 403–420. doi: 10.1007/BF02163027

Gorsás, A., Stenlund, H., Persson, P., and Trygg, J. (2011). Cell-specific chemotyping and multivariate imaging by combined FT-IR microspectroscopy and orthogonal projections to latent structures (OPLS) analysis reveals the chemical landscape of secondary xylem. *Plant J.* 66, 903–914. doi: 10.1111/j.1365-313X.2011.04542.x

Higuchi, T. (1997). *Biochemistry and Molecular Biology of Wood*. Ed. T. Higuchi, Springer Series in Wood Science (Berlin, Heidelberg: Springer).

Hillis, W. (1985). "Ocurrence of extractives in wood tissue," in *Biosynthesis and biodegradation of wood components*. Ed. T. Higuchi (Orlando: Academic Press, Inc.) 209–228.

Hillis, W. (1987). *Heartwood and Tree Exudates*, Springer Series in Wood Science (Berlin: Springer-Verlag).

Hinterstoisser, B., Stefke, B., and Schwanninger, M. (2000). Wood: raw material-materiak-source of energy for the future. *Lignovisionen* 2, 29–36. doi: 10.1016/S0924-2031(99)00063-6

Imai, T., Tanabe, K., Kato, T., and Fukushima, K. (2005). Localization of ferruginol, a diterpene phenol, in Cryptomeria japonica heartwood by time-

of-flight secondary ion mass spectrometry. *Planta* 221, 549–556. doi: 10.1007/s00425-004-1476-2

Ivanova, S. Z., Gorshlov, A. G., Kuzmin, A. V., Gordienko, I. I., and Babkin, V. A. (2012). Phenolic compounds of Siberian and Dahurian larch phloem. *Russ. J. Bioorg. Chem.* 38, 769–774. doi: 10.1134/S1068162012070096

Janik, E., Bednarska, J., Zubik, M., Puzio, M., Luchowski, R., Grudzinski, W., et al. (2013). Molecular architecture of plant thylakoids under physiological and light stress conditions: a study of lipid-light-harvesting complex II model membranes. *Plant Cell* 25, 2155–2170. doi: 10.1105/tpc.113.113076

Jaumot, J., Gargallo, R., de Juan, A., and Tauler, R. (2005). A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometr. Intell. Lab. Syst.* 76, 101–110. doi: 10.1016/j.chemolab.2004.12.007

Kampe, A., and Magel, E. (2013). "New insights into heartwood and heartwood formation," in *Cellular Aspects of Wood Formation*. Ed. J. Fromm (Berlin: Springer), 71–95.

Keeling, C. I., and Bohlmann, J. (2006). Diterpene resin acids in conifers. *Phytochemistry* 67, 2415–2423. doi: 10.1016/J.PHYTOCHEM.2006.08.019

Kocábová, J., Fiedler, J., Degano, I., and Sokolová, R. (2016). Electrochimica Acta Oxidation mechanism of fl avanone taxifolin. Electrochemical and spectroelectrochemical investigation. *Electrochim. Acta* 187, 358–363. doi: 10.1016/j.electacta.2015.11.077

Kwon, M., Davin, L. B., and Lewis, N. G. (2001). *In situ* hybridization and immunolocalization of lignan reductases in woody tissues: implications for heartwood formation and other forms of vascular tissue preservation. *Phytochemistry* 57, 899–914. doi: 10.1016/S0031-9422(01)00108-X

Lawther, J. M., Sun, R., and Banks, W. B. (1996). Fractional Characterization of Wheat Straw Lignin Components by Alkaline Nitrobenzene Oxidation and FT-IR Spectroscopy. *Appl. Spectrosc.* 44, 1241–1247. doi: 10.1021/jf9502764

Li, M., Pu, Y., and Ragauskas, A. J. (2016). Current understanding of the correlation of lignin structure with biomass recalcitrance. *Front. Chem.* 4, 1–8. doi: 10.3389/fchem.2016.00045

Lievens, C., Mourant, D., He, M., Gunawan, R., Li, X., and Li, C. Z. (2011). An FT-IR spectroscopic study of carbonyl functionalities in bio-oils. *Fuel* 90, 3417–3423. doi: 10.1016/j.fuel.2011.06.001

Liu, Z., Wei, M., Cui, G., Yang, X., Gu, H., and Yang, L. (2018). Optimization of arabinogalactan and taxifolin extraction process from Dahurian larch (*Larix gmelinii*) and evaluation of the effects on activities of α-amylase, α-glycosidase, and pancreatic lipase in vitro. *J. Food Biochem.* 42, 1–14. doi: 10.1111/jfbc.12607

Magel, E., Jay-Allemand, C., and Ziegler, H. (1994). Formation of heartwood substances in the stemwood of Robinia pseudoacacia L. II. Distribution of nonstructural carbohydrates and wood extractives across the trunk. *Trees* 8, 165–171. doi: 10.1007/BF00196843

Mayer, I., Koch, G., and Puls, J. (2006). Topochemical investigations of wood extractives and their influence on colour changes in American black cherry (*Prunus serotina* Borkh.). *Holzforschung* 60, 589–594. doi: 10.1515/HF.2006.100

Miklečić, J., Španić, N., and Jirouš-Rajković, V. (2012). Wood color changes by ammonia fuming. *BioResources* 7, 3767–3778.

Monties, B. (1991). Plant cell walls as fibrous lignocellulosic composites: relations with lignin structure and function. *Anim. Feed Sci. Technol.* 32, 159–175. doi: 10.1016/0377-8401(91)90019-O

Nair, M. N., Shah, J., and Pandalai, R. C. (1981). Wood anatomy and histochemical changes of sapwood during heartwood formation in Bridelia retusa Spreng. *Proc. Plant Sci.* 90, 425–433. doi: 10.1007/BF03052940

Nisula, L. (2018). Wood extractives in conifers: a study of stemwood and knots of industrially important species. Åbo, Finland: Åbo Akademi University Press. 372. Available at: https://www.doria.fi/handle/10024/149340.

Pereira, L., Flores-Borges, D. N. A., Bittencourt, P. R. L., Mayer, J. L. S., Kiyota, E., Araújo, P., et al. (2018). Infrared nanospectroscopy reveals the chemical nature of pit membranes in water-conducting cells of the plant xylem. *Plant Physiol.* 177, 1629–1638. doi: 10.1104/pp.18.00138

Piqueras, S., Duponchel, L., Offroy, M., Jamme, F., Tauler, R., and de Juan, A. (2013). Chemometric strategies to unmix information and increase the spatial description of hyperspectral images: a single-cell case study. *Anal. Chem.* 85, 6303–6311. doi: 10.1021/ac4005265

Piqueras, S., Duponchel, L., Tauler, R., and de Juan, A. (2014). Monitoring polymorphic transformations by using *in situ* Raman hyperspectral imaging and image multiset analysis. *Anal. Chim. Acta* 819, 15–25. doi: 10.1016/j.aca.2014.02.027

Piqueras, S., Krafft, C., Beleites, C., Egodage, K., von Eggeling, F., Guntinas-Lichius, O., et al. (2015). Combining multiset resolution and segmentation for hyperspectral image analysis of biological tissues. *Anal. Chim. Acta* 881, 24–36. doi: 10.1016/j.aca.2015.04.053

Pomar, F., Merino, F., and Barceló, A. R. (2002). O -4-Linked coniferyl and sinapyl aldehydes in lignifying cell walls are the main targets of the Wiesner (phloroglucinol-HCl) reaction. *Protoplasma* 220, 17–28. doi: 10.1007/s00709-002-0030-y

Popescu, C., Popescu, M., Singurel, G., Vasile, C., Argyropoulos, D. S., and Willfor, S. (2007). Spectral characterization of eucalyptus wood. *Appl. Spectrosc.* 61, 1168–1177. doi: 10.1366/000370207782597076

Popescu, C., Popescu, M., and Vasile, C. (2010). Characterization of fungal degraded lime wood by FT-IR and 2D IR correlation spectroscopy Characterization of fungal degraded lime wood by FT-IR and 2D IR correlation spectroscopy. *Microchem. J.* 95, 377–387. doi: 10.1016/j.microc.2010.02.021

Romeo, M., Mohlenhoff, B., and Diem, M. (2006). Infrared micro-spectroscopy of human cells: causes for the spectral variance of oral mucosa (buccal) cells. *Vib. Spectrosc.* 42, 9–14. doi: 10.1016/j.vibspec.2006.04.009.Infrared

Ruddick, J. N. R., and Xie, C. (1994). Why does Douglas-fir heartwood turn black when treated with ammoniacal copper preservatives? *For. Prod. J.* 44, 57–61.

Saito, K., Mitsutani, T., Imai, T., Matsushita, Y., Yamamoto, A., and Fukushima, K. (2008). Chemical differences between sapwood and heartwood of Chamaecyparis obtusa detected by ToF-SIMS. *Appl. Surf. Sci.* 255, 1088–1091. doi: 10.1016/J.APSUSC.2008.05.145

Savitzky, A., and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36, 1627–1639. doi: 10.1021/ac60214a047

Schultz, T. P., and Nicholas, D. D. (2000). Naturally durable heartwood: evidence for a proposed dual defensive function of the extractives. *Phytochemistry* 54, 47–52. doi: 10.1016/S0031-9422(99)00622-6

Schwanninger, M., Rodrigues, J. C., Pereira, H., and Hinterstoisser, B. (2004). Effects of short-time vibratory ball milling on the shape of FT-IR spectra of wood and cellulose. *Vib. Spectrosc.* 36, 23–40. doi: 10.1016/j.vibspec.2004.02.003

Shi, J., Xing, D., and Li, J. (2012). FTIR Studies of the changes in wood chemistry from wood (1640 C = O).pdf. *Energy Proc.* 16, 758–762. doi: 10.1016/j.egypro.2012.01.122

Stewart, D., Wilson, H., Hendra, P., and Morrison, M. (1995). Fourier-transform infrared and raman spectroscopic study of biochemical and chemical treatments of oak wood (Quercus rubra) and barley (Hordeum vulgare) Straw. *J. Agric. Food Chem.* 43, 2219–2225. doi: 10.1021/jf00056a047

Tauler, R., Smilde, A., and Kowalski, B. (1995). Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemometr.* 9, 31–58. doi: 10.1016/0169-7439(95)00047-X

Tauler, R., Maeder, M., and de Juan, A. (2009). "Multiset Data Analysis: Extended Multivariate Curve Resolution," in *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*. Eds. S. D. Brown, R. Tauler and B. Walczak (Amsterdam, The Netherlands: Elsevier), 473–506. Available at: http://www.iasbs.ac.ir/chemistry/chemometrics/history/11th/multiset data analysis.pdf [Accessed February 16, 2015].

Taylor, A. M., Gartner, B. L., and Morrell, J. J. (2002). Heartwood formation and natural durability - A review. *Wood Fiber Sci.* 34, 587–611.

Tomppo, L., Tiitta, M., Laasko, T., Harju, A., Venäläinen, M., and Lappalainen, R. (2011). Study of stilbene and resin acid content of Scots pine heartwood by electrical impedance spectroscopy (EIS). *Holzforschung* 65, 643–649. doi: 10.1515/HF.2011.111

Traoré, M., Kaal, J., and Martínez Cortizas, A. (2018). Differentiation between pine woods according to species and growing location using FTIR-ATR. *Wood Sci. Technol.* 52, 487–504. doi: 10.1007/s00226-017-0967-9

Umesh, P., and Rajai, H. (2010). "Vibrational Spectroscopy," in *Lignin and Lignans: Advances in Chemistry*. Eds. C. Heitner, D. Dimmel and J. A. Schmidt (Boca Raton, Florida: CRC Press Taylor&Francis group), 104–129.

Umesh, P., James, D., and Sally, A. (2011). FT – Raman investigation of Milled-Wood Lignins: softwood, hardwood, and chemically modified black spruce lignins. *J. Wood Chem. Technol.* 31, 324–344. doi: 10.1080/02773813.2011.562338

Valette, N., Perrot, T., Sormani, R., Gelhaye, E., and Morel-Rouhier, M. (2017). Antifungal activities of wood extractives. *Fungal Biol. Rev.* 31, 113–123. doi: 10.1016/j.fbr.2017.01.002

Wang, Q., Kretlow, A., Beekes, M., Naumann, D., and Miller, L. (2005). *In situ* characterization of prion protein structure and metal accumulation in scrapie-infected cells by synchrotron infrared and X-ray imaging. *Vib. Spectrosc.* 38, 61–69. doi: 10.1016/j.vibspec.2005.02.023

Wang, X., Deng, Y., Li, Y., Kjoller, K., Roy, A., and Wang, S. (2016). *In situ* identification of the molecular-scale interactions of phenol-formaldehyde resin and wood cell walls using infrared nanospectroscopy. *RSC Adv.* 6, 76318–76324. doi: 10.1039/c6ra13159j

Weiland, J. J., and Guyonnet, R. (2003). Study of chemical modifications and fungi degradation of thermally modified wood using DRIFT spectroscopy. *Holz als Roh und Werkstoff.* 61, 216–220. doi: 10.1007/s00107-003-0364-y

Willför, S., and Holmbom, B. (2004). Isolation and characterisation of water soluble polysaccharides from Norway spruce and Scots pine. *Wood Sci. Technol.* 38, 173–179. doi: 10.1007/s00226-003-0200-x

Windeisen, E., Wegener, G., Lesnino, G., and Schumacher, P. (2002). Investigation of the correlation between extractives content an natural durability in 20 cultivated larch trees. *Holz als Roh - und Werkst.* 60, 373–374. doi: 10.1007/s00107-002-0314-0

Windig, W., and Guilment, J. (1991). Interactive self-modeling mixture analysis. *Anal. Chem.* 63, 1425–1432. doi: 10.1021/ac00014a016

Yamauchi, S., Iijima, Y., and Doi, S. (2005). Spectrochemical characterization by FT-Raman spectroscopy of wood heat-treated at low temperatures: Japanese larch and beech. *Japan Wood Res. Soc* 51, 498–506. doi: 10.1007/s10086-004-0691-6

Zu, S., Yang, L., Huang, J., Ma, C., Wang, W., Zhao, C., et al. (2012). Micronization of Taxifolin by supercritical antisolvent process and evaluation of radical scavenging activity. *Int. J. Mol. Sci.* 13, 8869–8881. doi: 10.3390/ijms13078869

Zule, J., Čufar, K., and Tišler, V. (2015). Lipophilic Extractives in Heartwood of European Larch (Larix decidua Mill.). *Drv. Ind.* 66, 305–313. doi: 10.5552/drind.2015.1442

Zule, J., Čufar, K., and Tišler, V. (2017). Hydrophilic Extractives in Heartwood of European Larch (Larix decidua Mill.). *Drv. Ind.* 67, 363–370. doi: 10.5552/drind.2016.1618

ATR-FTIR Microspectroscopy Brings a Novel Insight Into the Study of Cell Wall Chemistry at the Cellular Level

Clément Cuello[1], Paul Marchand[1†], Françoise Laurans[1,2], Camille Grand-Perret[1], Véronique Lainé-Prade[1,2], Gilles Pilate[1] and Annabelle Déjardin[1*]

[1] INRAE, ONF, BioForA, Orléans, France, [2] INRAE, Phenobois, Orléans, France

Wood is a complex tissue that fulfills three major functions in trees: water conduction, mechanical support and nutrient storage. In Angiosperm trees, vessels, fibers and parenchyma rays are respectively assigned to these functions. Cell wall composition and structure strongly varies according to cell type, developmental stages and environmental conditions. This complexity can therefore hinder the study of the molecular mechanisms of wood formation, underlying the construction of its properties. However, this can be circumvented thanks to the development of cell-specific approaches and microphenotyping. Here, we present a non-destructive microphenotyping method based on attenuated total reflectance–Fourier transformed infrared (ATR-FTIR) microspectroscopy. We applied this technique to three types of poplar wood: normal wood of staked trees (NW), tension and opposite wood of artificially tilted trees (TW, OW). TW is produced by angiosperm trees in response to mechanical strains and is characterized by the presence of G fibers, exhibiting a thick gelatinous extra-layer, named G-layer, located in place of the usual S2 and/or S3 layers. By contrast, OW located on the opposite side of the trunk is totally deprived of fibers with G-layers. We developed a workflow for hyperspectral image analysis with both automatic pixel clustering according to cell wall types and identification of differentially absorbed wavenumbers (DAWNs). As pixel clustering failed to assign pixels to ray S-layers with sufficient efficiency, the IR profiling and identification of DAWNs were restricted to fiber and vessel cell walls. As reported elsewhere, this workflow identified cellulose as the main component of the G-layers, while the amount in acetylated xylans and lignins were shown to be reduced. These results validate ATR-FTIR technique for *in situ* characterization of G layers. In addition, this study brought new information about IR profiling of S-layers in TW, OW and NW. While OW and NW exhibited similar profiles, TW fibers S-layers combined characteristics of TW G-layers and of regular fiber S-layers. Unexpectedly, vessel S-layers of the three kinds of wood showed significant differences in IR profiling. In conclusion, ATR-FTIR microspectroscopy offers new possibilities for studying cell wall composition at the cell level.

Keywords: ATR-FTIR microspectroscopy, phenotyping, poplar, cell wall, G-layer, wood

INTRODUCTION

Trees are able to live long and reach a considerable height, partly thanks to the remarkable properties of their wood. Indeed, wood —or secondary xylem—fulfills three major functions in trees: (i) water conduction from the roots to the crown, (ii) mechanical support of the ever-increasing mass of the growing tree submitted to different environmental cues (wind, slope, light, …), and (iii) storage of temporary reserves, important for tree perennial growth (Plomion et al., 2001; Déjardin et al., 2010). In Angiosperm trees, vessels, fibers and parenchyma rays are respectively assigned to each of these functions. These different xylem cell types originate from the differentiation of cambial cells. When reaching their final dimensions, xylem cells build a secondary cell wall that will be deposited onto the primary wall and middle lamella. Vessels are rapidly submitted to programmed cell death after lignification of their cell walls, then fibers have the same destiny (Courtois-Moreau et al., 2009), while ray cells remain alive for several years (Nakaba et al., 2012). Therefore, wood results in the complex assembly of the cell walls of dead fibers and vessels, connected to living parenchyma rays (Déjardin et al., 2010).

Wood cell walls exhibit a multi-layered structure, determinant for wood mechanical properties: middle lamella and primary cell wall are overlaid by three layers of secondary cell walls (SCW), named S1, S2, and S3, S2 being the thickest. Each layer results from the assembly of semi-crystalline cellulose microfibrils embedded in an amorphous matrix of polysaccharides and lignins. They differ according to cellulose microfibril orientation and the nature and proportion of the matrix components (pectins, hemicelluloses and lignins) (Mellerowicz and Sundberg, 2008). The middle lamella and primary cell wall are known to be rich in lignins, pectins and xyloglucans, with apparently cellulose microfibrils oriented at random. By contrast, the S-layers are rich in xylans with oriented cellulose microfibrils: the angle formed between the cell axis and cellulose microfibrils, named microfibril angle (MFA) is wide in the S1- and S3-layers, while it is fairly low in the S2-layer (Barnett and Bonham, 2004). At the end of cell wall deposition, lignification occurs, leading to the impermeabilization of the cell walls. Lignins are phenolic polymers merely resulting, in hardwood species, from the polymerization of two classes of monolignols (coniferyl alcohol or G unit and synapyl alcohol or S unit) that differ in their degree of methoxylation (Freudenberg and Neish, 1968). They are synthesized in the cytoplasm, then transported to the cell wall, turned into chemical radicals by the activity of laccases and peroxidases and probably get polymerized in a random manner. Lignin structure is therefore highly diverse and difficult to predict (van Parijs et al., 2010).

Wood cell walls are highly variable according to developmental processes or various environmental stresses. First, it is known that secondary cell wall composition differs between vessels, fibers and ray cells (e.g. Donaldson et al., 2001). The cell wall of vessel elements is enriched in G-units while fiber cell wall is richer in S-units (Terashima et al., 1993). Wood cell wall composition can also be affected by various environmental cues. For example, in response to mechanical strains,

Angiosperm trees produce tension wood on the upper side of inclined trunks or branches in order to reorient them or at least to reach a new equilibrium. In poplar, tension wood fibers exhibit a thick gelatinous extra-layer, named G-layer (Jourez, 1997). Located in place of the usual S2 and/or S3 layers, G-layer is rich in cellulose, rhamnogalacturonan type I pectins (RG-I), and arabinogalactan proteins (AGP) (Gorshkova et al., 2015; Guedes et al., 2017), with nearly no lignin (Joseleau et al., 2004; Pilate et al., 2004; Gierlinger and Schwanninger, 2006). In G-layer, cellulose is more crystalline and MFA is close to 0° (Norberg and Meier, 1966; Gorshkova et al., 2015).

Thus, wood is both a highly complex and variable tissue, which does not facilitate the study of the molecular mechanisms underlying its formation and the building of its properties. However, this hindrance may be alleviated thanks to the development of cell-specific approaches and microphenotyping. Particularly, the development of microphenotyping methods may provide extensive biochemical information at the cell wall level. In the last decade, Fourier Transform InfraRed (FTIR) spectroscopy made possible high throughput biochemical analyses of lignocellulosic biomass. Indeed, the wavenumbers emitted by infrared beams are differentially absorbed according to the chemical bonds, therefore the functional groups, present in the sample. Cell wall composition may be determined by prediction equations based on near-infrared absorbance spectra of ground plant materials. This technique has been successfully used to predict cell wall composition in *Arabidopsis thaliana* (Jasinski et al., 2016), forage crops (Fairbrother and Brink, 1990; Molano et al., 2016; Baldy et al., 2017; Li et al., 2017), rice (Huang et al., 2017) and poplar (Gebreselassie et al., 2017). Mid-infrared (MIR) spectroscopy is complementary to predictive near-infrared (NIR) spectroscopy. Indeed, MIR spectroscopy targets the fundamental vibrations of the molecules, which gives a good visualization of the molecular bonds present in the sample (Bertrand and Dufour, 2006; Allison, 2011) and also to assign variations in the absorbance at a given wavenumber to a class of compounds (McCann et al., 1992; Kacurakova and Wilson, 2001). For example, FTIR spectroscopy was used to discriminate between different Arabidopsis cell wall mutants (Chen et al., 1998). FTIR spectroscopy coupled with microscopy imaging makes possible *in situ* analysis of cell wall assembly (Mouille et al., 2003; Gierlinger et al., 2008; Gorzsás et al., 2011; Chazal et al., 2014; Gierlinger, 2018).

In this methodological study, we present a non-destructive high-throughput microphenotyping method based on ATR-FTIR microspectroscopy, one of the most resolving methods currently available. The objective was to develop a method to obtain IR spectra specific to wood cell types, fibers, vessels and rays. Therefore, we have developed a high-throughput pipeline for hyperspectral image analysis with both automatic pixel clustering and identification of differentially absorbed wavenumbers between samples. We have implemented this pipeline on different types of wood, including tension wood, a model that has been extensively characterized from a biochemical point of view. We confirmed previously known results, which validates this *in situ* approach, and we generated for the first time an IR profile for vessel walls in three different types of wood.

MATERIALS AND METHODS

Plant Material and Sample Preparation

The study was carried out on six three-month-old ramets, originating from *in vitro* micropropagated shoots of the INRA 717-1B4 clone (*Populus tremula* L. x *Populus alba* L.). Two ramets were grown in a greenhouse in spring 2011 and the four others in the autumn and winter 2012. In the latter case, the plants were supplemented with sufficient amount of artificial light to keep the cambium active. In order to induce the production of TW, three of these trees (TT_11-A in 2011, TT_12-A and TT_12-B in 2012) were tilted at 45° two months before sampling. The three other poplar plants (UT_11-C, UT_12-C and UT_12-D) were grown upright in the same conditions and produce no or very few TW. Three-centimeter-long stem fragments were sampled at the base of each tree. NW was collected on upright trees, TW and OW on the upper and lower side of artificially tilted trees, respectively. The samples were stored at –20°C until use.

For each tree, 20 μm-thick cross-sections were cut from the frozen wood samples using a RM2155 microtome (LEICA, Wetzlar, Germany). The sections were ethanol-dried between two glass slides to ensure flatness and stored at 23°C and 45% RH until ATR-FTIR analysis.

Hyperspectral Imaging

ATR-FTIR images were produced from stem cross-sections by a mid-IR light ($1,800-850$ cm^{-1}) at a 4 cm^{-1} spectral resolution using a Spotlight 400 FTIR imaging system coupled to a Spectrum 400 FTIR spectrophotometer (PERKIN ELMER, Wellesley, USA). According to the manufacturer's procedure, the wood section was pushed by pressure into direct contact with the tip of the 600-μm diameter plane Germanium crystal. **Supplementary Figure 1** shows the contact area after pressure release showing the intimate contact of the sample with the crystal. Sixteen scans per pixel were taken in order to enhance signal-to-noise ratio. For each type of wood (TW, OW and NW), three (100 x 100) μm² images from each cross section were taken at a (1.56 x 1.56) μm² pixel size. The acquisition time of an image of this size was 40 minutes. This maximal pixel resolution corresponded to an oversampling factor of two, compared with the diffraction-limited spatial resolution of 3.1 μm. At this pixel resolution, the image size was 4,096 pixels. Therefore, for each type of wood, the data set was composed of 36,864 spectra (3 different trees x 3 images x 4,096 pixels). The spectra were then corrected using different functions of the SpectrumImage software (v.1.6.4): background correction during acquisition, noise reduction and atmospheric correction, using default parameters. No ATR correction was performed.

Multivariate Image Analysis

We analyzed ten classes of cell wall corresponding to the existing associations between cell wall layers (G- or S-layers), cell types (fiber, ray or vessel) and wood types (TW, OW or NW): TW fiber G-layer, TW fiber S-layer, OW fiber S-layer, NW fiber S-layer, TW ray S-layer, OW ray S-layer, NW ray S-layer, TW vessel S-layer, OW vessel S-layer and NW vessel S-layer. Multivariate image analysis has been performed independently for each type of wood using R (v. 3.4.4) (**Figure 1**). Principal component analysis (PCA) was applied on raw spectra using PCA function implemented in FactoMineR package (v. 1.40). False RGB images were then obtained using quilt.plot (fields, v.9.6). Red, blue and green intensities were based on pixel contribution to first, second and third principal component, respectively. Based on these false RGB images, specific mean representative spectrum (MRS) were built for each cell wall class, from the average of 90 selected spectra (that is 30 spectra manually selected in one of the three images of each biological replicate of each wood type). Both MRS and pixel spectra were corrected using the detrend function implemented in the prospectr package (v. 0.1.3). Correlations between each pixel spectrum and MRS of its type of wood were determined using the Spearman correlation test implemented in cor.test (stats, v.3.5.2). The correlation score was used to assign pixels to a cell wall class or to lumen. A pixel was considered as lumen when its correlation scores with all MRS were below the threshold value of 0.90. Pixels with very similar correlation scores with all MRS (less than 1% difference) were classified as "not assigned" (NA). Finally, the pixel was assigned to the cell wall class whose MRS has the highest correlation score with the pixel.

Identification of Differentially Absorbed Wavenumbers (DAWNs)

DAWNs were determined using the pairwise Wilcoxon test from the R stats package (v.3.5.2). A Bonferroni adjusted *P*-value of 0.001 was used as cut-off criterion. Only DAWNs located at a local maximum ($\pm$ 3 cm^{-1}) were considered for further analysis. Local maxima were identified on the average spectra resulting from all pixels assigned to a class of cell wall. A single wavenumber (σ) was considered as a local maximum when its absorbance was higher compared to σ-2 and σ+2 cm^{-1}. The assignation of DAWNs to cell wall components was based on an *in-lab* database (**Table S1**) compiling data from the literature and from the Histochem database (Durand et al., 2019; https://pfl. grignon.inra.fr/shistochem/).

Statistical Analysis

For each cell type, a PCA was carried out on the whole spectra to discriminate between the different types of wood. Loadings were calculated as follow: $Loadings = Eigenvector \cdot \sqrt{Eigenvalues}$. Due to the huge size of the data set, pixels corresponding to the cell wall of each cell type in each image were averaged to produce heatmaps using heatmap.2 function (gplots package, v. 3.0.1.1), scaling on columns. Heatmaps dendrograms were calculated considering Pearson's correlation between wavenumbers as a distance.

RESULTS

From Raw Infra-Red Images to Pixel Clusters of the Different Cell Wall Types

One of the most important steps of the workflow was to produce a good assignation of the pixels to the different cell
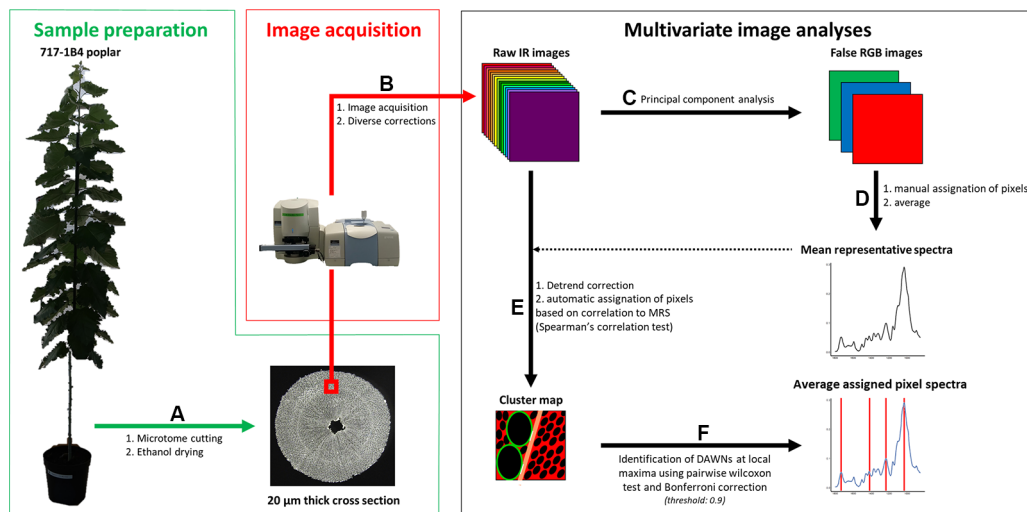
FIGURE 1 | Workflow. Sample preparation. Poplar trees were grown in a greenhouse for three months. **(A)** 20 μm-thick ethanol dried cross section were produced. Image acquisition. **(B)** (100 x 100) μm² images were acquired from 20 μm thick cross sections using an infrared (IR) microscope. Background subtraction, noise reduction and atmospheric correction were applied on raw IR images using the SpectrumImage software (v.1.6.4). Multivariate image analyses were carried out using homemade scripts encoded in R (v. 3.4.4). **(C)** Principal component analyses (PCA) were performed on raw spectra. RGB intensities were derived from pixel contribution to PC1 for Red, PC2 for Green and PC3 for Blue. **(D)** Ninety pixels were manually attributed to each of ray, fiber and vessel cell walls and to G-layers. **(E)** Mean representative spectra (MRS), which is the pixel average of each of these classes, and pixel spectra were detrend. Correlations between pixel spectra and MRS were determined using Spearman's correlation test. Pixel were assigned to a given class using a 0.9 correlation score threshold. Green: vessel, tan: ray, red: fiber, black: lumen, white: NA. **(F)** Differentially absorbed wavenumbers (DAWNs) were identified using pairwise Wilcoxon test. A Bonferroni adjusted P-value of 0.001 was used as cut-off criterion.

types from the different types of wood. Classic clustering methods, such as hierarchical clustering based on principal components, Ascending Hierarchical Classification, k-nearest neighbor and k-means were not conclusive. The most effective clustering method was the assignation based on Spearman's correlation score between pixel spectra and a mean representative spectrum (MRS) for each class of cell wall.

We applied this method to the spectra used to calculate the MRS of the different classes of cell wall (**Table S2**). When applied to pixels with a known assignment, our method correctly assigned 48% to 100% of the pixels depending on the cell wall category (**Figure 2A**). All the spectra from TW fiber G-layer were properly assigned, whereas, no spectrum from any TW S-layer class was wrongly assigned to the G-layer class. Likewise, all spectra of OW fiber S-layers were correctly assigned, while only three of them were identified as ray S-layers in NW and two of them were also identified as ray S-layers in TW. Our method correctly assigned a large part of vessel spectra: 78% in OW, 82% in NW and 73% in TW. In TW, nearly all wrongly assigned vessel spectra (23 out of 24) were classified as ray S-layers. In NW, only six vessel spectra were classified as ray S-layers, while four others had a correlation score below 0.90 with all MRS and were therefore classified as lumen. Finally, six spectra that presented similar correlation values to all MRS were in consequence not assigned. In OW, six vessel spectra were

categorized as ray S-layers, three as fiber S-layers, four as lumen and seven were not assigned. Conversely, the wrong assignment of ray spectra was rather high reaching 52% in OW, 28% in NW and 38% in TW.

Subsequently, we used this workflow to assign the 36,864 spectra from the pixels of each type of wood. In NW, 50% of the spectra were categorized as fiber S-layers, 14% as ray S-layers, 9% as vessel S-layers and 23% as lumen, while 4% of the spectra could not be assigned. In OW, 41% of the spectra corresponded to fiber S-layers, 16% to ray S-layers and 11% to vessel S-layers, while 26% of the spectra, with a correlation score below 0.90, were considered as lumen. Finally, 6% of the spectra with similar correlation levels to MRS from all classes, were not assigned and therefore excluded from the analysis. In TW, the fiber G-layer was the largest class with 44% of the spectra. The fiber S-layers, ray S-layers, vessel S-layers and lumen categories were composed of 34%, 8%, 2% and 12% of the spectra, respectively. Interestingly, unlike OW and NW, all TW spectra have been assigned (**Figure 2B**). The comparison between the raw IR images and the corresponding cluster maps (**Figure 3**) revealed a correct adjustment especially for fibers and vessels, indicative of the efficiency of the clustering method used to predict these two cell types. The pixels that could not be assigned mainly corresponded to pixels at the interface between cell types. As the assignation for ray S-layers was not very efficient, IR profiling was restricted to fiber and vessel SCW.
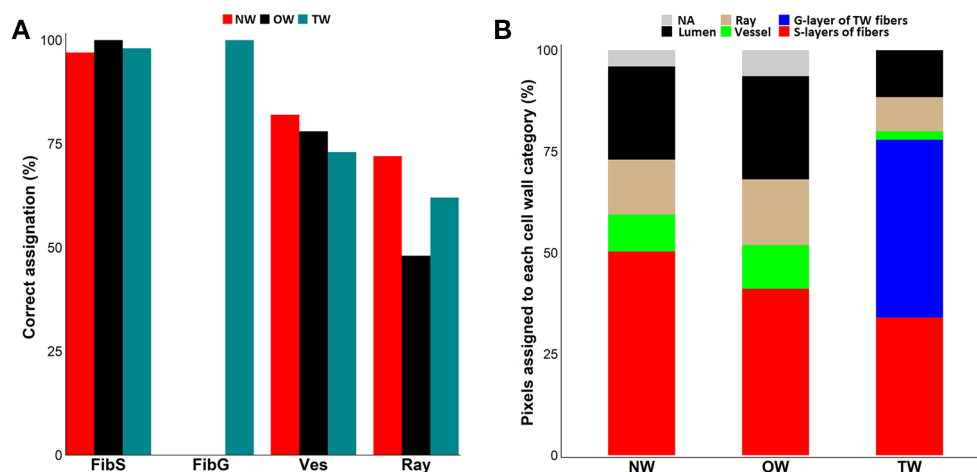
FIGURE 2 | Assignation of pixel to cell-wall categories. **(A)** Clustering efficiency evaluated on 90 manually assigned pixels per cell wall type. Red: Normal wood (NW), Black: Opposite wood (OW), Pigeon blue: Tension wood (TW), FibS: fiber S-layers, FibG: TW fiber G-layers, Ves: vessel S-layers, Ray: ray S-layers. **(B)** Distribution of the 36,864 pixels in cell wall classes after assignation using Spearman's correlation score between pixel spectra and mean representative spectra. NW, normal wood; OW, opposite wood; TW, tension wood; Red, fiber S-layers; Blue, TW fiber G-layer; Green, vessel S-layers; Tan, ray S-layers; Black, lumen; Grey, not assigned.

IR Profiling of Fiber and Vessel SCW

Fibers

The fiber spectra, assigned by automatic clustering, were first analysed by PCA. When performed on all the images, PCA clearly discriminated three groups of spectra: TW fiber G-layer, TW fiber S-layers and a group combining spectra from both NW and OW fiber S-layers. However, in the latter group, two populations of spectra were detected, with the minor one gathering spectra all coming from the same 4 images (3 from NW and 1 from OW) (**Supplementary Figure S2**). These 4 images were therefore removed from the subsequent analysis as they may induce unwanted variability for technical or biological reasons. A new PCA was performed (**Figure 4A**), whose first dimension explained 42.5% of the variance and discriminated rather well TW fiber G-layer from the fiber S-layers of both OW and NW, whereas TW fiber S-layers stood in between. **Figure 4B** shows the average infrared spectrum of pixels assigned to the four types of fiber layers. The TW fiber G-layer spectrum shows significant changes compared to the OW and NW fiber S-layers spectra, including the disappearance of two bands at 1,594 and 1,506 cm^{-1}, corresponding to aromatic skeletal vibration of lignins (**Table S1**) and the appearance of a well-resolved double band at 1,336 and 1,316 cm^{-1}, that can be attributed to cellulose (**Table S1**, O-H in-plane bending and CH2 rocking vibration, respectively). It is also characterized by the appearance of a band at 1200 cm^{-1}, that can be attributed to OH in-plane deformation in cellulose. This band is masked in the other S-layers because of the overlap of a more intense band at 1236 cm^{-1}, corresponding to acetyl and carboxyl vibration in xylan and to C-C, C-O and C=O stretch in lignins. Finally, a last band is clearly visible in TW fiber G and S-layers at 1,052 cm^{-1}, corresponding to C-O valence vibration, mainly from C3-O3H, most probably from cellulose. This latter band is detectable only

as a shoulder on the spectra of NW and OW fiber S-layers. The spectrum of TW fiber S-layers is intermediate, in particular with weak bands at 1,506 cm^{-1}, resembling the spectra of NW and OW; it also presents the G-layer characteristic bands at 1,136, 1,316, and 1,052 cm^{-1}. In terms of absorbance differences, the bands at 1,736 and 1,236 cm^{-1} mainly differentiate TW and OW/NW fibers, with a higher absorbance in the latter (**Figure 4B**): they can be mainly attributed to acetylated xylans, and to a lesser extent to lignins (**Table S1**). The PCA loading plots underline the spectral features associated with the first and second dimensions, which are responsible for most of the variance (**Supplementary Figure S3A**). The samples were discriminated only by the first dimension (**Figure 4A**). The loading for PC1 is quite complex and cannot be easily interpreted by simple changes in few compounds. The loading is positive mostly for six ranges of absorbance bands (1,576–14, 1,492–74, 1,446–28, 1,412–1,384, 1,360–1,280, 1,014–992 cm^{-1}). The first four ranges of bands do not correspond to known IR bands, while the two last groups may correspond to cellulose (crystalline cellulose at 1,335 cm^{-1} and 1,318–12 cm^{-1}, C–O valence vibration at 996–85 cm^{-1}) and lignins (1,330–24 cm^{-1}, S ring plus G ring condensed). The loading is negative for four ranges of bands (1,758–08, 1,268–06, 1,136–16, 1,098–68 cm^{-1}). The first three groups contain bands attributed to hemicelluloses, cellulose and lignins (**Table S1**). In order to identify more precisely differences between cell wall, we have identified DAWNs in the vicinity of a local maximum (± 3 cm^{-1}), that were further used to generate a heatmap (**Figure 4C**, **Table S3**). The analysis was limited to these DAWNs because the most relevant biological information in an IR spectrum is carried by the wavenumbers at the level of the most intense bands. The heatmap shows 2 clusters of DAWNs that have a very contrasted absorbance profile between TW G-layer and S-layers of OW or NW fibers, while
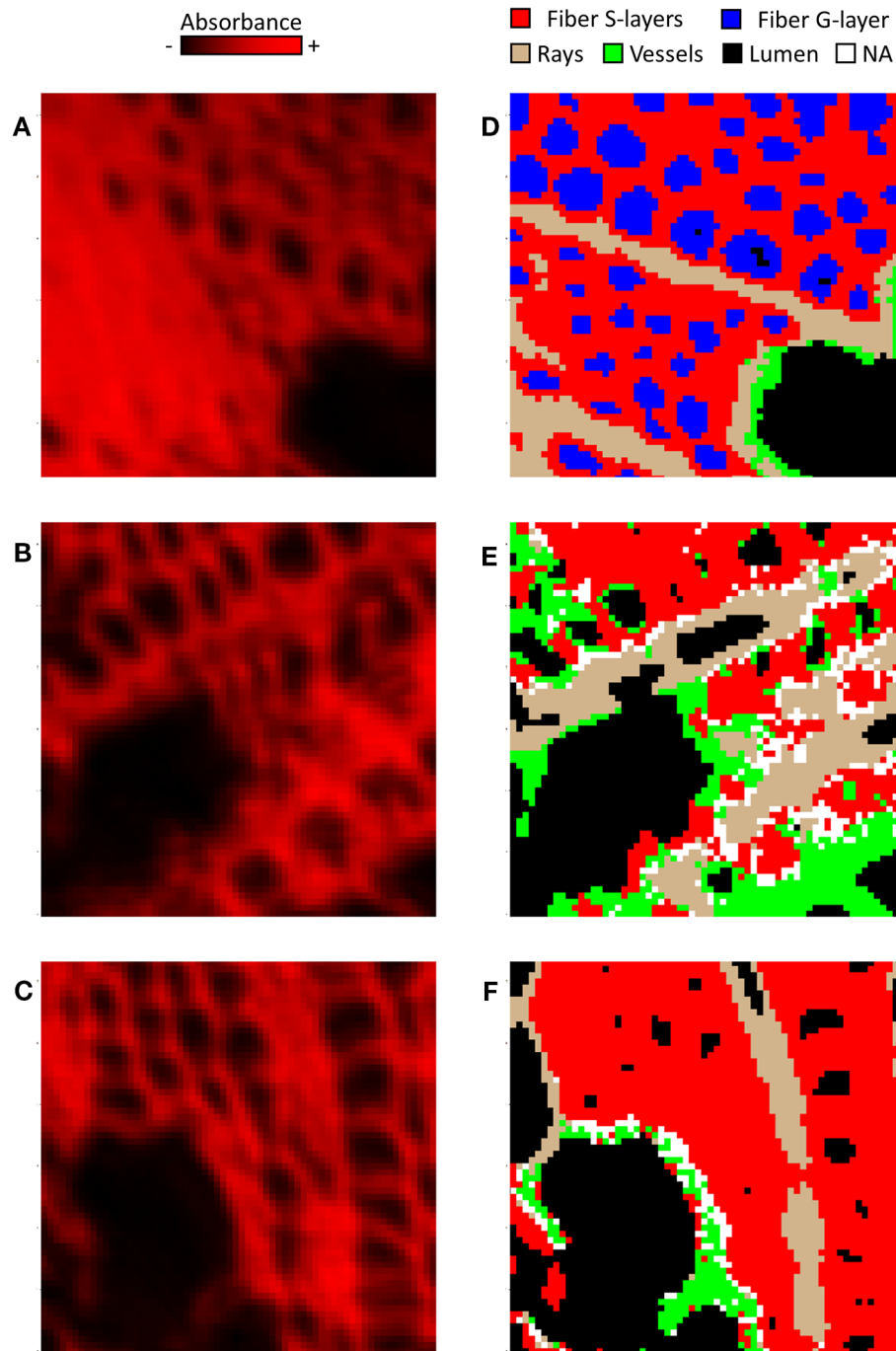
FIGURE 3 | Comparison of raw images **(A–C)** and cluster map **(D–F)** of tension wood **(A, D)**, opposite wood **(B, E)** and normal wood **(C, F)** using one representative image per type of wood. Note that these images were not the ones used to determine MRS. **(A–C)** In raw images, the intensity in red reflects the mean absorbance of the pixel. **(D–F)** Red, fiber S-layers; Blue, fiber G-layer; Green, vessel S-layers; Tan, ray S-layers; White, NA; Black, lumen.

TW fiber S-layer has an intermediate profile. Cluster B gathers DAWNs with a high absorbance in NW and OW S-layers and a low absorbance in TW S- and G-layers. On the reverse, Cluster E groups DAWNs with a high absorbance in TW G-layer and to a lesser extent in TW S-layers and a low absorbance in NW and OW

S-layers. Cluster B contains DAWNs in the range of 1,742–28 and 1,238–32 cm^{-1} related mainly to acetylated xylans, and in the range of 1,596–90 and 1,508–04 cm^{-1} related to lignins. Surprisingly, this latter group was weakly absorbed in NW fiber S-layers, compared to OW fiber S-layers. Cluster B also contains
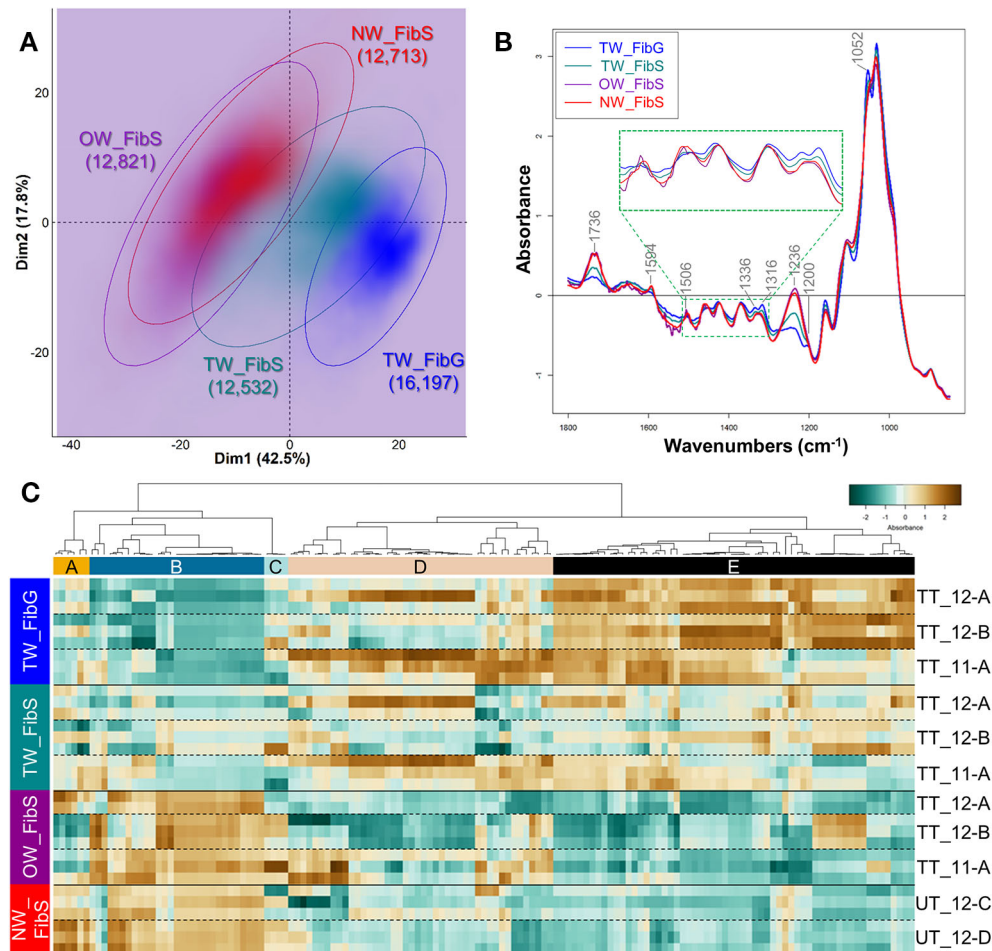
FIGURE 4 | Chemical differences between the G-layer from tension wood fibers and the S-layers from fibers of tension, opposite and normal wood. **(A)** PCA score plot. Blue, TW fiber G-layer; Turquoise, TW fiber S-layers; Purple, OW fiber S-layers; Red, NW fiber S-layers. Ellipses encompass 95% of the data in a normal distribution. Numbers in brackets refer to the number of pixels assigned to the different cell wall categories. Color intensity reflects the density of individuals. **(B)** Average ATR-FTIR spectra. Blue, TW fiber G-layer; Turquoise, TW fiber S-layers; Purple, OW fiber S-layers; Red, NW fiber S-layers. The green dotted line insertion represents a zoom of the [1,510–1,300] cm^{-1} region. **(C)** Heatmap of differentially absorbed wavenumbers. TW_FibG, TW fiber G-layer, TW_FibS; TW fiber S-layers; OW_FibS, OW fiber S-layers; NW_FibS, NW fiber S-layers.

DAWNs in the range of 1,248–1,244 cm^{-1}, that may be attributed to lignins (stretching of phenolics), and in the range of 1,464–58 cm^{-1}, where both cellulose, lignins and xylans may contribute (CH2 of pyran ring symmetric scissoring; OH- and CH-deformation). Cluster E contains DAWNs that may correspond to crystalline cellulose (1,428–26, 1,336–1,314 cm^{-1}, which appear as a double band in the TW G-layer spectra). Likewise, it contains DAWNs in the range of 1,282–78 and 1,162–56 cm^{-1} that may be also attributed to cellulose. Cluster E contains DAWNs in the range of 1,056–50 and 1,036–30 cm^{-1}, that can be attributed to cellulose, hemicelluloses (and also to lignins for the latter group). Interestingly, it contains a group of DAWNs at 1,518–12 cm^{-1} that can be attributed to lignins (aromatic skeletal vibration, higher absorbance in G lignin in comparison to S lignin, **Table S1**) and one in the range of 1,556–48 cm^{-1} than can be attributed to proteins (Amide II: N-H deformation + stretching contribution

from C-N stretching, **Table S1**). Finally, some groups of DAWNs cannot be easily related to known IR assignment (1,790–74, 1,696–92, 1,562, 1,540–30, 1,478–74 cm^{-1}).

Figure 5 shows images reconstructed on the basis of the absorbance level for each pixel at a given wavenumber. Three wavenumbers were chosen: 1,736 cm^{-1} (acetylated xylans), 1,316 cm^{-1} (crystalline cellulose), and 1,236 cm^{-1} (acetylated xylans and lignins). In particular, the images clearly show the absorbance gradient between fiber G-layer and S-layer in TW.

Vessels
PCA analysis on vessel S-layers spectra slightly discriminates the spectra from TW, OW and NW, according to the first two dimensions, explaining 36 and 17.7% of the variance (**Figure 6A**). When the first dimension partially separates TW vessel S-layers from both OW and NW vessel S-layers, the second one
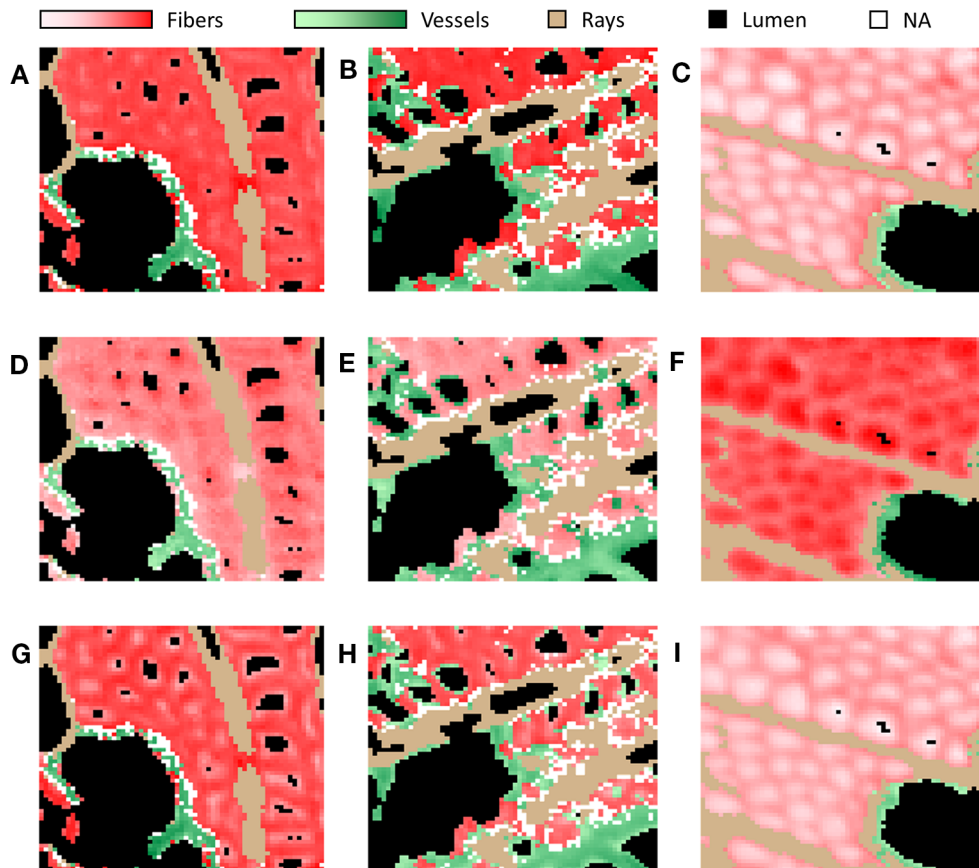
FIGURE 5 | Infrared images of selected DAWNs of representative images of NW **(A, D, G)**, OW **(B, E, H)** and TW **(C, F, I)**. Selected DAWNs: 1,236 cm^{-1} corresponding to acetylated xylans and lignins **(A–C)**, 1,316 cm^{-1} corresponding to crystalline cellulose **(D–F)** and 1,736 cm^{-1} corresponding to acetylated xylans **(G–I)**. Red, fiber S-/G-layers; Green, vessel S-layers; Tan, ray S-layers; White, NA; Black, lumen. Red, and green intensities are proportionate to the absorbance of pixels assigned to fiber S/G-layers and vessel S-layers, respectively.

slightly distinguishes OW and NW vessel S-layers. The average infrared spectra of the vessel S-layers from the three types of wood appear rather similar (**Figure 6B**). The average spectra of OW and TW vessel S-layers exhibit an important noise between approximately 1,800 and 1,450 cm^{-1}. Nevertheless, we can clearly distinguish in these two spectra, three double bands at 1,736, 1,460, and 1,374 cm^{-1}, while there is only one band in the NW spectrum at these wavenumbers. Acetylated xylans mainly contribute to the first band and cellulose, hemicelluloses and lignins to the two others. The bands at 1,736 (acetylated xylans), 1,332–26 (lignins, S ring plus G ring condensed), 1,236 (acetylated xylans and lignins) and 1,034 cm^{-1} (cellulose and lignins) makes possible to differentiate TW, NW and OW vessel S-layers (**Figure 6B**). The PCA loading plots (**Supplementary Figure S3B**) show that on PCA first dimension, the loading is positive for four spectral zones (1,580–26, 1,494–74, 1,450–1,380, and 976–904 cm^{-1}), that cannot be easily assigned to reported IR bands except the region of 1,430–21 cm^{-1} that can be assigned to cellulose, hemicelluloses and lignins (**Table S1**), 996–85 cm^{-1} to cellulose and 925–15 cm^{-1} to lignins. The loading is negative for

three ranges of bands (1,760–22 cm^{-1}, 1,256–36 cm^{-1} and 1,138–1,034 cm^{-1}), corresponding respectively to lignins and acetylated xylans, lignins, cellulose and lignins (**Table S1**). On the second dimension, two main ranges of bands (1,800–1,756 cm^{-1} and 1,206–1,164 cm^{-1}) presented positive loadings, the second one being potentially attributed to cellulose, hemicelluloses and lignins (**Table S1**). The loading is negative mainly for two groups of bands (1,676–56 and 1,650–1,590 cm^{-1}). They both correspond to lignins (C=O stretch in conjugated p-substituted aryl ketones/aromatic skeletal vibrations, higher absorbance in G lignin in comparison to S lignin, **Table S1**). As done for fiber analysis, the DAWNs in the vicinity of a local maximum (± 3 cm^{-1}) were used to generate a heatmap (**Figure 6C**, **Table S4**). The heatmap shows more variability between samples, compared to the heatmap for fiber, most likely because vessel/fiber area ratio is small in wood and in consequence, there were fewer vessel pixels available. Cluster A corresponds to DAWNs that have a low absorbance in NW vessel S-layers, and a high absorbance in TW vessel S-layers, while the absorbance is generally contrasted between the two sections of OW analyzed.
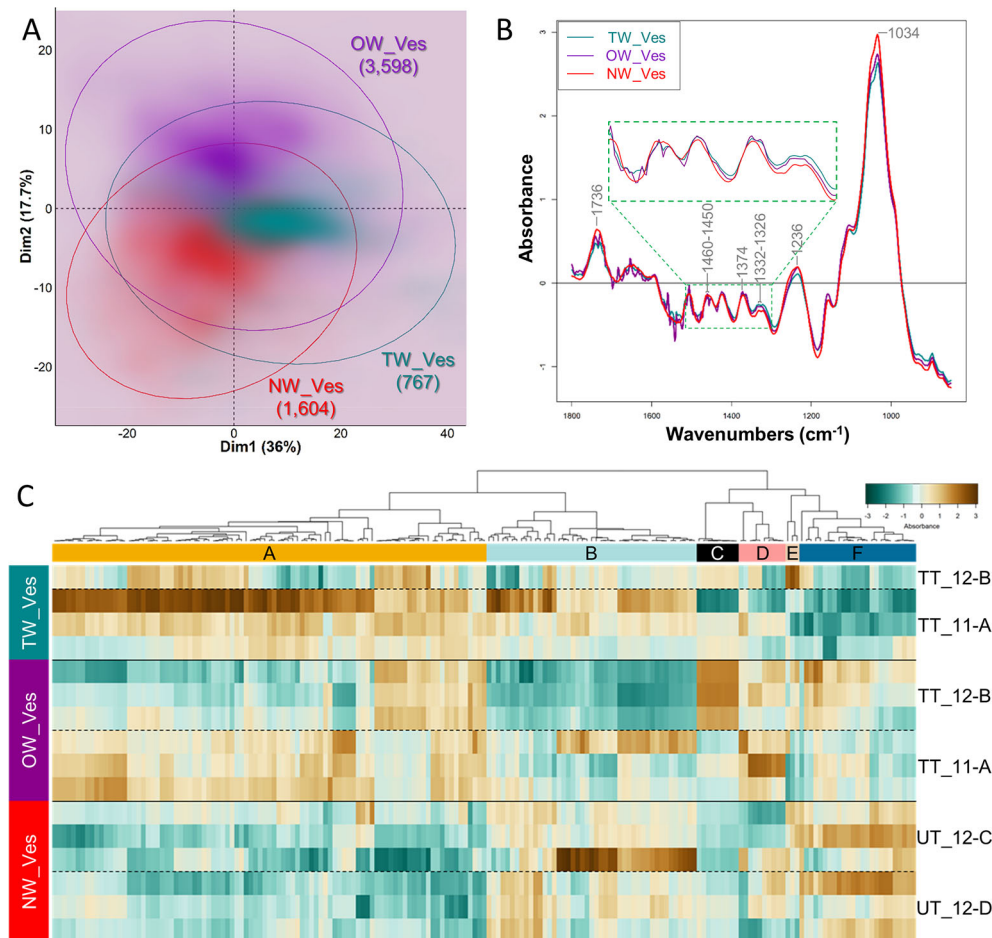
FIGURE 6 | Chemical differences between vessel S-layers of tension, opposite and normal wood. **(A)** PCA score plot. Turquoise, TW vessel S-layers; Purple, OW vessel S-layers; Red, NW vessel S-layers. Ellipses encompass 95% of the data in a normal distribution. Numbers in brackets refer to the number of pixels assigned to the different cell wall categories. Color intensity reflects the density of individuals. **(B)** Average ATR-FTIR spectra. Turquoise, TW vessel S-layers; Purple, OW vessel S-layers; Red, NW vessel S-layers. The green dotted line insertion represents a zoom of the [1,500–1,300] cm^{-1} region. **(C)** Heatmap of differentially absorbed wavenumbers. TW_Ves, TW vessel S-layers; OW_Ves, OW vessel S-layers; NW_Ves, NW vessel S-layers.

Cluster A contains DAWNs in the range of 1,364–60 and 1,346–42 cm^{-1}, that may be mostly attributed to cellulose and in the range of 1,336–22 cm^{-1}, that may be attributed to crystalline cellulose and also to lignins (S ring plus G ring condensed). DAWNs of 1,374–72 cm^{-1} may be attributed both to cellulose and hemicelluloses (CH deformation vibration and CH bending, **Table S1**). The isolated DAWN of 1,506 cm^{-1} can be assigned to lignins (aromatic skeletal vibrations; G > S). Besides, a large number of DAWNs cannot be easily related to known IR assignment (1,686–82, 1,572–68, 1,560–58, 1,538–28, 918, 914–10, 936–26 cm^{-1}). Cluster F gathers DAWNs that are less absorbed in TW vessel S-layers, when compared to OW and NW vessel S-layers. DAWNs in the range of 1,236–32 cm^{-1} correspond to acetylated xylans. The 3 groups of DAWNs in the same region of the IR spectra (1,744–36, 1,732–26, 1,722–10 cm^{-1}) are due to C=O stretching, that are mainly due to acetylated xylans too, but the involvement of esterified lignins cannot be excluded

(**Table S1**). On **Figure 6B**, the spectrum for NW shows a unique band around 1,734, while several bands and shoulders can be seen in TW and OW, probably related to a decreased amount of xylans, making visible other individual bands. Finally, both cellulose and lignins may contribute to DAWNs in the range of 1,040–32 cm^{-1} (C–O valence vibration, mainly from C3–O3H; aromatic C-H in-plane deformation, G > S).

DISCUSSION

In this study, we set up a non-destructive method for the microphenotyping of wood cell walls, based on ATR-FTIR microspectroscopy, that makes possible to obtain, in a rather short time, IR spectra from the different wood cell types, thus overcoming the complexity of wood tissue. We have developed a high-throughput pipeline for hyperspectral image analysis

(**Figure 1**) with both automatic pixel clustering according to cell wall types and identification of DAWNs between samples. This pipeline was used to compare the biochemical composition of the cell wall from both fibers and vessels originating from three kinds of wood in poplar: normal wood of upright trees, tension and opposite wood of artificially tilted trees. As the preparation of the samples is very simple—microtome sectioning at 20 µm, followed by ethanol drying—the acquisition of the images is fast—40 min for a (100 x 100) µm² hyperspectral image—and the analysis is mostly an automated process, this method is suitable to screen for differences in IR spectra, a large number of samples, at the fiber or vessel cell walls. Thanks to the high refractive index of the Germanium crystal (4.01), the spatial resolution of ATR-FTIR microspectrocopy is 3.1 µm, and based on an oversampling factor of two, it is possible to reach a pixel resolution of (1.56 x 1.56) µm². This is sufficient to get IR spectra at the level of secondary cell walls, and with respects to the possibility of signal contamination, we obtained the best results for fiber cell walls, due to their higher thickness compared to vessel and ray cell walls. However, this method is not resolutive enough to reach the level of the cell wall layer. For such studies, some other methods of IR nanospectroscopy, such as scattering scanning near-field optical microscopy (s-SNOM) or atomic force microscopy infrared nanospectroscopy (AFM-IR), would be more appropriate. Indeed, the resolution of such methods is beyond the classical diffraction limit of light. These techniques were recently used to investigate the chemical composition of primary cell walls in *Populus* vascular cambium (Pereira et al., 2018). The drawback of these techniques is that ultrathin sections need to be prepared, with potential chemical treatments like DMSO that may alter cell wall composition, and therefore IR spectra.

A critical step to identify differences in cell wall composition using ATR-FTIR microspectroscopy is to unequivocally assign pixels to the right type of cell wall. The clustering method used here makes possible such an assignment with a high efficiency for both fiber S- and G-layers and with an acceptable level of success for vessels. The low score for the assignation of ray pixels probably results from the high diversity in ray cell walls. Indeed, as rays are uniseriate in *Populus* (IAWA Committee, 1989), there are very few ray-to-ray cell walls on transverse sections and most ray cell walls are either neighbored by fiber or vessel cell walls. Therefore, in this study, the pixels classified as ray cell walls correspond in fact to a mixture of ray/vessel and ray/fiber cell walls, whose composition may differ accordingly. Ray cells can be classified into three different cell types, named contact cells, intermediate cells and isolated cells (Braun, 1967; Nakaba et al., 2012) that may have cell walls with rather different features. While contact cells are located predominantly within the upper and lower lines of individual ray lines and are connected to adjacent vessel elements through pits, intermediate cells are located within the same radial lines but are not adjacent to vessel elements. Isolated cells are located within the radial cell lines with no connections with vessels (Braun, 1967). In *Populus sieboldii x P. grandidentata*, it has been shown that secondary cell wall thickening is initiated earlier in the upper and lower cell lines of a ray (Nakaba et al., 2012). In

addition, parenchyma ray cells stay alive longer than vessels and fibers and their cell wall composition may also be subjected to modifications according to their age. These aspects may be at least partly circumvented by applying ATR-FTIR technique on longitudinal sections in place of transverse sections. The score for the assignation of vessel pixels is not as high as for fiber pixels: as for rays, vessel pixels may in fact correspond to a mixture of vessel/vessel, vessel/ray and vessel/fiber cell walls. The fact that vessels have thin cell walls may also contribute to some incorrect assignments. Finally, the lower score for the assignment of pixels in OW compared to NW and TW (**Figure 2B**) may be linked to the reduced growth of xylem cells resulting from the position of OW on the lower side of the tilted stem. The smaller number of pixels assigned to vessel cell wall in TW compared to OW and NW is in accordance with the observations from a number of studies using other approaches, that the number of vessels is decreased in TW in comparison to OW (Jourez et al., 2001; Ruelle et al., 2006; Tarmian et al., 2009) or to NW (Chow, 1947). We also observed in TW a decreased number of lumen-associated pixels, which certainly result in the reduced lumen size due to the presence of the G-layer (Clair et al., 2006; Ruelle et al., 2006; Déjardin et al., 2010; Gorshkova et al., 2010). Finally, the smaller number of pixels assigned to fiber S-layers in TW supports the fact that S-layers are reduced in G fibers, in relation to the deposition of the G-layer.

ATR-FTIR microscopy provided IR profiling of fibers and vessel cell wall layers and DAWNs were identified and used to characterize differences in biochemical composition between layers. Another critical step was to correctly assign these DAWNs to wood polymers as the absorption band from a given functional group may originate from several wood polymers and bands are overlapping in the spectrum. This is the reason why we focused on the DAWNs in the vicinity of visible peaks on the spectra to get information easier to interpret on a biological level. TW and G fibers have been extensively studied using various complementary methods (FTIR, Raman imaging, biochemical analyses, immunolocalization). These are valuable data to validate our results using *in situ* IR profiling. In this study, the IR spectra of TW fiber G-layer is indeed very similar to what was obtained by Olsson et al. (2011) on TW and isolated G-layers, with a large decrease of the bands at 1,736 and 1,236 cm⁻¹, no bands at 1,594 and 1,506 cm⁻¹, and the appearance of 3 specific bands at 1,136, 1,316, and 1,200 cm⁻¹. Basically, it has also the same profile as the one presented in Gierlinger et al. (2008), except that we did not observe, the large band at 1,645 cm⁻¹ attributed to the deformation of vibration of adsorbed and free water: this probably reflects a difference in the water status of the samples analyzed. Our results indicate that TW fiber G-layers contained higher amounts of crystalline cellulose and lower amounts of acetylated xylans and lignins than fiber S-layers, although it is not so striking for TW S-layers (see supra for discussion on this). This is consistent with a number of studies demonstrating that the G-layer was mainly composed of cellulose (e.g. Norberg and Meier, 1966; Nishikubo et al., 2007; Guedes et al., 2017). This is also in accordance with the absence or the low levels of xylans and lignins reported

elsewhere. The higher absorbance of a DAWN attributed to proteins is relevant since arabinogalactan proteins were reported as very abundant in TW in a number of studies (e.g. Lafarguette et al., 2004; Andersson-Gunnerås et al., 2006). Interestingly, while most lignin specific DAWNs exhibited lower absorbance in the G-layer, one of them presented a higher absorbance (1,518–12 cm^{-1}, aromatic skeletal vibration, higher absorbance in G lignin in comparison to S lignin). In addition, a positive loading on the first dimension of PCA was found for a group of bands corresponding also to lignins (1,330–24 cm^{-1}, S ring plus G ring condensed). This suggests that some lignins, likely different from the fiber S-layer lignins, are actually present in the G-layer, as previously reported in poplar (Joseleau et al., 2004; Gierlinger and Schwanninger, 2006) and in other species (Ghislain et al., 2016; Higaki et al., 2017). Thus, the *in situ* technique described in this paper is validated on TW fiber G-layers, since the results we obtained on individual G-layers were in accordance with other studies made at the tissue level or on isolated G-layers.

Regarding fiber S-layers, many DAWNs exhibited important differences in absorbance between TW fiber S-layers and fiber S-layers from OW or NW. From our observations, the former seemed to have a biochemical composition rather similar to what is described in the G-layer, while some features remained common to the S-layer from OW and NW. For example, the absorption bands for lignins and acetylated xylans were present in TW fiber S-layers but lower than in OW and NW. These observations may reflect biological differences in fiber cell wall assembly between the different wood types. This may be related to the fact that the S-layers in G fibers are not fully developed, in comparison to the S-layers of OW and NW fibers. Alternatively, it remains possible that the induction of tension wood formation, beside inducing G-layer formation, may also act on S-layer differentiation. However, we cannot rule out that these differences result from signal contaminations by the adjacent G-layer. Therefore, it remains to verify with a larger number of observations that our results reflect the biological reality or if the close proximity of the G-layer interferes with the measurements.

IR profiling was also obtained for vessel S-layers. The two first dimensions of PCA were able to discriminate between the three types of wood (**Figure 6A**) showing that there are some chemical differences in the cell wall of vessels, even between NW and OW, while fiber S-layers were very similar for both types of wood. The analysis of DAWNs between the different kinds of wood was not easy. Indeed, there was an important variability between the different sections analyzed [see for example differences between TT_11A and TT_12B, the 2 replicates for OW vessel cell wall (**Figure 6C**)]. This certainly reflects the low number of pixels analyzed (in comparison to fibers) that mainly results from the fact that vessels have thin cell wall and are not numerous (still in comparison to fibers). As for fiber cell walls, there are obviously important differences between vessel cell wall from NW and TW, with more cellulose and less acetylated xylans in the latter, two features also observed in TW fiber cell wall. Although, G-layer is absent from vessel cell wall, we cannot rule out a response at the vessel cell wall to stem tilting. The differences between NW and OW vessel S-layers, based on PC2 loadings, are mainly explained by differences in lignins. As we did not find any information on vessel cell wall composition in the bibliography, it is difficult to go further on these observations without performing measurements on a larger number of samples or using complementary techniques like immunocytochemistry.

CONCLUSION

Thanks to the high pixel resolution of ATR-FTIR hyperspectral images—(1.56 x 1.56) µm² —we were able to get IR profiling of fibers and vessels at the cell layer level. We have developed a high-throughput pipeline for hyperspectral image analysis with both automatic pixel clustering and identification of differentially absorbed wavenumbers between samples. We have implemented this pipeline on different types of wood, including tension wood, a model that has been extensively characterized from a biochemical point of view. We confirmed previously known results on G layers, which validates this *in situ* approach, and we generated an IR profile for vessel S-layers in 3 different types of wood, as well as for TW fiber S-layers, providing information on the composition of cell layers rarely described in the literature. This methodological study paves the way for the microphenotyping of cell walls to rapidly and finely characterize various genetic resources or trees submitted to various stresses, and thus to overcome the complexity of wood tissue for all studies aimed at understanding wood formation. We consider that this technique is useful to give some clues about the cell wall compounds underlying cell wall differences. However, with regards to the inherent limitations of this technique (spatial resolution, difficulties to unambiguously assign bands to specific cell wall compounds), the hypotheses raised have to be validated by a complementary low-throughput technique, like immunocytochemistry.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

CC, FL, GP, and AD designed the research. CC, PM, FL, CG-P, and VL-P acquired the data. CC, PM, and AD analyzed the data. CC, GP and AD wrote the article.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00105/full#supplementary-material

SUPPLEMENTARY FIGURE 1 | Crystal impacts on UT_12-C cross-section. *Green circles pinpoint the three impacts.*

SUPPLEMENTARY FIGURE 2 | PCA score plot of fibres on the whole data set. *Blue: TW fibre G-layer, Turquoise: TW fibre S-layers, Purple: OW fibre S-layers, Red: NW fibre S-layers. Ellipses encompass 95% of the data in a normal distribution. Color intensity reflects the density of individuals.*

SUPPLEMENTARY FIGURE 3 | PCA loading plots of fibre S- and G-layers **(A)** and vessel S-layers **(B)**. *Red: PC1, Green: PC2.*

REFERENCES

Allison, G. G. (2011). "Application of Fourier Transform Mid-Infrared Spectroscopy (FTIR) for Research into Biomass Feed-Stocks" in *Fourier Transforms - New Analytical Approaches and FTIR Strategies*. Ed. G. Nikolic (London: IntechOpen). doi: 10.5772/15785

Andersson-Gunnerås, S., Mellerowicz, E. J., Love, J., Segerman, B., Ohmiya, Y., Coutinho, P. M., et al. (2006). Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *Plant J. Cell Mol. Biol.* 45, 144–165. doi: 10.1111/j.1365-313X.2005.02584.x

Baldy, A., Jacquemot, M.-P., Griveau, Y., Bauland, C., Reymond, M., and Mechin, V. (2017). Energy values of registered corn forage hybrids in france over the last 20 years rose in a context of maintained yield increase. *Am. J. Plant Sci.* 08, 1449–1461. doi: 10.4236/ajps.2017.86099

Barnett, J. R., and Bonham, V. A. (2004). Cellulose microfibril angle in the cell wall of wood fibers. *Biol. Rev.* 79, 461–472. doi: 10.1017/S1464793103006377

Bertrand, D., and Dufour, E. (2006). *La spectroscopie infrarouge et ses applications analytiques* (Tec & Doc. Paris: Lavoisier).

Braun, H. J. (1967). Entwicklung und Bau der Holzstrahlen unter dem Aspekt der kontakt-isolations-differenzierung gegenüber dem hydrosystem. i. das prinzip der kontakt-isolations-differenzierung. *Holzforschung* 21, 33–37. doi: 10.1515/hfsg.1967.21.2.33

Chazal, R., Robert, P., Durand, S., Devaux, M.-F., Saulnier, L., Lapierre, C., et al. (2014). Investigating lignin key features in maize lignocelluloses using infrared spectroscopy. *Appl. Spectrosc.* 68, 1342–1347. doi: 10.1366/14-07472

Chen, L., Carpita, N. C., Reiter, W.-D., Wilson, R. H., Jeffries, C., and McCann, M. C. (1998). A rapid method to screen for cell-wall mutants using discriminant analysis of Fourier transform infrared spectra. *Plant J.* 16, 385–392. doi: 10.1046/j.1365-313x.1998.00301.x

Chow, K. Y. (1947). A comparative study of the structure and chemical composition of tension wood and normal wood in beech (*Fagus sylvatica* L.). *Forestry* 20, 62–77. doi: 10.1093/forestry/20.1.62

Clair, B., Ruelle, J., Beauchêne, J., Prévost, M. F., and Fournier, M. (2006). Tension wood and opposite wood in 21 tropical rain forest species. *IAWA J.* 27, 329–338. doi: 10.1163/22941932-90000158

Courtois-Moreau, C. L., Pesquet, E., Sjödin, A., Muñiz, L., Bollhöner, B., Kaneda, M., et al. (2009). A unique program for cell death in xylem fibers of *Populus* stem. *Plant J.* 58, 260–274. doi: 10.1111/j.1365-313X.2008.03777.x

Déjardin, A., Laurans, F., Arnaud, D., Breton, C., Pilate, G., and Leplé, J.-C. (2010). Wood formation in Angiosperms. *C. R. Biol.* 333, 325–334. doi: 10.1016/j.crvi.2010.01.010

Donaldson, L., Hague, J., and Snell, R. (2001). Lignin distribution in coppice poplar, linseed and wheat straw. *Holzforschung* 55, 379–385. doi: 10.1515/HF.2001.063

Durand, S., Papineau, P., Robert, P., Devaux, M.-F., and Guillon, F. (2019). HISTOCHEM: a database of reference spectra for plant cell wall polymers. *Berichte Aus Dem. Julius Kühn-Inst.* 204, 91–92. doi: 10.5073/berjki.2019.204.000

Fairbrother, T. E., and Brink, G. E. (1990). Determination of cell wall carbohydrates in forages by near infrared reflectance spectroscopy. *Anim. Feed Sci. Technol.* 28, 293–302. doi: 10.1016/0377-8401(90)90160-A

Freudenberg, K., and Neish, A. C. (1968). *Constitution and biosynthesis of lignin* (Berlin Heidelberg: Springer-Verlag), 129 pp.

Gebreselassie, M. N., Ader, K., Boizot, N., Millier, F., Charpentier, J.-P., Alves, A., et al. (2017). Near-infrared spectroscopy enables the genetic analysis of chemical properties in a large set of wood samples from Populus nigra (L.) natural populations. *Ind. Crops Prod.* 107, 159–171. doi: 10.1016/j.indcrop.2017.05.013

Ghislain, B., Nicolini, E.-A., Romain, R., Ruelle, J., Yoshinaga, A., Alford, M. H., et al. (2016). Multilayered structure of tension wood cell walls in Salicaceae *sensu lato* and its taxonomic significance. *Bot. J. Linn. Soc* 182, 744–756. doi: 10.1111/boj.12471

Gierlinger, N., and Schwanninger, M. (2006). Chemical imaging of poplar wood cell walls by confocal raman microscopy. *Plant Physiol.* 140, 1246–1254. doi: 10.1104/pp.105.066993

Gierlinger, N., Goswami, L., Schmidt, M., Burgert, I., Coutand, C., Rogge, T., et al. (2008). *In situ* FT-IR microscopic study on enzymatic treatment of poplar wood cross-sections. *Biomacromolecules* 9, 2194–2201. doi: 10.1021/bm800300b

Gierlinger, N. (2018). New insights into plant cell walls by vibrational microspectroscopy. *Appl. Spectrosc. Rev.* 53, 517–551. doi: 10.1080/05704928.2017.1363052

Gorshkova, T. A., Gurjanov, O. P., Mikshina, P. V., Ibragimova, N. N., Mokshina, N. E., Salnikov, V. V., et al. (2010). Specific type of secondary cell wall formed by plant fibers. *Russ. J. Plant Physiol.* 57, 328–341. doi: 10.1134/S1021443710030040

Gorshkova, T., Mokshina, N., Chernova, T., Ibragimova, N., Salnikov, V., Mikshina, P., et al. (2015). Aspen tension wood fibers contain β-(1→4)-galactans and acidic arabinogalactans retained by cellulose microfibrils in gelatinous walls. *Plant Physiol.* 169, 2048–2063. doi: 10.1104/pp.15.00690

Gorzsás, A., Stenlund, H., Persson, P., Trygg, J., and Sundberg, B. (2011). Cell-specific chemotyping and multivariate imaging by combined FT-IR microspectroscopy and orthogonal projections to latent structures (OPLS) analysis reveals the chemical landscape of secondary xylem: The chemical landscape of secondary xylem. *Plant J.* 66, 903–914. doi: 10.1111/j.1365-313X.2011.04542.x

Guedes, F. T. P., Laurans, F., Quemener, B., Assor, C., Lainé-Prade, V., Boizot, N., et al. (2017). Non-cellulosic polysaccharide distribution during G-layer formation in poplar tension wood fibers: abundance of rhamnogalacturonan I and arabinogalactan proteins but no evidence of xyloglucan. *Planta* 246, 857–878. doi: 10.1007/s00425-017-2737-1

Higaki, A., Yoshinaga, A., and Takabe, K. (2017). Heterogeneous distribution of xylan and lignin in tension wood G-layers of the S1+G type in several Japanese hardwoods. *Tree Physiol.* 37, 1767–1775. doi: 10.1093/treephys/tpx144

Huang, J., Li, Y., Wang, Y., Chen, Y., Liu, M., Wang, Y., et al. (2017). A precise and consistent assay for major wall polymer features that distinctively determine

biomass saccharification in transgenic rice by near-infrared spectroscopy. *Biotechnol. Biofuels* 10, 294. doi: 10.1186/s13068-017-0983-x

IAWA Committee (1989). *IAWA List of Microscopic Features for Hardwood Identification IAWA Bulletin* Vol. 10. Eds. E. A. Wheeler, P. Baas and P. Gasson (Leiden: Brill publishers), 219–332.

Jasinski, S., Lécureuil, A., Durandet, M., Bernard-Moulin, P., and Guerche, P. (2016). *Arabidopsis* Seed Content QTL Mapping Using High-Throughput Phenotyping: The Assets of Near Infrared Spectroscopy. *Front. Plant Sci.* 7, 1682. doi: 10.3389/fpls.2016.01682

Joseleau, J.-P., Imai, T., Kuroda, K., and Ruel, K. (2004). Detection *in situ* and characterization of lignin in the G -layer of tension wood fibers of *Populus deltoides*. *Planta* 219, 338–345. doi: 10.1007/s00425-004-1226-5

Jourez, B., Riboux, A., and Leclercq, A. (2001). Anatomical characteristics of tension wood and opposite wood in young inclined stems of poplar (*Populus euramericana* cv 'GHOY'). *IAWA J.* 22, 133–157. doi: 10.1163/22941932-90000274

Jourez, B. (1997). Le bois de tension. 1 Définition et distribution dans l'arbre. *Biotechnol. Agron. Société Environ.* 1, 100–112. http://hdl.handle.net/2268/95404

Kacurakova, M., and Wilson, R. (2001). Developments in mid-infrared FT-IR spectroscopy of selected carbohydrates. *Carbohydr. Polym.* 44, 291–303. doi: 10.1016/S0144-8617(00)00245-9

Lafarguette, F., Leplé, J.-C., Déjardin, A., Laurans, F., Costa, G., Lesage-Descauses, M.-C., et al. (2004). Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. *New Phytol.* 164, 107–121. doi: 10.1111/j.1469-8137.2004.01175.x

Li, K., Wang, H., Hu, X., Ma, F., Wu, Y., Wang, Q., et al. (2017). Genetic and quantitative trait locus analysis of cell wall components and forage digestibility in the Zheng58 × HD568 Maize RIL population at anthesis stage. *Front. Plant Sci.* 8, 1472. doi: 10.3389/fpls.2017.01472

McCann, M. C., Hammouri, M., Wilson, R., Belton, P., and Roberts, K. (1992). Fourier transform infrared microspectroscopy is a new way to look at plant cell walls. *Plant Physiol.* 100, 1940–1947. doi: 10.1104/pp.100.4.1940

Mellerowicz, E. J., and Sundberg, B. (2008). Wood cell walls: biosynthesis, developmental dynamics and their implications for wood properties. *Curr. Opin. Plant Biol.* 11, 293–300. doi: 10.1016/j.pbi.2008.03.003

Molano, M. L., Cortés, M. L., Ávila, P., Martens, S. D., and Muñoz, L. S. (2016). Near infrared spectroscopy (NIRS) calibration equations to predict nutritional quality parameters of tropical forages. *Trop. Grassl.-Forrajes Trop.* 4, 139–145. doi: 10.17138/tgft(4)139-145

Mouille, G., Robin, S., Lecomte, M., Pagant, S., and Höfte, H. (2003). Classification and identification of *Arabidopsis* cell wall mutants using Fourier-Transform InfraRed (FT-IR) microspectroscopy. *Plant J.* 35, 393–404. doi: 10.1046/j.1365-313X.2003.01807.x

Nakaba, S., Begum, S., Yamagishi, Y., Jin, H.-O., Kubo, T., and Funada, R. (2012). Differences in the timing of cell death, differentiation and function among three different types of ray parenchyma cells in the hardwood *Populus sieboldii × P. grandidentata*. *Trees* 26, 743–750. doi: 10.1007/s00468-011-0640-0

Nishikubo, N., Awano, T., Banasiak, A., Bourquin, V., Ibatullin, F., Funada, R., et al. (2007). Xyloglucan Endo-transglyxosylase (XET) Functions in gelatinous layers of tension wood fibers in poplar—a glimpse into the mechanism of the balancing act of trees. *Plant Cell Physiol.* 48, 843–855. doi: 10.1093/pcp/pcm055

Norberg, P. H., and Meier, H. (1966). Physical and chemical properties of the gelatinous layer in tension wood fibers of aspen (*Populus tremula* L.). *Holzforschung* 20, 174–178. doi: 10.1515/hfsg.1966.20.6.174

Olsson, A.-M., Bjurhager, I., Gerber, L., Sundberg, B., and Salmén, L. (2011). Ultra-structural organisation of cell wall polymers in normal and tension wood of aspen revealed by polarisation FTIR microspectroscopy. *Planta* 233, 1277–1286. doi: 10.1007/s00425-011-1384-1

Pereira, L., Flores-Borges, D. N. A., Bittencourt, P. R. L., Mayer, J. L. S., Kiyota, E., Araujo, P., et al. (2018). Infrared nanospectroscopy reveals the chemical nature of pit membranes in water-conducting cells of plant xylem. *Plant Physiol.* 177, 1629–1638. doi: 10.1016/j.crvi.2004.07.006

Pilate, G., Déjardin, A., Laurans, F., and Leplé, J.-C. (2004). Tension wood as a model for functional genomics of wood formation. *New Phytol.* 164, 63–72. doi: 10.1111/j.1469-8137.2004.01176.x

Plomion, C., Leprovost, G., and Stokes, A. (2001). Wood formation in trees. *Plant Physiol.* 127, 1513–1523. doi: 10.1104/pp.010816

Ruelle, J., Clair, B., Beauchêne, J., Prévost, M. F., and Fournier, M. (2006). Tension wood and opposite wood in 21 tropical rain forest species. *IAWA J.* 27, 341–376. doi: 10.1163/22941932-90000159

Tarmian, A., Remond, R., Faezipour, M., Karimi, A., and Perré, P. (2009). Reaction wood drying kinetics: tension wood in *Fagus sylvatica* and compression wood in *Picea abies*. *Wood Sci. Technol.* 43, 113–130. doi: 10.1007/s00226-008-0230-5

Terashima, N., Fukushima, K., He, L.-F., and Takabe, K. (1993). "Comprehensive Model of the Lignified Plant Cell Wall," in *Forage Cell Wall Structure and Digestibility* (Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America), 247–270. doi: 10.2134/1993.foragecellwall.c10

van Parijs, F. R. D., Morreel, K., Ralph, J., Boerjan, W., and Merks, R. M. H. (2010). Modeling lignin polymerization. I. simulation model of dehydrogenation polymers. *Plant Physiol.* 153, 1332–1344. doi: 10.1104/pp.110.154468

Chemical Analysis of Pollen by FT-Raman and FTIR Spectroscopies

Adriana Kenđel[1] and Boris Zimmermann[2,3]*

[1] *Division of Analytical Chemistry, Department of Chemistry, Faculty of Science, University of Zagreb, Zagreb, Croatia,*
[2] *Faculty of Science and Technology, Norwegian University of Life Sciences, Ås, Norway,* [3] *Division of Organic Chemistry and Biochemistry, Ruđer Bošković Institute, Zagreb, Croatia*

Pollen studies are important for the assessment of present and past environment, including biodiversity, sexual reproduction of plants and plant-pollinator interactions, monitoring of aeroallergens, and impact of climate and pollution on wild communities and cultivated crops. Although information on chemical composition of pollen is of importance in all of those research areas, pollen chemistry has been rarely measured due to complex and time-consuming analyses. Vibrational spectroscopies, coupled with multivariate data analysis, have shown great potential for rapid chemical characterization, identification and classification of pollen. This study, comprising 219 species from all principal taxa of seed plants, has demonstrated that high-quality Raman spectra of pollen can be obtained by Fourier transform (FT) Raman spectroscopy. In combination with Fourier transform infrared spectroscopy (FTIR), FT-Raman spectroscopy is obtaining comprehensive information on pollen chemistry. Presence of all the main biochemical constituents of pollen, such as proteins, lipids, carbohydrates, carotenoids and sporopollenins, have been identified and detected in the spectra, and the study shows approaches to measure relative and absolute content of these constituents. The results show that FT-Raman spectroscopy has clear advantage over standard dispersive Raman measurements, in particular for measurement of pollen samples with high pigment content. FT-Raman spectra are strongly biased toward chemical composition of pollen wall constituents, namely sporopollenins and pigments. This makes Raman spectra complementary to FTIR spectra, which over-represent chemical constituents of the grain interior, such as lipids and carbohydrates. The results show a large variability in pollen chemistry for families, genera and even congeneric species, revealing wide range of reproductive strategies, from storage of nutrients to variation in carotenoids and phenylpropanoids. The information on pollen's chemical patterns for major plant taxa should be of outstanding value for various studies in plant biology and ecology, including aerobiology, palaeoecology, forensics, community ecology, plant-pollinator interactions, and climate effects on plants.

Keywords: Raman spectroscopy, Fourier transform infrared spectroscopy, multivariate analysis, male gametophyte, flowering, pollen wall, pollination, palynology

INTRODUCTION

Pollen is multicellular haploid gametophyte life stage of seed plants (spermatophytes) and thus it has a key function in plant life cycle. Due to their high mobility by abiotic and biotic pollination vectors, pollen play an essential role in the gene flow within and among plant populations. Therefore, pollen studies are important for assessment of environment, including biodiversity, plant–pollinator interactions, and impact of climate and pollution on wild communities and cultivated crops. Moreover, pollen is seasonal air pollutant that can trigger allergy-related respiratory diseases, and thus pollen monitoring is needed for avoidance and timely treatment of symptoms. Finally, fossil pollen grains are often the most abundant and the best preserved remains of plant species, thus providing crucial information for the reconstruction of past terrestrial communities and climate conditions (Lindbladh et al., 2002; Jardine et al., 2016). In general, pollen studies can provide information on spatial and temporal distribution of organisms and populations, as well as on the biological and environmental processes influencing them. As a result, pollen studies have been extensively conducted in biology, ecology, palaeoecology, medicine, agronomy, and forensics.

Most of the pollen studies are focused on a quite limited number of traits, such as pollen morphology, pollen production per flower, pollen transfer, pollinator attraction, and pollen viability (Bell, 1959; Bassani et al., 1994; Molina et al., 1996; Pacini et al., 1997; Streiff et al., 1999; Tamura and Kudo, 2000; Oddou-Muratorio et al., 2005; Hall and Walter, 2011; Welsford et al., 2016). The most important reason for such deficiency of data, compared to information on female traits, is relative difficulty of quantitative and qualitative measurements of male traits (Williams and Mazer, 2016). In particular, pollen chemical composition has been rarely measured due to complex and time-consuming analyses. For example, triglyceride lipids (triacylglycerols) primarily serve as carbon and long-term energy reserves in a form of lipid bodies that play a crucial role, as a source of materials and energy, in germination of pollen as well as in pollen tube growth (Piffanelli et al., 1998; Rodriguez-Garcia et al., 2003). This is of importance since the reproduction of seed plants involves competition among growing pollen tubes to reach and penetrate the ovule. Carbohydrates, in the form of cytoplasmic saccharides, have a vital function in the resistance of pollen to dehydration and temperature stress, as well as serving as grain wall components (cellulose) and energy reserves for germination (starch and sucrose) (Pacini et al., 2006; Bokszczanin et al., 2013). Pollen's proteins have both structural and functional role, and have implication for both pollen-pistil and plant-pollinator interactions (Roulston et al., 2000). Pollen proteins are important source of dietary nitrogen for a majority of pollinators, while as enzymes they have crucial function during pollen tube growth (Roulston et al., 2000). Pigments, such as carotenoids and flavonoids, participate in light harvesting, serve as cellular membrane protectants from photooxidative damage, and as pathogen defense (Fambrini et al., 2010; Lutz, 2010). Finally, sporopollenins are complex and resilient grain wall biopolymers that protect the grain interior from environmental effects (Li et al., 2019), and can

provide valuable information on past environmental conditions (Lomax et al., 2012).

Vibrational spectroscopy of pollen offers a novel approach in plant phenomics via precise and comprehensive measurement of pollen's biochemical 'fingerprint.' Vibrational spectra of pollen contain specific signals of lipids, proteins, carbohydrates and water, and even some minor biochemical constituents, such as pigments, can be precisely measured (Schulte et al., 2008, 2009; Zimmermann, 2010; Pummer et al., 2013; Zimmermann and Kohler, 2014; Bagcioglu et al., 2015; Jardine et al., 2015). The compounds that are measurable by vibrational spectroscopy are the principal structural and nutritious components, and they are responsible for the majority of chemical phenotypic attributes of pollen. Therefore, chemical analysis of pollen by vibrational spectroscopy offers complementary information to the contemporary 'omics-based' approaches, such as genomics and transcriptomics by sequencing technologies, as well as to proteomics and metabolomics by mass spectrometry and NMR spectroscopy.

Vibrational studies of pollen, by diverse infrared and Raman techniques, have shown that vibrational spectroscopy achieves economical and rapid identification and classification of pollen according to taxonomy and phylogenetic relationship (Pappas et al., 2003; Ivleva et al., 2005; Gottardini et al., 2007; Schulte et al., 2008; Dell'anna et al., 2009; Zimmermann, 2010, 2018; Guedes et al., 2014; Zimmermann and Kohler, 2014; Zimmermann et al., 2015b; Julier et al., 2016; Seifert et al., 2016; Woutersen et al., 2018; Jardine et al., 2019; Mondol et al., 2019). For example, the recent FTIR microspectroscopy study on individual pollen grains has achieved more accurate classification than optical microscopy, which is the benchmark method in pollen identification (Zimmermann et al., 2016). Moreover, FTIR microspectroscopy enables chemical imaging of pollen grain ultrastructure (Zimmermann et al., 2015a), while even higher spatial resolution of Raman microspectroscopy enables monitoring of the molecular composition during pollen germination and pollen tube growth (Schulte et al., 2010; Joester et al., 2017). In addition to classification and identification studies, vibrational spectroscopies provide biochemical characterization of pollen with respect to environmental conditions. For example, differences in chemical phenotypes of pollen were measured with respect to nutrient availability (Zimmermann et al., 2017), heat stress (Lahlali et al., 2014; Jiang et al., 2015), pollution stress (Depciuch et al., 2016, 2017), location (Bağcıoğlu et al., 2017; Zimmermann et al., 2017), and season (Zimmermann and Kohler, 2014; Bağcıoğlu et al., 2017). In general, Raman and Fourier transform infrared (FTIR) spectroscopies provide chemically complementary information, and therefore measurement of samples by both techniques provides highly detailed biochemical characterization (Zimmermann, 2010; Pummer et al., 2013; Bagcioglu et al., 2015; Zimmermann et al., 2015a; Diehn et al., 2020).

One important advantage of Raman spectroscopy of pollen over FTIR approach is the obtained information on grain wall pigments, in particular carotenoids, which cannot be measured at all by FTIR due to low concentrations and weak signals (Schulte et al., 2009). Unfortunately, Raman measurements

are often hindered by laser-induced degradation (burning) of pollen grains, and by the strong fluorescence background that often masks any underlying Raman spectra (Ivleva et al., 2005; Schulte et al., 2008; Guedes et al., 2014). An interesting aspect of Raman spectroscopy is that pigments, such as carotenoids and chlorophylls, exhibit resonance Raman spectra of various intensities (De Oliveira et al., 2009). Resonance Raman spectra occur when the wavelength of the excitation laser coincides with electronic transition. For example, the conjugated nature of π-electrons from the polyene backbone of carotenoids results in electronic states of lower energy, often with absorption in the visible part of the spectrum. For this reason, carotenoids usually display strong yellow, orange and red colors. Moreover, this can cause strong enhancement of vibrational bands in carotenoids, especially those at 1530 (related to $-C=C-$ bonds) and 1160 cm^{-1} (related to $-C-C-$ bonds) that have strong electron-phonon coupling. Resonance Raman spectra of pigments enable measurement of very low concentration of pigments (Schulte et al., 2009). However, resonant Raman spectra can also mask completely the regular Raman spectral contributions from other compounds, thus hindering their analysis (Schulte et al., 2008). Use of visible excitation lasers, such as 633 nm, often results with strong light absorbance of pollen sample, leading to sample heating and even photodegradation (Schulte et al., 2008, 2009).

These problems, that are common in conventional (dispersive) Raman measurements, can be addressed by Fourier transform (FT) Raman spectroscopy that uses high-wavelength near infrared (NIR) laser excitation. In general, electronic transitions are weaker at longer wavelengths, and thus detrimental effect of sample heating can be avoided by use of NIR lasers. Moreover, the frequency of the NIR laser usually does not correspond to an electronic transition of the sample, thus diminishing possibility for the occurrence of fluorescence. Finally, use of longer Raman excitation wavelengths can significantly increase penetration depth, compared to short-wavelength lasers, thus more comprehensive information on pollen composition could be obtained (Moester et al., 2019). However, these important advantages of FT-Raman spectrometers can be overshadowed by sensitivity advantage of dispersive Raman spectrometers with short-wavelength laser excitation. Nevertheless, our preliminary study, employing FT-Raman spectroscopy for measurement of 43 conifer species, has shown that high-quality FT-Raman spectra of pollen can be recorded (Zimmermann, 2010). The spectra were devoid of detrimental fluorescence and heating effects, thus indicating great potential of FT-Raman spectroscopy for identification and analysis of plants.

In the paper at hand, we explore the use of FT-Raman and FTIR spectroscopy for chemical characterization of pollen. The study was conducted on a diverse set of plants, comprising 219 species, belonging to 42 families, and covering all major taxa of seed plants. Firstly, we wanted to demonstrate that high-quality Raman spectra of pollen can be obtained by FT-Raman technique. In particular, Raman spectra of pollen samples with high pigment content, which regularly cannot be measured intact with dispersive Raman, were obtained by FT-Raman spectroscopy. Secondly, the study highlights unique pollen chemistry information obtained by either FT-Raman or FTIR approach, and thus demonstrates advantages of the combined approach with both techniques. The biochemical characterization of pollen for the major plant lineages is provided, in particular regarding lipid, protein, carbohydrate, carotenoid and phenylpropanoid content. Pollen chemistry, obtained by the spectroscopic approach, was discussed in relation to the results of the standard chemical analysis pollen studies. Moreover, quantitative measurement of pollen protein content has been provided by combining spectroscopy and chemometrics, clearly demonstrating potential of vibrational spectroscopy for not only qualitative but also quantitative chemical analysis of pollen.

MATERIALS AND METHODS

Samples

Samples of pollen were collected at two facilities of the University of Zagreb; the Botanical Garden of the Faculty of Science and the Botanical Garden "Fran Kušan" of the Faculty of Pharmacy and Biochemistry. Both locations are situated within 1.5 km radius and can be considered the same climate area. 219 samples were collected altogether, each belonging to different plant species (**Table 1**, **Supplementary Table S1**, and **Supplementary Figure S1**). Pollen samples were collected during 2011 and 2012 pollination seasons. The pollen samples were collected directly from plants at flowering time, either by shaking flowers (anemophilous species) or collecting mature anthers (entomophilous species). Only one sample per species was created, either by collecting pollen from only one plant or by collecting pollen from several individuals of the same species followed by merging of all the collected pollen into one sample. The samples were kept in paper bags at r.t. for 24 h (together with anthers for entomophilous species), and afterward transferred to vials as dry powder and stored at $-15°C$. For the spectroscopic measurements, three replicates per technique were measured, each replicate comprising approx. 0.5–1.0 mg of pollen sample. Approx. 10^3–10^5 pollen grains per replicate were measured, and considering that each pollen grain has a unique genotype (and implicitly phenotype), each measurement comprised biologically distinct pollen population. However, in a majority of cases (for example for all tree species) genetic pool was very limited since it originated from the same sporophyte parent plant. Therefore, the presented variation in the spectral sets can be considered a preliminary estimate for the measured plant groups. Influence of larger genetic pools, growth conditions, location and year of pollination on pollen chemical composition and pollen classification for a number of plant groups (e.g., grasses, pines, and oaks) were covered in our previous studies (Zimmermann and Kohler, 2014; Bağcıoğlu et al., 2017; Zimmermann et al., 2017; Diehn et al., 2020).

For identification of basic biochemicals in pollen a set of model compounds was measured to correlate with high positive or negative values in the principal component analyses loadings plots. Spectra of crystal lipids and carbohydrates were recorded above their melting temperature, and again at r.t. after cooling to obtain spectrum of amorphous

TABLE 1 | List of analyzed plant taxa with number of genera and species covered by the study (see **Supplementary Table S1** for details).

Clade	Order	Family	No. of genera	No. of species
Eudicots	Fagales	Betulaceae	5	13
		Fagaceae	2	11
		Juglandaceae	3	6
	Malpighiales	Salicaceae	2	3
		Euphorbiaceae	1	1
	Lamiales	Oleaceae	1	3
		Plantaginaceae	2	4
		Scrophulariaceae	1	1
		Acanthaceae	1	1
	Solanales	Solanaceae	3	4
	Proteales	Platanaceae	1	2
	Saxifragales	Altingiaceae	1	2
		Paeoniaceae	1	2
	Sapindales	Sapindaceae	1	1
		Anacardiaceae	1	2
		Rutaceae	1	1
	Dipsacales	Adoxaceae	2	3
		Dipsacaceae	1	1
	Buxales	Buxaceae	1	2
	Rosales	Rosaceae	4	5
		Urticaceae	1	1
	Ranunculales	Ranunculaceae	5	10
		Papaveraceae	3	5
	Caryophyllales	Polygonaceae	1	4
	Asterales	Asteraceae	2	3
		Campanulaceae	1	1
	Malvales	Malvaceae	3	3
Magnoliids	Magnoliales	Magnoliaceae	2	2
Monocots	Poales	Cyperaceae	3	11
		Poaceae	9	18
		Juncaceae	1	2
	Asparagales	Iridaceae	1	17
		Xanthorrhoeaceae	2	4
	Liliales	Liliaceae	2	10
	Arecales	Arecaceae	1	1
Gymnosperms	Pinales	Cupressaceae	11	23
		Taxaceae	1	2
		Cephalotaxaceae	2	2
		Pinaceae	5	29
		Podocarpaceae	1	1
	Ginkgoales	Ginkgoaceae	1	1
	Ephedrales	Ephedraceae	1	1

phase (liquid and/or glass phase). Lipids: Tristearin (2,3-di(octadecanoyloxy)propyl octadecanoate), triolein (2,3-bis[[(*Z*)-octadec-9-enoyl]oxy]propyl (*Z*)-octadec-9-enoate), tri-heptadecanoin (2,3-di(heptadecanoyloxy)propyl heptadeca-noate), phosphatidistearoylcholine (1,2-distearoyl-rac-glycero-3-phosphocholine), phosphatidioleylcholine (1,2-dioleoyl-sn-glycero-3-phosphocholine), stearic acid (octadecanoic acid), oleic acid ((9Z)-octadec- 9-enoic acid). Pigments

and phenylpropanoids: rutin, β-carotene, *p*-coumaric acid, ferulic acid, caffeic acid, sinapic acid, hydro-*p*-coumaric acid, hydroferulic acid, hydrocaffeic acid. Carbohydrates: cellulose, amylose, amylopectin, arabinoxylan, pectin, β-D-glucan, sucrose, trehalose, fructose, glucose. Proteins: gluten. All chemicals were purchased from Merck (Darmstadt, Germany) and Sigma-Aldrich (St. Louis, United States), and used without further purification.

Spectroscopic Analyses

The Raman spectra in backscattering geometry were recorded on a FT-Raman FRA 106/S model, coupled with a Bruker Equinox 55 IR spectrometer, equipped with a neodymium-doped yttrium aluminum garnet (Nd:YAG) laser (1064 nm, 9394 cm^{-1}), and germanium detector cooled with liquid nitrogen. The spectra were recorded with a resolution of 4 cm^{-1}, with a digital resolution of 1.9 cm^{-1}, and with a total of 64 scans, using Blackman–Harris 4- term apodization and with a laser power of 400 mW. Each pollen sample was measured in three replicates.

The infrared spectra were recorded on an ABB Bomem (Quebec City, Canada) MB102 single-beam spectrometer, equipped with cesium iodide optics and deuterated triglycine sulfate (DTGS) detector. The reflectance spectra were recorded by using the single-reflection attenuated total reflectance (SR-ATR) accessory with the horizontal diamond prism and with 45° angle of incidence. The SR-ATR infrared spectra were measured with a Specac (Slough, United Kingdom) Golden Gate ATR Mk II or a Specac High Temperature Golden Gate ATR Mk II. The spectra were recorded with a spectral resolution of 4 cm^{-1}, with a digital resolution of 1.9 cm^{-1}, and with a total of 30 scans, using cosine apodization. Each spectrum was recorded as the ratio of the sample spectrum to the spectrum of the empty ATR plate. Each pollen sample was measured in three replicates.

Spectral Pre-processing and Data Analysis

The spectra were pre-processed prior to calibration: all spectra were smoothed by the Savitzky–Golay algorithm using a polynomial of degree two and a window size of 11 points in total, followed by normalization by extended multiplicative signal correction (EMSC), an MSC model extended by a linear and quadratic component (Zimmermann and Kohler, 2013; Guo et al., 2018). The following spectral regions were selected for data analysis: 1900-800 cm^{-1} for infrared spectra, and 2000-500 cm^{-1} for Raman spectra. In the EMSC pre-processing, the spectral region of chemical absorbance was down-weighted, and spectral regions devoid of any chemical absorbance were up-weighted, by applying a weighting vector. Vector value 1 was used in the whole spectral region, except the regions 1900-1800 cm^{-1} (for IR spectra) and 2000-1800 cm^{-1} (for Raman spectra), where the weighting vector was set to 10. These two regions are devoid of any chemical signals. Therefore, these regions should have the same baseline values in all pollen spectra when interferent signals, due to light reflection or fluorescence, have been removed. Up-weighting of this region, by applying a weighting vector, is constraining the EMSC pre-processing

and ensuring a stable baseline in all pollen spectra. Thus pre-processed spectra were designated *Datasets I* and subsequently used to evaluate biochemical similarities between pollen samples by calculating correlation coefficients (Pearson product-moment correlation coefficients) or by using principal component analysis (PCA). For better viewing the figures depicting correlation matrixes and PCA plots were based on averaged spectra, where spectra of 3 replicates were averaged.

The estimates of relative chemical composition of pollen were obtained by deflating the data matrix containing complete set of spectra from *Datasets I* by using spectra of standard compounds (Zimmermann and Kohler, 2014). In general, matrix deflation modifies a data matrix to eliminate the influence of a given eigenvector (White, 1958). However, here we used model compounds as eigenvectors while the corresponding eigenvalues were used to estimate the relative content of those compounds. For the deflation, the data matrix was centered while the vectors were normalized. Tristearin, gluten and amylose FTIR spectra were used as eigenvectors for FTIR dataset to estimate relative amounts of triglicerides, proteins and carbohydrates respectively. β-carotene and amylose FT-Raman spectra were used as eigenvectors for Raman dataset to estimate relative amounts of carotenoids and carbohydrates respectively. The corresponding eigenvalues, as well as ratios of eigenvalues (carbohydrate-to-protein ration), were plotted in order to visualize chemical composition of pollen.

Datasets I were used for the analysis of pollination strategy by denoting the following taxa: (1) anemophilous: Fagales (except Fagaceae), Pinales (except Podocarpaceae), Poales, Proteales, Anacardiaceae, Asteraceae (except *Taraxacum*), Polygonaceae, Urticaceae, *Plantago*; (2) entomophilous: Asparagales, Dipsacales, Liliales, Magnoliales, Malvales, Ranunculales, Sapindales (except Anacardiaceae), Solanales, Acanthaceae, Campanulaceae, Paeoniaceae, Rosaceae, Scrophulariaceae, *Digitalis*, *Taraxacum*; (3) double-strategy: Arecales, Buxales, Ephedrales, Malpighiales, Altingiaceae, Fagaceae, Ginkgoaceae, Oleaceae, Podocarpaceae.

The protein content of pollen from Roulston et al. (2000) was used as a chemical reference values for regression in the Partial Least Squares Regression (PLSR) modeling of spectral data from *Datasets I*, where spectra of replicates were averaged. The optimal number of components (i.e., PLSR factors) of the calibration models (A_{Opt}) was determined using full cross-validation. The PLSR coefficient of determination (R^2), correlation value (R), and root-mean-square error (RMSE) were used to evaluate the calibration models. The following 35 species were included in the PLSR: *Fagus sylvatica, Quercus rubra, Quercus robur, Corylus avellana, Alnus incana, Alnus glutinosa, Betula pendula, Juglans nigra, Juglans regia, Carya illinoinensis, Zea mays, Secale cereale, Festuca pratensis, Poa pratensis, Poa nemoralis, Dactylis glomerata, Holcus lanatus, Juniperus communis, Thuja occidentalis, Picea abies, Pinus mugo, Pinus sylvestris, Pinus ponderosa, Eschscholzia californica, Magnolia x sonlangiana, Liriodendron tulipifera, Fraxinus excelsior, Plantago lanceolata, Salix alba, Taraxacum officinale, Populus nigra, Aesculus hippocastanum, Buxus sempervirens, Artemisia vulgaris, Rumex acetosa.*

The following spectral regions were selected for analysis of chemical composition of aromatics in pollen grain wall: 810 – 860 cm^{-1} for FTIR spectra, and 1580 – 1650 cm^{-1} for FT-Raman spectra. Prior to the selection of spectral regions, the EMSC pre-processing was conducted by applying a weighting vector: Vector value 1 was used in the whole spectral region, except the regions 1900-1800 cm^{-1} (for IR spectra) and 1800-1660 cm^{-1} (for FT-Raman spectra), where the weighting vector was set to 10. Thus pre-processed spectra were designated *Datasets II* and subsequently analyzed by PCA.

All pre-processing methods and data analyses were performed using The Unscrambler X 10.3 (CAMO Software, Oslo, Norway), as well as functions and in-house developed routines written in MATLAB 2014a.8.3.0.532 (The MathWorks, Natick, MA, United States).

RESULTS AND DISCUSSION

Vibrational Spectra of Pollen

As mentioned in the Introduction, the major problem in Raman spectroscopy of pollen is sample heating and fluorescence, resulting with complex background and low signal-to-noise ratio. In contrast, the spectra of all 219 pollen samples covered by this study are devoid of strong fluorescence background and have high signal-to-noise ratio (**Supplementary Figure S2**). The vibrational spectra of pollen of representative species show influence of different biochemicals on an overall spectral fingerprint (**Figure 1**) (Gottardini et al., 2007; Schulte et al., 2008; Zimmermann, 2010; Bagcioglu et al., 2015; Zimmermann et al., 2015b). In some species, such as gymnosperms (e.g., **Figure 1** *Pinus ponderosa* and *Ephedra major*), the most prominent features are phenylpropanoid-associated signals of sporopollenins around 1630, 1605, 1585, 1205, 1170, 855, and 830 cm^{-1} in the Raman spectra, and around 1605, 1515, 1205, 1170, 855, 830 and 815 cm^{-} in the FTIR spectra[1] (all vibrations are related to phenyl ring vibrations). Furthermore, some taxa, such as grasses (e.g., **Figure 1** *Festuca amethystina*) and sedges, have strong carbohydrate signals around 1450-1300 (CH$_2$ and CH deformations) and 1150-900 cm^{-1} (C−O−C, C−C and C−O stretching vibrations) in the Raman spectra, and around 1200-900 cm^{-1} (C−O, C−C, C−O−C, and C−OH stretches and deformations) in the FTIR spectra. Signals related to lipids (e.g., **Figure 1** *Fagus sylvatica*), around 1745 (C=O stretch), 1440 and 1300 (CH$_2$ deformation), and 1070 cm^{-1} (C−C stretch) in the Raman, and around 1745 (C=O stretch), 1460 (CH$_2$ deformation) and 1165 cm^{-1} (C−O−C stretching in esters) in the FTIR spectra, often show large variation within related plant species. All species show prominent protein signals around 1655 (amide I), 1450 (CH$_2$ deformation), and 1260 cm^{-1} (amide III) in the Raman spectra, and around 1645 (amide I), 1535 (amide II), and 1445 cm^{-1} (CH$_2$ deformation) in the FTIR spectra, with taxon-specific ratio of protein-to-carbohydrate signals. In addition to these signals, a number of species (e.g., **Figure 1**, *Lilium bulbiferum*) show carotenoid-associated strong Raman signals around 1520 (C=C stretching), 1155 (C-C stretching), and 1005 cm^{-1} (C-CH$_3$
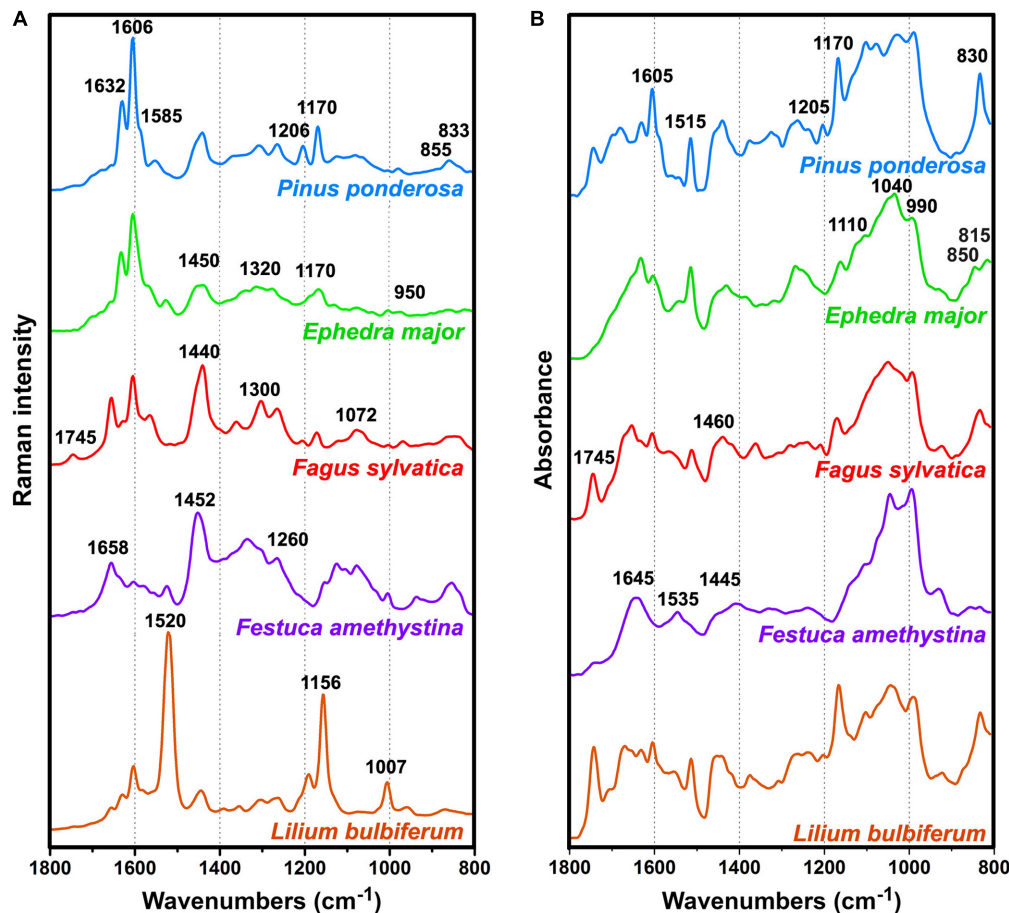
FIGURE 1 | Pre-processed **(A)** FT-Raman and **(B)** FTIR spectra of representative species: *Pinus ponderosa* (with denoted sporopollenin bands), *Ephedra major* (with denoted carbohydrate signals), *Fagus sylvatica* (with denoted lipid signals), *Festuca amethystina* (with denoted protein signals), and *Lilium bulbiferum* (with denoted carotenoid signals).

deformation) (Schulte et al., 2009). Carotenoids are present in low concentration in pollen, and thus they cannot be detected by FTIR spectroscopy. However, resonant Raman effect enables their measurement by FT-Raman spectroscopy, which sometimes results with the complete dominance of these carotenoid signals over spectral contributions of other biochemicals, such as proteins, sporopollenins and carbohydrates (e.g., **Figure 1**, *Lilium bulbiferum*).

Compared to the published results of Raman measurements with 633 nm laser excitation (see **Supplementary Figure S1** in Guedes et al., 2014), the FT-Raman spectra have rather simple background, which can be easily corrected with EMSC pre-processing (**Supplementary Figure S2**). Moreover, compared to the published results of Raman measurements with 785 nm laser excitation (Schulte et al., 2008), the FT-Raman spectra show significantly weaker resonant Raman effect. For example, the spectrum of *Aesculus hippocastanum* excited with 785 nm shows predominant carotenoid bands at 1518 and 1156 cm^{-1}, while strong signals of proteins, carbohydrates and sporopollenins were recorded only after the prolonged photodestruction of the sample with 633 nm laser (see Figure 1 in Schulte et al., 2008). On the

other hand, the FT-Raman spectrum of *Aesculus hippocastanum* has weak carotenoid signals and strong signals of proteins, carbohydrates and sporopollenins (**Supplementary Figure S3**).

Overall Assessment of Pollen Composition

The large and extremely diverse set of measured species, covering 42 plant families, has enabled assessment of major biochemical differences and similarities between pollen species. The correlation coefficients of spectra were calculated in order to assess major patterns within and between taxa. In addition, PCA was used to estimate predominant spectral differences, and indirectly to assess principal differences in chemical composition of pollen.

The matrices of correlation coefficients (**Figure 2** and **Supplementary Figure S4**) show that plant families have relatively uniform and specific pollen composition, specifically that pollen of related species share common chemical features. Such property is apparent, for example, for pollens of Pinaceae, Cupressaceae, Fagaceae, Betulaceae, Poaceae and Cyperaceae.
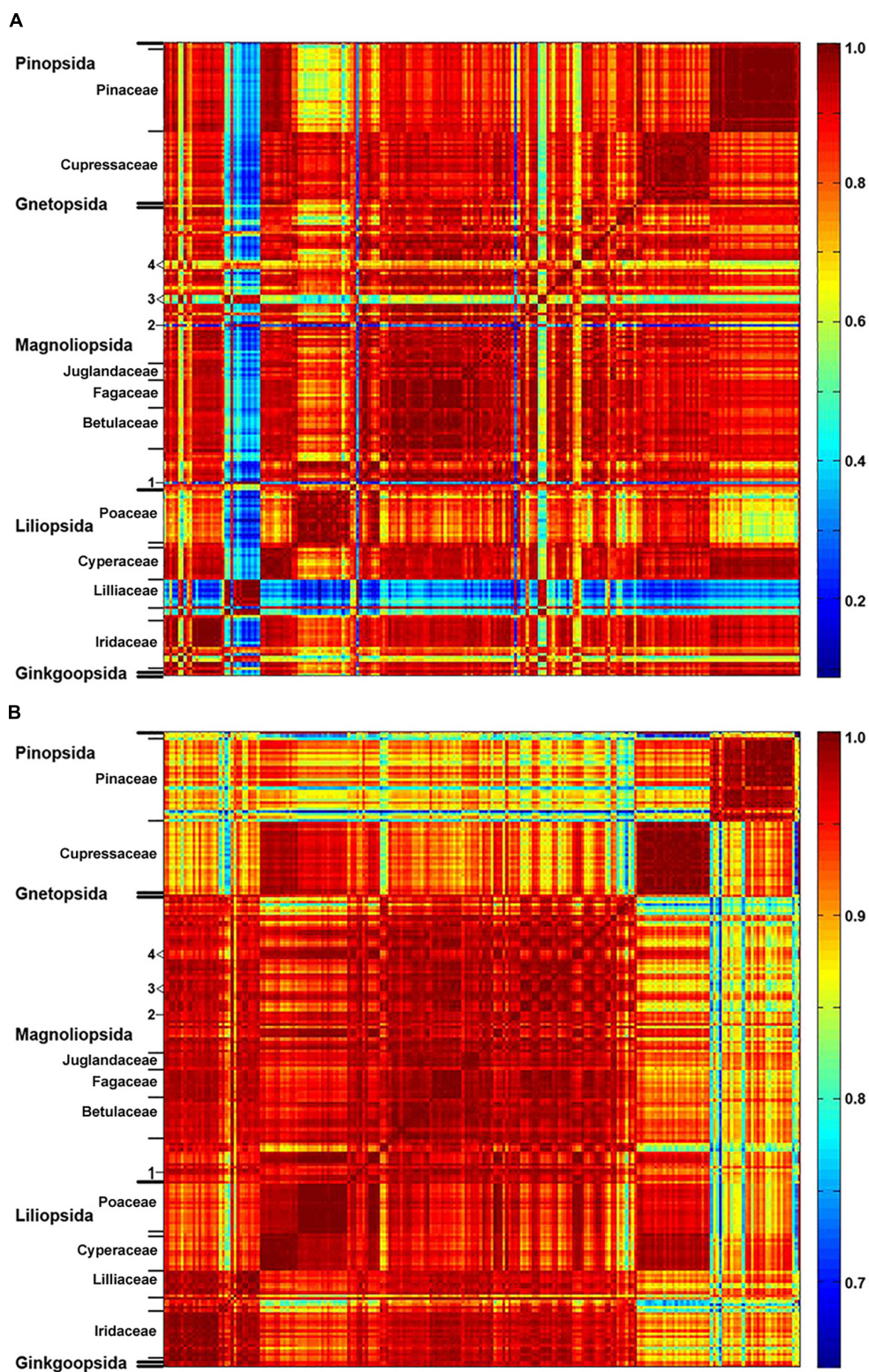
FIGURE 2 | Correlation between spectroscopic data and taxonomy. Matrices of correlation coefficients calculated from: **(A)** FT-Raman and **(B)** FTIR spectra of 219 species (*Dataset I*, average spectra of 3 replicates), with depiction of plant classes and families (in addition: 1, *Taraxacum officinale*; 2, *Hibiscus trionum*; 3, *Eschscholzia californica*, *Glaucium flavum*, *Papaver lapponicum*; 4, *Ranunculus repens*, *Ranunculus acris*, *Ranunculus lanuginosus*).

However, FT-Raman spectral set has higher spectral variability than infrared set, as shown by the larger range of correlation coefficients, with a number of taxa showing specific spectral patterns (**Figure 2A**). For example, FT-Raman spectra of Liliaceae (**Figure 2A**) are extremely different compared to spectra of a majority of angiosperms (including other monocots), while the corresponding infrared spectra (**Figure 2B**) show considerably lower level of variability. Moreover, a number of species show specific FT-Raman spectral fingerprints, such as large variations for congeneric species of *Iris* and *Papaver* (see Iridaceae and number 3 markings in **Figure 2A**). In all the cases the large spectral variability within the FT-Raman data set is driven by the strong Raman signals of the carotenoids that overshadow spectral contributions from other chemicals, as illustrated previously by the spectrum of *Lilium bulbiferum* (**Figure 1A**).

The PCA of FT-Raman data shows that the predominant spectral differences are the result of variations of bands associated with carotenoids, sporopollenins, carbohydrates and proteins (**Figures 3, 4**). The PCA plots have high factor loadings associated with carotenoids (positive loadings) at 1523, 1155, and 1005 cm^{-1}, and proteins (negative) at 1655, 1455, and 1260 cm^{-1} in PC 1, and sporopollenins at 1630, 1605, 1585, 1205, and 1170 cm^{-1}, and proteins (negative) at 1650 and 1453 cm^{-1} in PC 2 (**Figure 4**). Therefore, it is evident that the predominant information from FT-Raman spectral data is on pollen grain wall chemicals. The PCA score plot in **Figure 3** indicates scores for the selected plant families with relatively high number of species represented in the data set. Similar as the matrices of correlation coefficients (**Figure 2** and **Supplementary Figure S4**), the score plot shows that the majority of Liliaceae, as well as the number of Iridaceae species, have quite different pollen chemistry (in particular, carotenoid content) compared to the rest of measured pollen species. All the major families, apart from Iridaceae, show relatively good clustering, indicating taxon-specific chemistry. For example, the separation of relatively related clades Poaceae and Cyperaceae (both Poales), as well as Pinaceae and Cupressaceae (both Pinales), is mostly driven by the difference in their sporopollenin and carbohydrate content and composition.

Furthermore, the PCA score plot of PCs 1 and 4 indicates a trend in pollen chemistry composition based on pollination mode (**Figure 3**). Anemophilous (wind pollinated) and entomophilous (insect pollinated) species show different tendencies based on relative content of protein, carbohydrates and carotenoids. The PCA loading plots have high factor loadings associated with carotenoids and proteins in PC 1, and proteins (positive) at 1655 and 1450 cm^{-1}, and carbohydrates (negative) at 1450-1300 and 1150-1050 cm^{-1} in PC 4 (**Figure 4**). In general, anemophilous species have low content of carotenoids and proteins, and high content of carbohydrates, compared to entomophilous species. This is in agreement with the published studies, showing that insect foragers prefer plants with high-protein pollen content (Roulston et al., 2000), while anemophilous species produce pollen with high carbohydrate content (Speranza et al., 1997; Wang et al., 2004). The evolutionary explanation is that production of proteins has higher metabolic cost, and

thus plants with non-rewarding pollen (anemophilous, self-pollinators, and nectar rewarders) have pollen with higher carbohydrate content. Although the spectroscopy results are in agreement with previous studies on pollination mode and pollen chemistry, it is possible that the trends present in **Figure 3** are driven by plant relatedness, specifically that pollen of related species share common chemical features. Therefore, further studies are needed, preferably on a group of closely related species presenting different pollination modes.

The PCA of the FTIR data shows that the predominant spectral differences are the result of variations of bands associated with proteins, carbohydrates and sporopollenins (**Figure 5**). The PC loading plots have high factor loadings associated with proteins (negative loadings) at 1650 and 1540 cm^{-1}, carbohydrates (positive) at 1050-950 cm^{-1} and lipids (positive) at 1745 and 1165 cm^{-1} in PC 1, and carbohydrates (negative) at 1050-950 cm^{-1}, sporopollenins (positive) at 1605, 1515, 1170 and 833 cm^{-1}, and lipids (positive) at 1745 and 1165 cm^{-1} in PC 2. The PCA score plot in **Figure 5** indicates scores for the selected plant families highlighted in the PCA score plot of Raman data (**Figure 3**). Similar as for the Raman data, the major plant families show taxon-specific clustering. For example, analogous to the Raman data set, the separation of relatively related clades Poaceae and Cyperaceae (both Poales), as well as Pinaceae and Cupressaceae (both Pinales), is mostly driven by the difference in their sporopollenin and carbohydrate content and composition. This issue has already been mentioned in our previous studies (Zimmermann and Kohler, 2014), and it will be discussed in more details later in this paper. The main difference between the FTIR and FT-Raman data is lack of the carotenoid-driven outliers in the FTIR that were present in the Raman data (in particular, Liliaceae). Another difference is relative large variation in the FTIR data driven by the lipid content, which was mostly lacking in the FT-Raman data. The issue of carotenoid and lipid content will be tackled in more details below when we discuss strategies for quantification of relative chemical composition of pollen.

Relative Chemical Composition of Pollen

A primary drawback of PCA of vibrational spectral data is its reduced interpretability due to complex loadings. Therefore, the data matrices containing complete set of spectra were deflated by using spectra of standard compounds (Zimmermann and Kohler, 2014). That way, estimates of chemical composition of pollen were obtained regarding principal type of compounds: triglyceride lipids, proteins, carbohydrates and carotenoids. It should be noted that this procedure is not a replacement for aforementioned PCA, particularly if residual spectral component is large and contains important variability information. However, the obtained eigenvalues can be good proxies for estimating relative chemical composition of pollen and for simple visualization of pollen composition.

Figures 6, 7 show that relative composition of pollen has big variations regarding carotenoids and triglycerides. These type of compounds show substantial variations even for congeneric species. For example, a number of far-related genera, such as *Quercus*, *Iris*, *Pinus*, and *Juniperus* show large variations for congeneric species regarding triglycerides, as well as *Papaver*,

FIGURE 3 | Correlation between FT-Raman data and taxonomy. PCA score plots of FT-Raman dataset containing 219 species (*Dataset I*, average spectra of 3 replicates) with: **(A)** depiction of plant classes (■ monocots; ♦ gymnosperms; ● eudicots and magnoliids) and families (Pinaceae, Cupressaceae, Betulaceae, Fagaceae, Poaceae, Cyperaceae, Liliaceae, and Iridaceae), and **(B)** depiction of pollination strategies (blue ■ entomophilous; red ● anemophilous; purple ♦ double-strategy) species. Vectors are approximating the increase in relative amount of proteins (P), sporopollenins (S), carbohydrates (C) and carotenoids (Cr). The percent variances for the first five PCs are 54.77, 25.08, 6.43, 4.41, and 2.31.

Lilium and *Iris* regarding carotenoids (**Supplementary Figures S5, S6**, respectively). On the other hand, relative composition regarding carbohydrates and proteins is quite taxa-specific. These two type of compounds show negative linear correlation in the FTIR dataset (**Supplementary Figure S7**). It has been known that the pollen protein content is very similar for congeneric species, and it can be even similar for confamiliar species (Roulston et al., 2000). Based on this and the aforementioned linear correlation between pollen carbohydrates and proteins, it can be concluded that pollen carbohydrate content is highly conserved within genera and families as well. Moreover, there

is a clear trend in chemical composition for anemophilous and entomophilous species, which is consistent with FT-Raman results in **Figure 3** and our previous finding (Zimmermann and Kohler, 2014). Pollen of anemophilous plants have much higher relative content of carbohydrates, defined as carbohydrate-to-protein ratio, as compared with entomophilous plants (**Supplementary Figure S7**).

These results are in agreement with the published studies. As mentioned previously, insect pollinated species have in general higher protein content and lower carbohydrate content than anemophilous plants (Speranza et al., 1997; Roulston

FIGURE 4 | Loadings plots on the principal components 1, 2 and 4 of FT-Raman dataset containing 219 species (Dataset I, average spectra of 3 measurements).

et al., 2000). Similarly to our results, the study by Roulston et al. (2000), has revealed that anemophilous pollens contain significantly less protein (average value 25.8% protein dry mass content) than zoophilous (animal pollinated) pollens (average value 39.3%). However, the authors have stipulated that this discrepancy could arise from a sampling and measurement bias. They have stated that due to analytical limitations and relative ease of collecting anemophilous pollens, anemophilous species are always overrepresented in the data set. The main reason is that standard analyses of pollen protein content requires 1-1000 mg of pollen (Roulston et al., 2000), thus favoring anemophilous plants that produce large quantities of pollen. Although in our vibrational study the sampling set was relatively balanced regarding number of anemophilous and entomophilous species, the sampling bias cannot be entirely disregarded. It

should be noted that vibrational microspectroscopy can measure single pollen grains (Zimmermann et al., 2016), and therefore quantification of pollen proteins by spectroscopy approach would be equally applicable to anemophilous and zoophilous pollen.

Quantitative Measurement of Pollen Protein Content

The exploratory data analyses, as the ones presented above, are offering valuable information on relative chemical composition of pollen, as well as on chemical differences within and between taxa. However, the next important question to address is whether the vibrational data on pollen contains valuable quantitative biochemical information to allow the prediction of absolute chemical composition. Quantitative chemical analysis of complex biological samples, such as composition of biomass or biofluids, is readily obtained by combining vibrational spectroscopy with multivariate regression, such as PLSR (Zimmermann and Kohler, 2013; Kosa et al., 2017). Therefore, we have conducted PLSR analyses on FT-Raman and FTIR datasets for predicting protein mass fraction of pollen (percentage of protein by dry mass). PLSR models were validated, using the full cross validation method, against protein mass fraction values for 35 species obtained from Roulston et al. (2000). The analyzed species include all major taxa of seed plants (see **Supplementary Table S2**), and had extensive range of protein content, from 8.8 to 43.1% of protein content by dry mass.

The R^2 values for the PLSR models were 0.53 and 0.49 for FT-Raman and FTIR models respectively, with RMSE errors of approx. 15% (**Table 2**). PLSR regression coefficients are summarizing the relationship between spectral variables and protein mass fraction values. As can be seen, the spectral features associated with proteins are present in the regression coefficients at 1640-65 cm^{-1} (amide I), 1452 cm^{-1} (CH_2 deformation) and 1006 cm^{-1} (phenylalanine sidechain vibrations) for FT-Raman dataset, and at 1630-1670 cm^{-1} (amide I) and 1515-1560 cm^{-1} (amide II) for FTIR dataset (**Supplementary Figure S8**). FT-Raman model was based on a larger number of components (A_{opt} = 12) than the model based on FTIR data (A_{opt} = 6). This is probably due to relatively strong protein-related signals in FTIR spectra, compared to FT-Raman spectra where protein signals are often overlapped by stronger signals associated with sporopollenins and carotenoids. It should be noted that the reference data for the PLSR models was based on literature values, and not on an actual measurements of studied samples. It can be assumed that prediction models will improve when actual protein reference values for measured samples are used, and when they are restricted to phylogenetically related taxa, for example plant orders and families. Moreover, there is a great potential of vibrational spectroscopy for direct measurement of not only protein content of pollen, but other constituents, such as carbohydrates and carotenoids, as well.

Chemical Composition of Phenylpropanoids in Pollen Grain Wall

Pollen wall is extremely resilient structure, both physically and chemically, protecting generative cells from environmental stress,
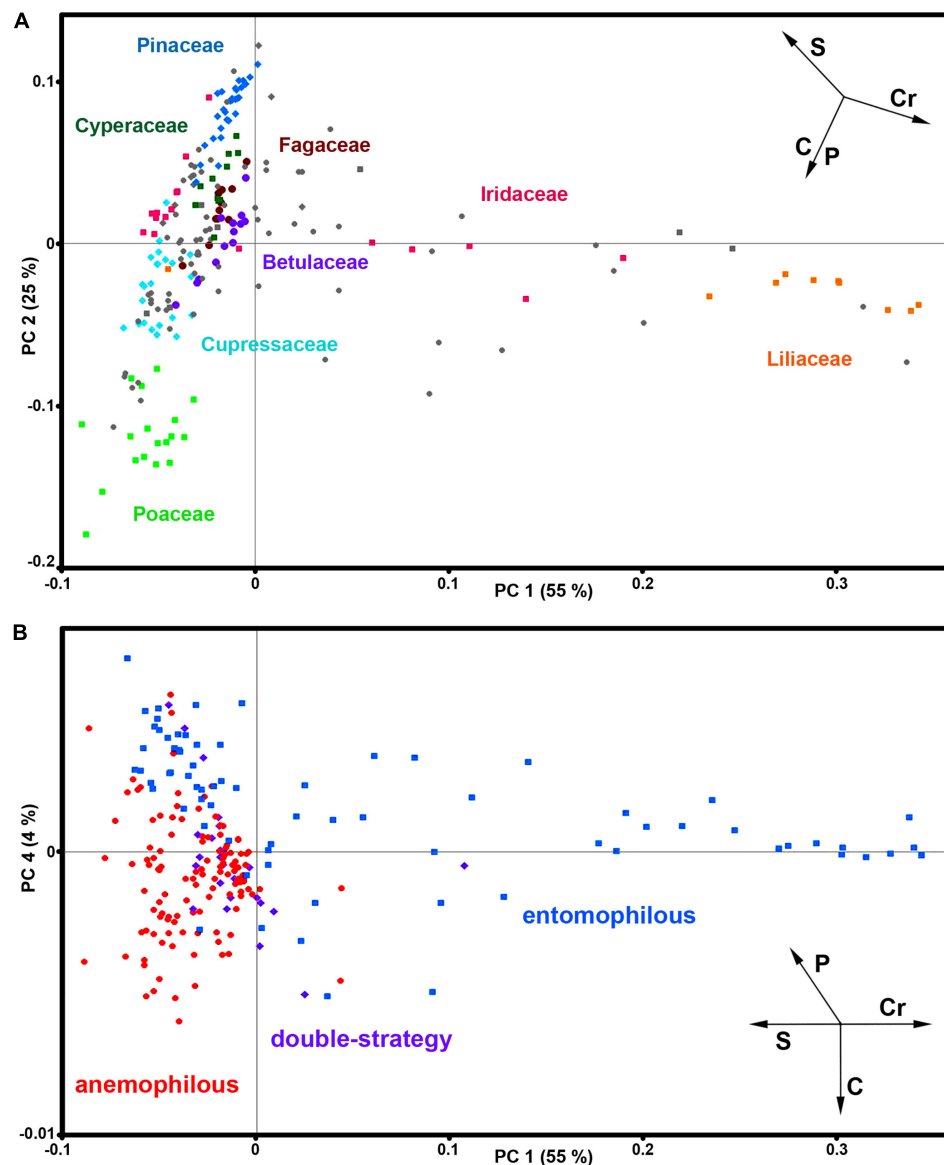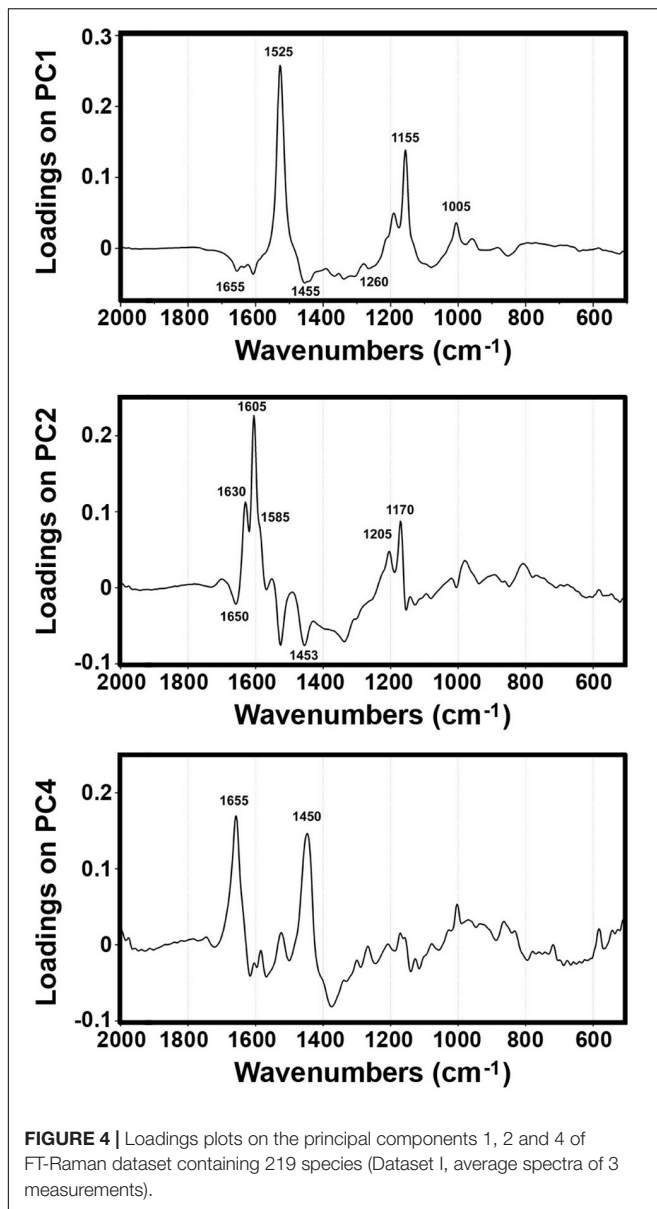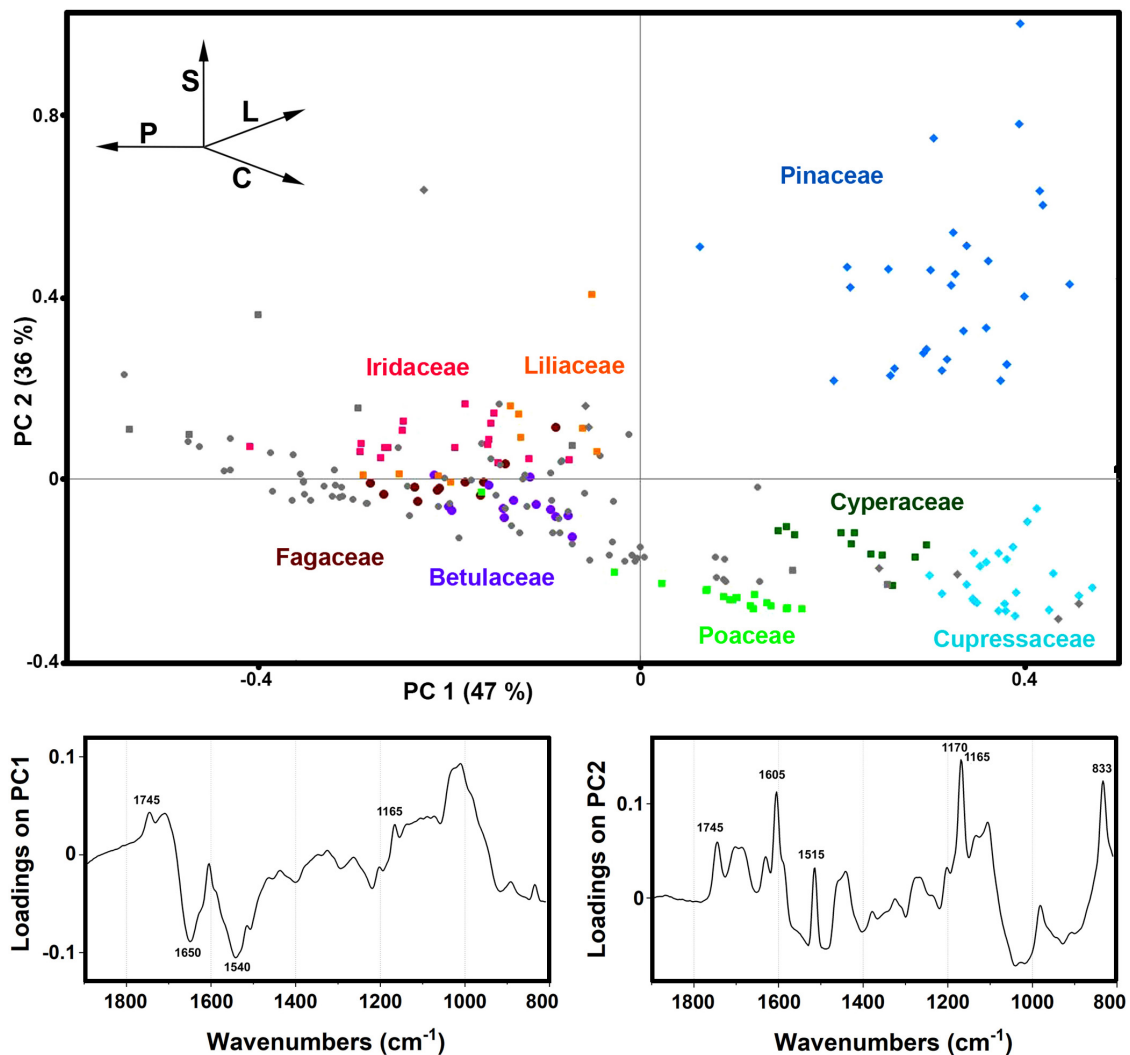
FIGURE 5 | Correlation between FTIR data and taxonomy. PCA score and loading plots of FTIR dataset containing 219 species (*Dataset I*, average spectra of 3 replicates) with: depiction of plant classes (■ monocots; ♦ gymnosperms; ● eudicots and magnoliids) and families (Pinaceae, Cupressaceae, Betulaceae, Fagaceae, Poaceae, Cyperaceae, Liliaceae, and Iridaceae). Vectors are approximating the increase in relative amount of proteins (P), sporopollenins (S), carbohydrates (C), and lipids (L). The percent variances for the first five PCs are 47.44, 35.59, 4.12, 2.53, and 2.18.

including ultraviolet light, temperature, excessive water loss and gain, and microbial damage. In general, pollen wall is comprised of two layers with distinct chemical composition: exine, an outer layer, and intine, an inner layer (Blackmore et al., 2007). Exine is the most complex and resilient plant extracellular matrix, and is predominantly composed of sporopollenins, an extremely robust, chemically resistant and complex biopolymers (Jiang et al., 2013). Sporopollenins are a group of chemically related polymers composed of covalently coupled derivatives of fatty acid and aromatic phenylpropanoid building blocks, with significant taxon-specific variations in chemical composition (Dominguez et al., 1999; Blackmore et al., 2007; Jiang et al., 2013; Li et al., 2019). Production of phenylpropanoids in plants is induced by solar ultraviolet radiation (UV-B) via the phenylpropanoid pathway, the same pathway responsible for synthesis of similar

complex biopolymers, such as lignin and suberin. Unlike the intine and the grain interior (i.e., vegetative and generative cells), which are synthesized under the control of the gametophytic genome, sporopollenins are synthesized in the tapetum under the control of the sporophytic genome (Piffanelli et al., 1998; Blackmore et al., 2007). Therefore, sporopollenin measurements reveal important information on parent plants (sporophytes), in particular concerning plant-environment interactions.

A number of studies have shown response of sporopollenin chemistry to variation in UV-B radiation levels received by sporophytes for a range of different plant species, such as conifers, grasses and legumes (Rozema et al., 2001, 2009; Willis et al., 2011; Lomax et al., 2012; Jardine et al., 2016; Bell et al., 2018). Recent studies have shown that the FTIR analysis of UV-B-absorbing phenylpropanoids in sporopollenins of pollen and

FIGURE 6 | Scatter plot of eigenvalues from the reduced FT-Raman data set (*Dataset I*, average spectra of 3 replicates), obtained by modeling spectral contribution of carbohydrates and carotenoids, with depiction of plant classes (■ monocots; ♦ gymnosperms; ● eudicots and magnoliids) and families (Pinaceae, Cupressaceae, Poaceae, Cyperaceae, Liliaceae, and Iridaceae). Axis scales are in relative units and are based on deflation eigenvalues.

FIGURE 7 | Scatter plot of eigenvalues from the reduced FTIR data set (*Dataset I*, average spectra of 3 replicates), obtained by modeling spectral contribution of carbohydrates and proteins (expressed as C/P ratio), and lipids with depiction of plant classes (■ monocots; ♦ gymnosperms; ● eudicots and magnoliids) and families (Pinaceae, Cupressaceae, Poaceae, Cyperaceae, Liliaceae, and Iridaceae). Axis scales are in relative units and are based on deflation eigenvalues.

spores could provide very valuable record of solar-UV radiation received by plants, which is of high interest in palaeoclimatic and palaeoecological fields. The primary phenylpropanoids in pollen, such as derivatives of *p*-coumaric, ferulic and sinapic acids, have specific vibrational bands in both infrared and FT-Raman spectra, and thus specific spectral regions can be selected and analyzed in detail in order to obtain characteristic chemical fingerprints of

pollen cell wall (Bagcioglu et al., 2015). The main spectral regions (i.e., aromatic regions) for characterization of phenylpropanoids are 860-800 cm^{-1} in infrared spectra, associated with phenyl C-H out-of-plane deformations, and 1650-1580 cm^{-1} in FT-Raman spectra, associated with phenyl C=C stretching vibrations.

The PCA data analysis of these spectral regions shows that the majority of taxa have phylogeny-based similarities

in chemical composition of phenylpropanoids (**Figure 8**). In accordance with our previous finding (Bagcioglu et al., 2015), *Cedrus* is a noteworthy outliers, showing quite different chemistry when compared to the rest of Pinaceae species, with a higher ratio of ferulic-to-*p*-coumaric acid derivatives in sporopollenin compared to the other species. In general, gymnosperms show much higher chemical variability of phenylpropanoids than angiosperms, with substantial differences between Cupressaceae, Cephalotaxaceae, Pinaceae, *Podocarpus*, *Ginkgo*, and *Ephedra* (For example, see differences in 900–800 cm^{-1} region for FTIR spectra of *Pinus ponderosa* and *Ephedra major* in **Figure 1B**). This is not surprising, since all major families of gymnosperms have diverged in Permian-Triassic periods (300-200 Ma) (Lu et al., 2014), much earlier that angiosperm families.

In addition, the analysis has revealed a difference in chemical composition of phenylpropanoids between sedges (Cyperaceae) and grasses (Poaceae). While in both cases the predominant signals belong to *p*-coumaric acid at 830 cm^{-1} in FTIR and 1605 cm^{-1} in FT-Raman, grasses have additional signals associated with ferulic acid at 850 and 1605 cm^{-1} in FTIR and FT-Raman respectively, while sedges have signals associated with sinapic acid at 815 and 1595 cm^{-1} in FTIR and FT-Raman respectively. This is in accordance with phenylpropanoid studies of plant vegetative tissues which have shown that grass cell walls are characterized by ferulic and *p*-coumaric acids, while sedges contain sinapic and *p*-coumaric acids (Bogucka-Kocka et al., 2011; De Oliveira et al., 2015). Sinapic acid is rarely detected in plant tissues, and it has been hypothesized that its presence in tissues of a number of *Carex* species can be associated with the humidity of plants' habitats (Bogucka-Kocka et al., 2011).

Comparative Assessment of FT-Raman and FTIR: A Case Study on Monocots and *Iris*

Here we will demonstrate the benefits of pollen phenotyping by both FT-Raman and FTIR methods by taking a more detailed look on spectral data of monocots (Monocotyledons, Liliopsida). Monocots are large clade, covering a variety of habitats, and include quite diverse group of plants such as lilies, agaves and sedges, as well as grasses which are economically the most important group of plants. The PCA analyses of FT-Raman and FTIR data reveal corresponding and complementary information on pollen chemistry (**Figure 9** and **Supplementary Figures S4, S9**).

The PCA plots have high factor loadings associated with carotenoids (positive loadings) at 1526, 1156 and 1007 cm^{-1}, and proteins (negative) at 1657 and 1450 cm^{-1} in PC 1, and sporopollenins (positive loadings) at 1630, 1603, 1204 and 1171 cm^{-1}, and carbohydrates (negative) at 1460-1300 and 1150-1000 cm^{-1} and carotenoids (negative) at 1526, and 1156 cm^{-1} in PC 2 (**Supplementary Figure S9a**). In contrast, FTIR data shows large variation in carbohydrate-to-protein and carbohydrate-to-lipid ratios. The PC loading plots have high factor loadings associated with proteins (negative loadings) at 1640 and 1530 cm^{-1}, carbohydrates (positive) at 1050-950 cm^{-1}, lipids (negative) at 1744 and 1165 cm^{-1}, and sporopollenins

TABLE 2 | PLS regression results between vibrational spectra and protein content for 35 pollen species (N = 35).

PLSR parameter	FTIR	FT-Raman
RMSE	6.45%$_{wt}$ (15%)	6.19%$_{wt}$ (14%)
R^2	0.487	0.527
R	0.698	0.726
A_{opt}	6	12

%$_{wt}$ = protein mass fraction (% of protein by dry mass).

(negative) at 1605, 1514, 1167, and 831 cm^{-1} in PC 1, and sporopollenins (positive) at 1603, 1512, 1165, and 831 cm^{-1}, lipids (positive) at 1744 and 1165 cm^{-1}, and proteins (negative) at 1645 and 1530 cm^{-1} in PC 2 (**Supplementary Figure S9b**).

Both FT-Raman and FTIR can easily distinguish between the two, predominantly anemophilous, families: grasses (Poaceae) and sedges (Cyperaceae). However, this discrimination is based on relative protein-to-lipid ration in FTIR data, which is high in grasses and low in sedges, while in FT-Raman data it is based on carbohydrate-to-sporopollenin ratio, which is high in grasses and low in sedges. Regarding the three, predominantly entomophilous, families (Xanthorrhoeaceae, Liliaceae and Iridaceae) it is evident that information obtained by FT-Raman has no equivalence in FTIR. The predominant spectral variability in FT-Raman data belongs to relative amount of carotenoids, and, to a less extent, to sporopollenins as well. These results are in accordance with previous studies that have shown significant reserves of starch nutrients in Poaceae and Cyperaceae pollen, while pollen grains of Iridaceae, Xanthorrhoeaceae, Liliaceae, and Arecaceae is predominantly starchless (Franchi et al., 1996).

The large dataset has enabled us to study differences for congeneric species, in particular *Iris* genus comprising spectral data on 17 species. *Iris* is the largest genus of Iridaceae, comprising approx. 250 entomophilous species that inhabit the Earth's North Temperate Zone (Mathew, 1981). *Iris* flowers have elaborate and versatile pollinator attractants, including different color patterns of tepals and sepals with specific orientation and distinctive nectar guides, floral odors, nutritive nectars and pollen, as well as non-nutritive forms of reward, such as shelter and thermal energy (Sapir et al., 2006; Vereecken et al., 2013; Imbert et al., 2014; Guo, 2015; Pellegrino, 2015). Regarding pollen chemistry, triglycerides are the primary nutrient reserve in *Iris* pollen since the grains are starchless (Franchi et al., 1996).

Our study shows that *Iris* species have a large variation of pollen chemistry (**Figure 10** and **Supplementary Figure S4**). The species with large content of carotenoids, such as *I. graminea*, *I. orientalis*, *I. japonica* and *I. crocea*, have relatively low content of lipids and carbohydrates, and high content of proteins (**Figure 10** and **Supplementary Figure S10**). It should be noted that our previous FTIR study has shown that pollen lipids between various species of *Iris* can vary tenfold, as for example between *I. pallida* and *I. graminea* (Zimmermann and Kohler, 2014). This could indicate differences in germination and pollen tube growth, considering the roles of triglycerides in other plants (Rodriguez-Garcia et al., 2003).
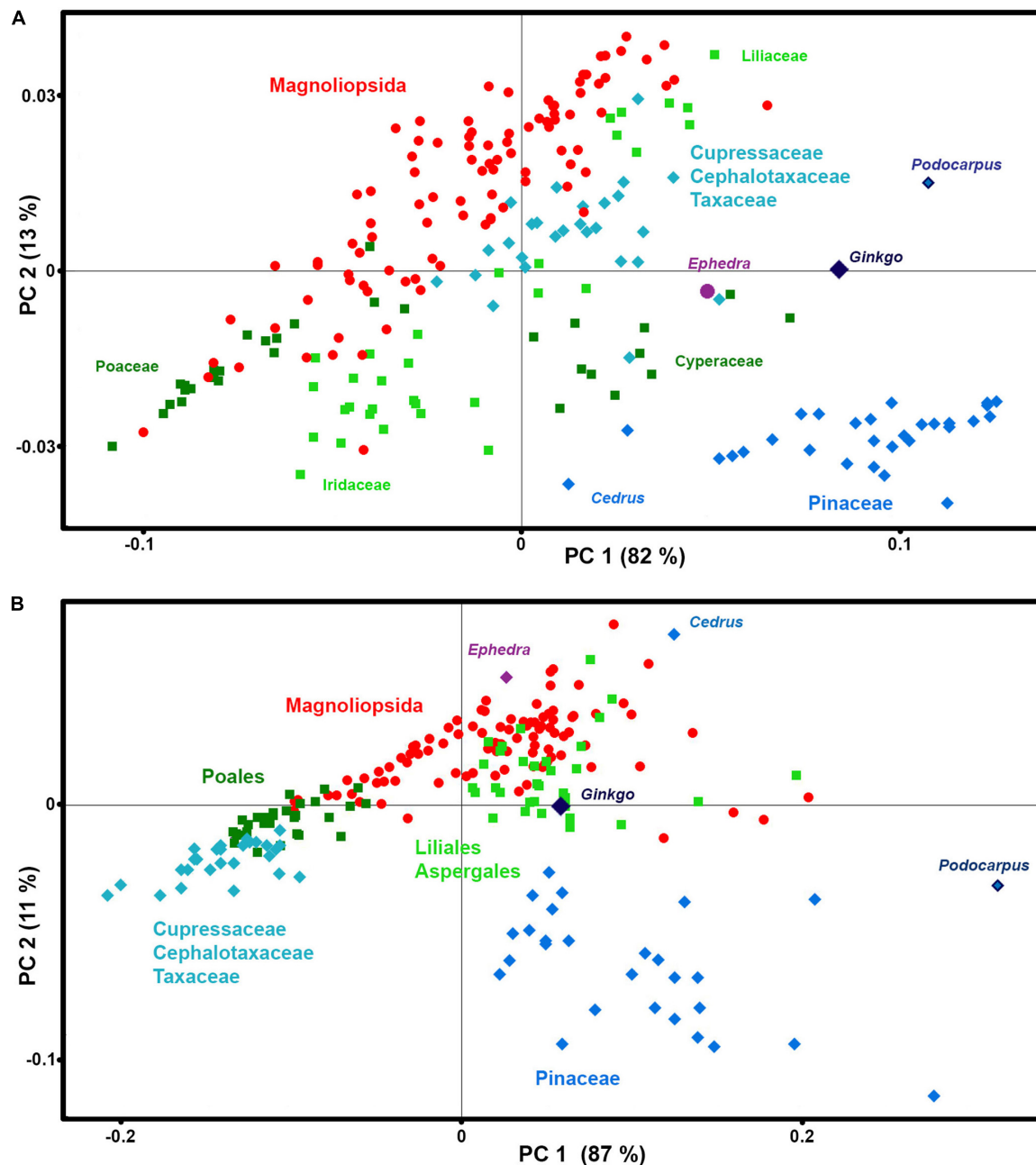
FIGURE 8 | Correlation between phenylpropanoids-related spectral data and taxonomy. PCA score plots of: **(A)** FT-Raman and **(B)** FTIR datasets, containing 219 species (*Datasets II*) with depiction of plant classes (■ monocots; ◆ gymnosperms; ● eudicots and magnoliids), orders, families and selected species.

Pollen with high amount of lipid and carbohydrate nutrients is colorless, such as *I. sikkimensis, I. pallida, I. unguicularis,* and *I. spuria* (**Figure 10** and **Supplementary Table S3**). The spectral results show no clear clustering with phylogeny of Iris genus. For example, *I. unguicularis* is considered to belong to clade Siphonostylis, a sister group (and proposed separate genus) to the rest of *Iris* taxa (Mavrodiev et al., 2014). However, in PCA scores plots for both FT-Raman and FTIR spectral data, *I. unguicularis* PC1 and PC2 scores are close to the median

values (**Figure 10**). Limniris clade (Mavrodiev et al., 2014), comprising *I. versicolor, I. pseudacorus, I. sanguinea, I. sibirica,* and *I. bulleyana,* shows relatively good clustering, though several far-related taxa, such as *I. unguicularis* and *I. spuria,* have similar score values (**Figure 10**). In contrast, Chamaeiris clade (Mavrodiev et al., 2014), comprising *I. graminea, I. spuria,* and *I. orientalis,* shows large difference of score values for both FTIR and FT-Raman data (**Figure 10**). These results indicate that chemical phenotype of *Iris* pollen, as measured by
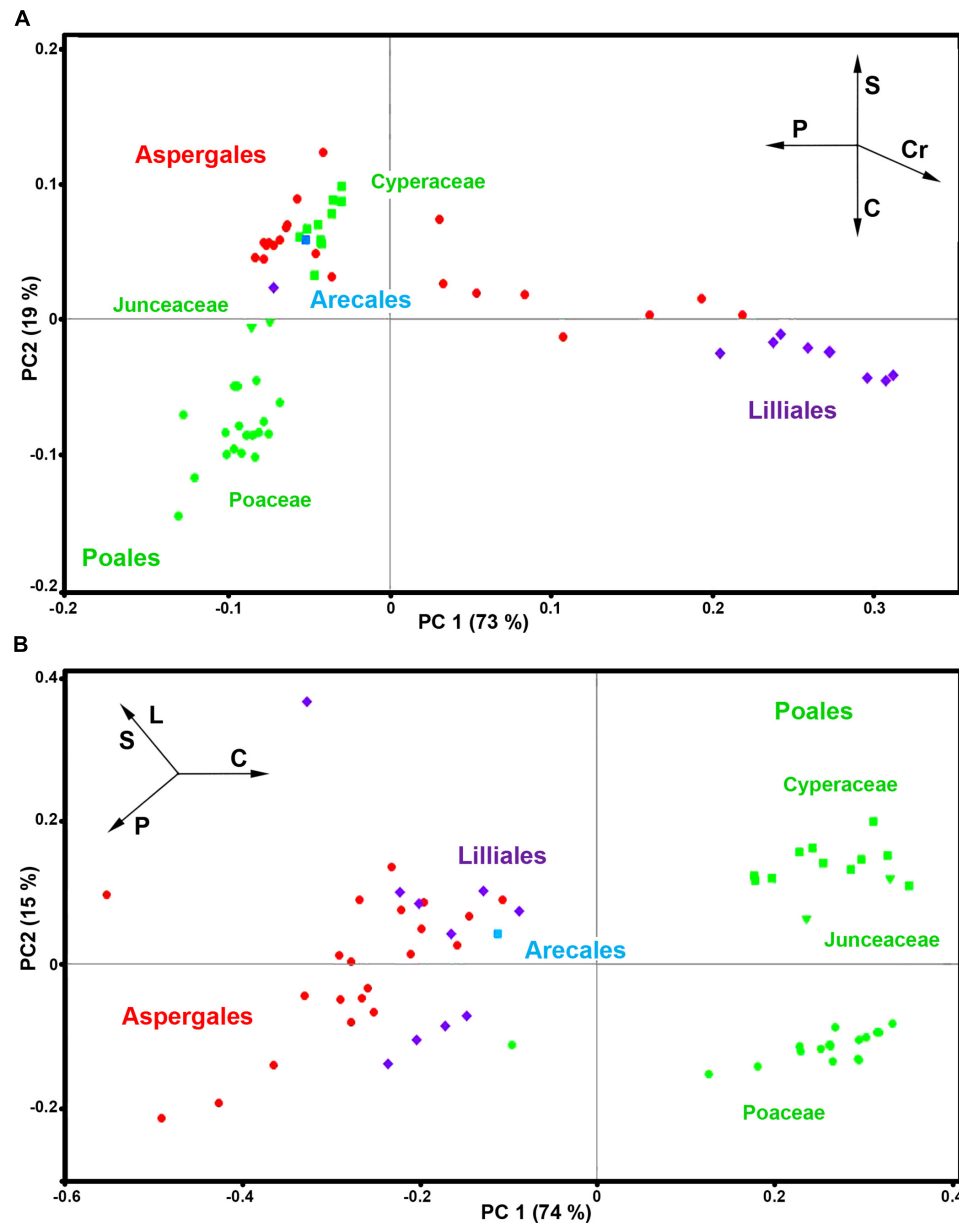
FIGURE 9 | PCA score plots for monocots (Liliopsida): **(A)** FT-Raman and **(B)** FTIR datasets, containing 63 species (*Datasets I*, average spectra of 3 replicates) with depiction of plant order (green - Poales; red - Asperagales; purple – Liliales; blue - Arecales) and families (■ Cyperaceae; ▼ Juncaceae; ● Poaceae). Vectors are approximating the increase in relative amount of proteins (P), sporopollenins (S), carbohydrates (C), lipids (L), and carotenoids (Cr). The percent variances for the first five PCs are 73.09, 18.67, 3.20, 1.78, and 1.39 for FT-Raman, and 73.92, 14.82, 3.88, 2.66, and 1.46 for FTIR.

vibrational spectroscopy, is somewhat unrelated to the genotype. In general, the results clearly show a large variability in pollen chemistry for congeneric species, in particular regarding content of proteins, lipids and carotenoids. In fact, the study indicates large variations even for subgenus clades, such as Chamaeiris clade (Mavrodiev et al., 2014). These extensive differences in pollen biochemistry indicate that congeneric species can employ different reproductive strategies.

Pollen pigments, in the form of carotenoids and flavonoids, are predominantly accumulated in pollenkitt, a sticky lipid-rich pollen coat that covers exine and which is developed under control of sporophytic genome (Pacini and Hesse, 2005; Fambrini et al., 2010). The function of pollen pigments has not been sufficiently studied, but they probably have several functional roles, such as light screening and oxidative stress defense (Stanley and Linskens, 1974; Fambrini et al., 2010). Our study indicates that, in addition to aforementioned functions, pollen carotenoids can also have role in plant signaling, serving as attractor for pollinators to indicate protein-rich pollen. Conversely, pollen coloration could have protective role as nectar guides to direct
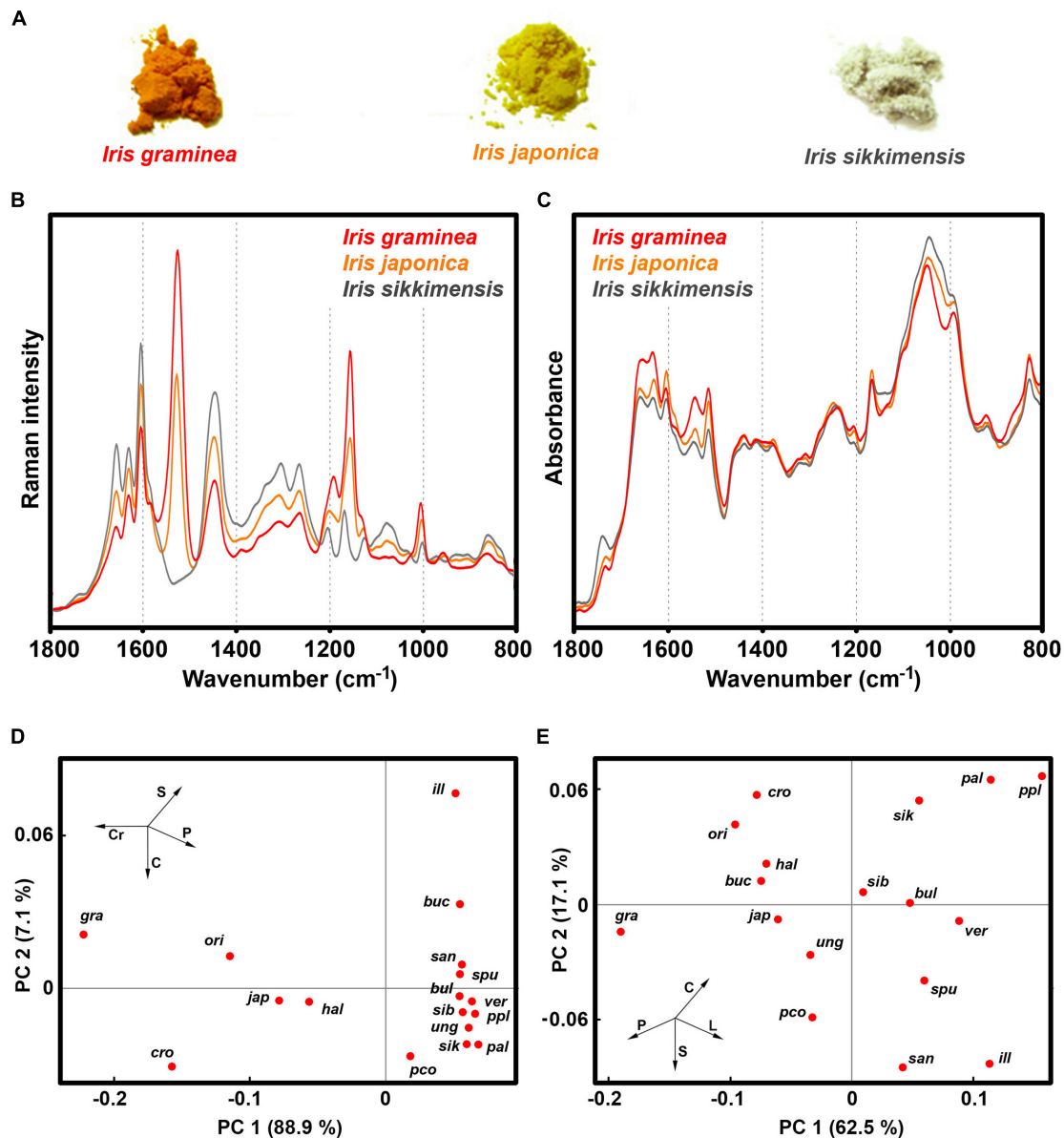
FIGURE 10 | Vibrational spectroscopy of *Iris* pollen: **(A)** Bulk samples of pollen for *I. graminea*, *I. japonica* and *I. sikkimensis*; **(B)** FT-Raman, and **(C)** FTIR spectra of the aforementioned pollen samples; PCA score plots for **(D)** FT-Raman, and **(E)** FTIR spectral data of pollen comprising 17 *Iris* species: *ill - I. illyrica, ppl - I. pseudopallida, pal - I. pallida, sik - I. sikkimensis, ver - I. versicolor, pco - I. pseudacorus, san - I. sanguinea, sib - I. sibirica, bul - I. bulleyana, gra - I. graminea; spu - I. spuria, cro - I. crocea, ori - I. orientalis, hal - I. halophila, ung - I. unguicularis, jap - I. japonica, buc - I. bucharica.*

pollinators toward nectar reward. This notion is partly supported by the fact that brightly colored pollen predominantly belongs to nectar-bearing *Iris* species, such as those belonging to *Iris* subg. *Limniris* series *Spuriae* (see **Supplementary Table S3**), i.e., Chamaeiris clade (Mavrodiev et al., 2014). In those species, primarily floral reward is most likely nectar and not pollen. It has been observed that bumblebees form expectations, based on flower color, on what type of reward a plant will offer (Nicholls and De Ibarra, 2014; Muth et al., 2015). For example, bumblebees can simultaneously learn floral cues associated with pollen and nectar rewards (Muth et al., 2015). However, it is possible

that pollen color is the indirect target of selection through genetic associations with some other trait under selection, such as petal color.

CONCLUSION

Vibrational spectroscopies, coupled with multivariate data analysis, have shown great potential for simple and economical chemical characterization, identification and classification of pollen. This study has demonstrated that high-quality Raman

spectra of pollen, comprising all principal taxa of seed plants, can be obtained by FT-Raman spectroscopy. All FT-Raman spectra were devoid of strong fluorescence background, have high signal-to-noise ratio, and contain clear signals of not only pollen intracellular constituents (lipids, carbohydrates and proteins), but grain wall constituents as well (pigments and sporopollenins). Thus, FT-Raman spectra are superior than corresponding spectra obtained by dispersive Raman spectrometers. In combination with FTIR spectroscopy, FT-Raman spectroscopy is obtaining comprehensive information on pollen biochemistry. Specifically, FT-Raman spectra are strongly biased toward chemical composition of pollen wall constituents, namely sporopollenins (namely phenylpropanoids) and carotenoids, while FTIR spectra are over-representing chemical constituents of the grain interior, such as lipids and carbohydrates. Since chemical composition of pollen depends both on sporophytic genome, which controls expression of sporopollenins and carotenoids, and on gametophytic genome, which controls expression of intracellular lipid and carbohydrate nutrients, it means that each technique can provide unique information on certain aspects of plant genome and pollen development.

The study has demonstrated that the main biochemical constituents of pollen can be identified, and that relative chemical content of pollen can be estimated. Moreover, absolute values of protein content, and probably other chemicals as well, can be obtained by multivariate regression. Our results show a large variability in pollen chemistry for families, genera and even congeneric species, revealing wide range of reproductive strategies. The information on pollen's chemical patterns for major plant taxa should be of value for various studies in plant biology and ecology, including aerobiology, community ecology, palaeoecology, plant-pollinator interactions, and climate effects on plants.

DATA AVAILABILITY STATEMENT

All measured FTIR and FT-Raman spectral data is available in the **Supplementary Material**.

AUTHOR CONTRIBUTIONS

BZ conceived the research idea, contributed to the pollen sampling, performed the FTIR measurements, analyzed the data, and wrote the manuscript. AK and BZ conceived and designed the experiments, discussed and revised the manuscript. AK performed the FT-Raman measurements.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00352/full#supplementary-material

REFERENCES

Bağcıoğlu, M., Kohler, A., Seifert, S., Kneipp, J., and Zimmermann, B. (2017). Monitoring of plant–environment interactions by high-throughput FTIR spectroscopy of pollen. *Methods Ecol. Evol.* 8, 870–880. doi: 10.1111/2041-210x.12697

Bagcioglu, M., Zimmermann, B., and Kohler, A. (2015). A multiscale vibrational spectroscopic approach for identification and biochemical characterization of pollen. *PLoS One* 10:0137899. doi: 10.1371/journal.pone.0137899

Bassani, M., Pacini, E., and Franchi, G. G. (1994). Humidity stress responses in pollen of anemophilous and entomophilous species. *Grana* 33, 146–150. doi: 10.1080/00173139409428991

Bell, B. A., Fletcher, W. J., Ryan, P., Seddon, A. W. R., Wogelius, R. A., and Ilmen, R. (2018). UV-B-absorbing compounds in modern *Cedrus atlantica* pollen: the potential for a summer UV-B proxy for Northwest Africa. *Holocene* 28, 1382–1394. doi: 10.1177/0959683618777072

Bell, C. R. (1959). Mineral nutrition and flower to flower pollen size variation. *Am. J. Bot.* 46, 621–624. doi: 10.1002/j.1537-2197.1959.tb07062.x

Blackmore, S., Wortley, A. H., Skvarla, J. J., and Rowley, J. R. (2007). Pollen wall development in flowering plants. *New Phytol.* 174, 483–498. doi: 10.1111/j.1469-8137.2007.02060.x

Bogucka-Kocka, A., Szewczyk, K., Janyszek, M., Janyszek, S., and Ciesla, L. (2011). RP-HPLC analysis of phenolic acids of selected Central European Carex L. (Cyperaceae) species and its implication for taxonomy. *J. Aoac Int.* 94, 9–16. doi: 10.1093/jaoac/94.1.9

Bokszczanin, K. L., Fragkostefanakis, S., and Itn, S. (2013). Perspectives on deciphering mechanisms underlying plant heat stress response and thermotolerance. *Front. Plant Sci.* 4:315. doi: 10.3389/fpls.2013.00315

De Oliveira, D. M., Finger-Teixeira, A., Mota, T. R., Salvador, V. H., Moreira-Vilar, F. C., Molinari, H. B. C., et al. (2015). Ferulic acid: a key component in grass lignocellulose recalcitrance to hydrolysis. *Plant Biotechnol. J.* 13, 1224–1232. doi: 10.1111/pbi.12292

De Oliveira, V. E., Castro, H. V., Edwards, H. G. M., and De Oliveira, L. F. C. (2009). Carotenes and carotenoids in natural biological samples: a Raman spectroscopic analysis. *J. Raman Spectrosc.* 41, 642–650. doi: 10.1002/jrs.2493

Dell'anna, R., Lazzeri, P., Frisanco, M., Monti, F., Malvezzi Campeggi, F., Gottardini, E., et al. (2009). Pollen discrimination and classification by Fourier transform infrared (FT-IR) microspectroscopy and machine learning. *Anal. Bioanal. Chem.* 394, 1443–1452. doi: 10.1007/s00216-009-2794-9

Depciuch, J., Kasprzyk, I., Roga, E., and Parlinska-Wojtan, M. (2016). Analysis of morphological and molecular composition changes in allergenic Artemisia vulgaris L. pollen under traffic pollution using SEM and FTIR spectroscopy. *Environ. Sci. Pollut. Res.* 23, 23203–23214. doi: 10.1007/s11356-016-7554-8

Depciuch, J., Kasprzyk, I., Sadik, O., and Parlinska-Wojtan, M. (2017). FTIR analysis of molecular composition changes in hazel pollen from unpolluted and urbanized areas. *Aerobiologia* 33, 1–12. doi: 10.1007/s10453-016-9445-3

Diehn, S., Zimmermann, B., Tafintseva, V., Seifert, S., Bağcıoğlu, M., Ohlson, M., et al. (2020). Combining chemical information from grass pollen in multimodal characterization. *Front. Plant Sci.* 10:1788. doi: 10.3389/fpls.2019.01788

Dominguez, E., Mercado, J. A., Quesada, M. A., and Heredia, A. (1999). Pollen sporopollenin: degradation and structural elucidation. *Sex. Plant Reprod.* 12, 171–178. doi: 10.1007/s004970050189

Fambrini, M., Michelotti, V., and Pugliesi, C. (2010). Orange, yellow and white-cream: inheritance of carotenoid-based colour in sunflower pollen. *Plant Biol.* 12, 197–205. doi: 10.1111/j.1438-8677.2009.00205.x

Franchi, G. G., Bellani, L., Nepi, M., and Pacini, E. (1996). Types of carbohydrate reserves in pollen: localization, systematic distribution and ecophysiological significance. *Flora* 191, 143–159. doi: 10.1016/s0367-2530(17)30 706-5

Gottardini, E., Rossi, S., Cristofolini, F., and Benedetti, L. (2007). Use of Fourier transform infrared (FT-IR) spectroscopy as a tool for pollen identification. *Aerobiologia* 23, 211–219. doi: 10.1007/s10453-007-9065-z

Guedes, A., Ribeiro, H., Fernandez-Gonzalez, M., Aira, M. J., and Abreu, I. (2014). Pollen Raman spectra database: application to the identification of airborne pollen. *Talanta* 119, 473–478. doi: 10.1016/j.talanta.2013.11.046

Guo, J. Y. (2015). Comparative micromorphology and anatomy of crested sepals in Iris (Iridaceae). *Int. J. Plant Sci.* 176, 627–642. doi: 10.1086/682135

Guo, S. X., Kohler, A., Zimmermann, B., Heinke, R., Stockel, S., Rosch, P., et al. (2018). Extended multiplicative signal correction based model transfer for Raman spectroscopy in biological applications. *Anal. Chem.* 90, 9787–9795. doi: 10.1021/acs.analchem.8b01536

Hall, J. A., and Walter, G. H. (2011). Does pollen aerodynamics correlate with pollination vector? Pollen settling velocity as a test for wind versus insect pollination among cycads (Gymnospermae: Cycadaceae: Zamiaceae). *Biol. J. Linn. Soc.* 104, 75–92. doi: 10.1111/j.1095-8312.2011.01695.x

Imbert, E., Wang, H., Conchou, L., Vincent, H., Talavera, M., and Schatz, B. (2014). Positive effect of the yellow morph on female reproductive success in the flower colour polymorphic Iris lutescens (Iridaceae), a deceptive species. *J. Evol. Biol.* 27, 1965–1974. doi: 10.1111/jeb.12451

Ivleva, N. P., Niessner, R., and Panne, U. (2005). Characterization and discrimination of pollen by Raman microscopy. *Anal. Bioanal. Chem.* 381, 261–267. doi: 10.1007/s00216-004-2942-1

Jardine, P. E., Fraser, W. T., Lomax, B. H., and Gosling, W. D. (2015). The impact of oxidation on spore and pollen chemistry. *J. Micropalaeontol.* 34, 139–149. doi: 10.1144/jmpaleo2014-022

Jardine, P. E., Fraser, W. T., Lomax, B. H., Sephton, M. A., Shanahan, T. M., Miller, C. S., et al. (2016). Pollen and spores as biological recorders of past ultraviolet irradiance. *Sci. Rep.* 6:39269. doi: 10.1038/srep39269

Jardine, P. E., Gosling, W. D., Lomax, B. H., Julier, A. C. M., and Fraser, W. T. (2019). Chemotaxonomy of domesticated grasses: a pathway to understanding the origins of agriculture. *J. Micropalaeontol.* 38, 83–95. doi: 10.5194/jm-38-83-2019

Jiang, J., Zhang, Z., and Cao, J. (2013). Pollen wall development: the associated enzymes and metabolic pathways. *Plant Biol.* 15, 249–263. doi: 10.1111/j.1438-8677.2012.00706.x

Jiang, Y. F., Lahlali, R., Karunakaran, C., Kumar, S., Davis, A. R., and Bueckert, R. A. (2015). Seed set, pollen morphology and pollen surface composition response to heat stress in field pea. *Plant Cell Environ.* 38, 2387–2397. doi: 10.1111/pce.12589

Joester, M., Seifert, S., Emmerling, F., and Kneipp, J. (2017). Physiological influence of silica on germinating pollen as shown by Raman spectroscopy. *J. Biophot.* 10, 542–552. doi: 10.1002/jbio.201600011

Julier, A. C. M., Jardine, P. E., Coe, A. L., Gosling, W. D., Lomax, B. H., and Fraser, W. T. (2016). Chemotaxonomy as a tool for interpreting the cryptic diversity of Poaceae pollen. *Rev. Palaeobot. Palynol.* 235, 140–147. doi: 10.1016/j.revpalbo.2016.08.004

Kosa, G., Kohler, A., Tafintseva, V., Zimmermann, B., Forfang, K., Afseth, N. K., et al. (2017). Microtiter plate cultivation of oleaginous fungi and monitoring of lipogenesis by high-throughput FTIR spectroscopy. *Microb. Cell Fact.* 16:101. doi: 10.1186/s12934-017-0716-7

Lahlali, R., Jiang, Y., Kumar, S., Karunakaran, C., Liu, X., Borondics, F., et al. (2014). ATR-FTIR spectroscopy reveals involvement of lipids and proteins of intact pea pollen grains to heat stress tolerance. *Front. Plant Sci.* 5:747. doi: 10.3389/fpls.2014.00747

Li, F. S., Phyo, P., Jacobowitz, J., Hong, M., and Weng, J. K. (2019). The molecular structure of plant sporopollenin. *Nat. Plants* 5, 41–46. doi: 10.1038/s41477-018-0330-7

Lindbladh, M. S., O'connor, R., and Jacobson, G. L. (2002). Morphometric analysis of pollen grains for paleoecological studies: classification of Picea from eastern North America. *Am. J. Bot.* 89, 1459–1467. doi: 10.3732/ajb.89.9.1459

Lomax, B. H., Fraser, W. T., Harrington, G., Blackmore, S., Sephton, M. A., and Harris, N. B. W. (2012). A novel palaeoaltimetry proxy based on spore and pollen wall chemistry. *Earth Planet. Sci. Lett.* 353, 22–28. doi: 10.1016/j.epsl.2012.07.039

Lu, Y., Ran, J. H., Guo, D. M., Yang, Z. Y., and Wang, X. Q. (2014). Phylogeny and divergence times of gymnosperms inferred from single-copy nuclear genes. *PLoS One* 9:0107679. doi: 10.1371/journal.pone.0107679

Lutz, C. (2010). Cell physiology of plants growing in cold environments. *Protoplasma* 244, 53–73. doi: 10.1007/s00709-010-0161-5

Mathew, B. (1981). *The Iris*. New York, NY: Universe Books.

Mavrodiev, E. V., Martinez-Azorin, M., Dranishnikov, P., and Crespo, M. B. (2014). At Least 23 Genera Instead of one: the case of iris L. s.l. (Iridaceae). *PLoS One* 9:0106459. doi: 10.1371/journal.pone.0106459

Moester, M. J. B., Zada, L., Fokker, B., Ariese, F., and De Boer, J. F. (2019). Stimulated Raman scattering microscopy with long wavelengths for improved imaging depth. *J. Raman Spectrosc.* 50, 1321–1328. doi: 10.1002/jrs.5494

Molina, R. T., Rodriguez, A. M., Palacios, I. S., and Lopez, F. G. (1996). Pollen production in anemophilous trees. *Grana* 35, 38–46. doi: 10.1080/00173139609430499

Mondol, A. S., Patel, M. D., Ruger, J., Stiebing, C., Kleiber, A., Henkel, T., et al. (2019). Application of high-throughput screening Raman spectroscopy (HTS-RS) for label-free identification and molecular characterization of Pollen. *Sensors* 19:4428. doi: 10.3390/s19204428

Muth, F., Papaj, D. R., and Leonard, A. S. (2015). Colour learning when foraging for nectar and pollen: bees learn two colours at once. *Biol. Lett.* 11:20150628. doi: 10.1098/rsbl.2015.0628

Nicholls, E., and De Ibarra, N. H. (2014). Bees associate colour cues with differences in pollen rewards. *J. Exp. Biol.* 217, 2783–2788. doi: 10.1242/jeb.106120

Oddou-Muratorio, S., Klein, E. K., and Austerlitz, F. (2005). Pollen flow in the wildservice tree, Sorbus torminalis (L.) Crantz. II. Pollen dispersal and heterogeneity in mating success inferred from parent-offspring analysis. *Mol. Ecol.* 14, 4441–4452. doi: 10.1111/j.1365-294x.2005.02720.x

Pacini, E., Franchi, G. G., Lisci, M., and Nepi, M. (1997). Pollen viability related to type of pollination in six angiosperm species. *Ann. Bot.* 80, 83–87. doi: 10.1006/anbo.1997.0421

Pacini, E., Guarnieri, M., and Nepi, M. (2006). Pollen carbohydrates and water content during development, presentation, and dispersal: a short review. *Protoplasma* 228, 73–77. doi: 10.1007/s00709-006-0169-z

Pacini, E., and Hesse, M. (2005). Pollenkitt – its composition, forms and functions. *Flora* 200, 399–415. doi: 10.1016/j.flora.2005.02.006

Pappas, C. S., Tarantilis, P. A., Harizanis, P. C., and Polissiou, M. G. (2003). New method for pollen identification by FT-IR spectroscopy. *Appl. Spectrosc.* 57, 23–27. doi: 10.1366/000370203321165160

Pellegrino, G. (2015). Pollinator limitation on reproductive success in Iris tuberosa. *Aob Plants* 7:lu089. doi: 10.1093/aobpla/plu089

Piffanelli, P., Ross, J. H. E., and Murphy, D. J. (1998). Biogenesis and function of the lipidic structures of pollen grains. *Sex. Plant Reprod.* 11, 65–80. doi: 10.1186/1471-2229-11-122

Pummer, B. G., Bauer, H., Bernardi, J., Chazallon, B., Facq, S., Lendl, B., et al. (2013). Chemistry and morphology of dried-up pollen suspension residues. *J. Raman Spectrosc.* 44, 1654–1658. doi: 10.1002/jrs.4395

Rodriguez-Garcia, M. I., M'rani-Alaoui, M., and Fernandez, M. C. (2003). Behavior of storage lipids during development and germination of olive (Olea europaea L.) pollen. *Protoplasma* 221, 237–244. doi: 10.1007/s00709-002-0076-x

Roulston, T. H., Cane, J. H., and Buchmann, S. L. (2000). What governs protein content of pollen: pollinator preferences, pollen-pistil interactions, or phylogeny? *Ecol. Monogr.* 70, 617–643. doi: 10.1890/0012-9615(2000) 070[0617:wgpcop]2.0.co;2

Rozema, J., Blokker, P., Fuertes, M. A. M., and Broekman, R. (2009). UV-B absorbing compounds in present-day and fossil pollen, spores, cuticles, seed coats and wood: evaluation of a proxy for solar UV radiation. *Photochem. Photobiol. Sci.* 8, 1233–1243. doi: 10.1039/b904515e

Rozema, J., Noordijk, A. J., Broekman, R. A., Van Beem, A., Meijkamp, B. M., De Bakker, N. V. J., et al. (2001). (Poly)phenolic compounds in pollen and spores of Antarctic plants as indicators of solar UV-B - A new proxy for the

reconstruction of past solar UV-B? *Plant Ecol.* 154, 9–26. doi: 10.1007/978-94-017-2892-8_2

Sapir, Y., Shmida, A., and Ne'eman, G. (2006). Morning floral heat as a reward to the pollinators of the *Oncocyclus irises*. *Oecologia* 147, 53–59. doi: 10.1007/s00442-005-0246-6

Schulte, F., Lingott, J., Panne, U., and Kneipp, J. (2008). Chemical characterization and classification of pollen. *Anal. Chem.* 80, 9551–9556. doi: 10.1021/ac801791a

Schulte, F., Mader, J., Kroh, L. W., Panne, U., and Kneipp, J. (2009). Characterization of pollen carotenoids with in situ and high-performance thin-layer chromatography supported resonant Raman spectroscopy. *Anal. Chem.* 81, 8426–8433. doi: 10.1021/ac901389p

Schulte, F., Panne, U., and Kneipp, J. (2010). Molecular changes during pollen germination can be monitored by Raman microspectroscopy. *J. Biophotonics* 3, 542–547. doi: 10.1002/jbio.201000031

Seifert, S., Merk, V., and Kneipp, J. (2016). Identification of aqueous pollen extracts using surface enhanced Raman scattering (SERS) and pattern recognition methods. *J. Biophotonics* 9, 181–189. doi: 10.1002/jbio.201500176

Speranza, A., Calzoni, G. L., and Pacini, E. (1997). Occurrence of mono- or disaccharides and polysaccharide reserves in mature pollen grains. *Sex. Plant Reprod.* 10, 110–115. doi: 10.1007/s004970050076

Stanley, R. G., and Linskens, H. F. (1974). *Pollen: Biology, Biochemistry, Management.* New York, NY: Springer-Verlag.

Streiff, R., Ducousso, A., Lexer, C., Steinkellner, H., Gloessl, J., and Kremer, A. (1999). Pollen dispersal inferred from paternity analysis in a mixed oak stand of Quercus robur L-and Q-petraea (Matt.) Liebl. *Mol. Ecol.* 8, 831–841. doi: 10.1046/j.1365-294x.1999.00637.x

Tamura, S., and Kudo, G. (2000). Wind pollination and insect pollination of two temperate willow species, Salix miyabeana and Salix sachalinensis. *Plant Ecol.* 147, 185–192.

Vereecken, N. J., Dorchin, A., Dafni, A., Hotling, S., Schulz, S., and Watts, S. (2013). A pollinators eye view of a shelter mimicry system. *Ann. Bot.* 111, 1155–1165. doi: 10.1093/aob/mct081

Wang, Z. Y., Ge, Y. X., Scott, M., and Spangenberg, G. (2004). Viability and longevity of pollen from transgenic and nontransgenic tall fescue (*Festuca arundinacea*) (Poaceae) plants. *Am. J. Bot.* 91, 523–530. doi: 10.3732/ajb.91.4.523

Welsford, M. R., Hobbhahn, N., Midgley, J. J., and Johnson, S. D. (2016). Floral trait evolution associated with shifts between insect and wind pollination in the dioecious genus *Leucadendron* (Proteaceae). *Evolution* 70, 126–139. doi: 10.1111/evo.12821

White, P. A. (1958). The computation of eigenvalues and eigenvectors of a matrix. *J. Soc. Indus. Appl. Math.* 6, 393–437.

Williams, J. H., and Mazer, S. J. (2016). Pollen-Tiny and ephemeral but not forgotten: new ideas on their ecology and evolution. *Am. J. Bot.* 103, 365–374. doi: 10.3732/ajb.1600074

Willis, K. J., Feurdean, A., Birks, H. J. B., Bjune, A. E., Breman, E., Broekman, R., et al. (2011). Quantification of UV-B flux through time using UV-B-absorbing compounds contained in fossil *Pinus sporopollenin*. *New Phytol.* 192, 553–560. doi: 10.1111/j.1469-8137.2011.03815.x

Woutersen, A., Jardine, P. E., Bogota-Angel, R. G., Zhang, H. X., Silvestro, D., Antonelli, A., et al. (2018). A novel approach to study the morphology and chemistry of pollen in a phylogenetic context, applied to the halophytic taxon Nitraria L.(Nitrariaceae). *PeerJ* 6:e5055. doi: 10.7717/peerj.5055

Zimmermann, B. (2010). Characterization of pollen by vibrational spectroscopy. *Appl. Spectrosc.* 64, 1364–1373. doi: 10.1366/000370210793561664

Zimmermann, B. (2018). Chemical characterization and identification of Pinaceae pollen by infrared microspectroscopy. *Planta* 247, 171–180. doi: 10.1007/s00425-017-2774-9

Zimmermann, B., Bagcioglu, M., Sandt, C., and Kohler, A. (2015a). Vibrational microspectroscopy enables chemical characterization of single pollen grains as well as comparative analysis of plant species based on pollen ultrastructure. *Planta* 242, 1237–1250. doi: 10.1007/s00425-015-2380-7

Zimmermann, B., Bagcioglu, M., Tafinstseva, V., Kohler, A., Ohlson, M., and Fjellheim, S. (2017). A high-throughput FTIR spectroscopy approach to assess adaptive variation in the chemical composition of pollen. *Ecol. Evol.* 7, 10839–10849. doi: 10.1002/ece3.3619

Zimmermann, B., and Kohler, A. (2013). Optimizing savitzky-golay parameters for improving spectral resolution and quantification in infrared spectroscopy. *Appl. Spectrosc.* 67, 892–902. doi: 10.1366/12-06723

Zimmermann, B., and Kohler, A. (2014). Infrared spectroscopy of pollen identifies plant species and genus as well as environmental conditions. *PLoS One* 9:e95417. doi: 10.1371/journal.pone.0095417

Zimmermann, B., Tafintseva, V., Bagcioglu, M., Hoegh Berdahl, M., and Kohler, A. (2016). Analysis of allergenic pollen by FTIR microspectroscopy. *Anal. Chem.* 88, 803–811. doi: 10.1021/acs.analchem.5b03208

Zimmermann, B., Tkalcec, Z., Mesic, A., and Kohler, A. (2015b). Characterizing aeroallergens by infrared spectroscopy of fungal spores and pollen. *PLoS One* 10:e0124240. doi: 10.1371/journal.pone.0124240

Herbal Medicine Characterization Perspectives Using Advanced FTIR Sample Techniques – Diffuse Reflectance (DRIFT) and Photoacoustic Spectroscopy (PAS)

Agnese Brangule[1]*, Renāte Šukele[2] and Dace Bandere[2]

[1] *Department of Human Physiology and Biochemistry, Riga Stradiņš University, Riga, Latvia,* [2] *Department of Pharmaceutical Chemistry, Riga Stradiņš University, Riga, Latvia*

This study demonstrates the significant potential of the Fourier transform infrared spectroscopy (FTIR) sampling methods: cantilever-enhanced Fourier transform infrared photoacoustic spectroscopy (FTIR PAS) and diffuse reflectance infrared spectroscopy (FTIR DRIFT) in the field of herbal medicines (HM). In the present work we investigated DRIFT and PAS sampling methods because they do not require sample preparation, samples may be opaque or dark, require small amounts, both liquid and solid samples can be measured, and solid samples can be analyzed on a small scale. Experiments conducted prove high sensitivity, reproducibility and capability in combination with an unsupervised multivariate analysis technique to discriminate important characteristics of HM, such as the identification of plant parts, differentiation of samples by types, and determination of the concentration of extractable compounds in HM.

Keywords: herbal medicines, FTIR DRIFT, FTIR PAS, cluster analysis, herbal differentiation

INTRODUCTION

In today's world, people are increasingly focusing on healthy lifestyles and the use of herbs. They understand that chemical compounds in herbs can not only help to fight specific diseases but can also be preventative, improving overall health (Eniņa, 2017).

The chemical composition of herbs may vary depending on the species, location of growth, age, harvesting season, drying conditions, and other factors (Heinrich, 2015). Therefore, comprehensive studies of effective analytical methods are required to make quick and reliable quality control at any stage of herbal medicine (HM) production as well as during the storage process, to obtain feedback (Bostijn et al., 2018; Rehrl et al., 2018).

The World Health Organization and European Pharmacopeia provide guidelines for the assessment of the quality of HM (Bunaciu et al., 2011; Kitanov et al., 2015). In previous studies, there were various techniques that were used to obtain a complete overview of a herbal product, for example, chromatography methods (CM) (Yang et al., 2013; Kitanov et al., 2015). However, CM methods have some critical disadvantages: complicated sample preparation procedures and a long analysis time. Moreover, these are sample-destructive methods (Peerapattana et al., 2015).

Another commonly used method is Fourier transform infrared spectroscopy (FTIR). FTIR methods have been widely used since the 1960s and can be used for both qualitative and quantitative

analysis (Stuart, 2004; Smith, 2011). In the field of HMs, the FTIR fingerprint spectra have been used since early 1987, and are used less frequently than CM (Zou et al., 2005). Until now, the introduction of FTIR methods was limited by the complexity of spectra and its interpretation (Rohman et al., 2014). On the other hand, FTIR spectroscopy, in conjunction with multidimensional statistical analysis (Chemometrics), offers a wide scope for HM studies (Kadiroğlu et al., 2018; Miaw et al., 2018). Chemometrics is defined as the application of mathematics and statistics to treat chemical data (Gemperline, 2006), and provide a good opportunity for mining more useful chemical information from the original spectral data by using unsupervised [Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA)] and supervised classification methods (Rohman et al., 2019).

The major advantages of FTIR methods are the following: methods are sensitive and non-destructive or only slightly damage the sample; they require minimal sample preparation; small sample quantities are necessary for measuring (Smith, 2011). One of the essential features of FTIR is the possibility to simultaneously determine different components in the same sample from a single instrumental measurement (Moros et al., 2010). In previous studies, the FTIR spectroscopy's sampling methods transmission and the attenuated total reflection (ATR) was most widely used for diffuse reflectance (DRIFT) and cantilever-enhanced photoacoustic spectroscopy (FTIR PAS) (Legner et al., 2018). The significant benefit of ATR is the ability to measure a wide variety of solid and liquid samples without requiring complex preparations because the ATR measurement is independent of sample thickness and requires small amounts of sample material (Kazarian and Chan, 2006; Rodriguez-Saona and Allendorf, 2011). However, there are significant disadvantages of the ATR sampling method:

- dependence of intensities on a wavenumber and shifts has essential implications on the interpretation of spectra; a particular "ATR correction" function must be used to reduce these differences (Grdadolnik, 2002);
- the spectral absorption area for the ATR sampling method is narrower than other FTIR methods – it starts from 500 to 600 cm^{-1}, and important fingerprint region information might be lost.

In the present work, we used DRIFT and PAS sampling methods because they, like ATR, do not require sample preparation, samples may be opaque or dark, the methods require small amounts of sample, both liquid and solid samples can be measured, and solid samples can be analyzed on a small scale. Furthermore, PAS is a little-explored technique for plants (Dias et al., 2018). PAS is based on the photoacoustic effect. The sample is placed in the photoacoustic measurement cell and irradiated with modulated infrared light through a window that is absorbed by the sample at characteristic wavelengths. The heating of the sample generates a pressure wave whose amplitude is detected with a microphone (Gasera, 2010). Photoacoustic spectroscopy is an advantageous method

for the measurement and analysis of solid or semi-solid samples, also for powders, fibers, and samples of a very small size, the photoacoustic signal contains information of the surface and inner layers of samples. The shape of the photoacoustic spectrum is independent of the morphology of the sample (Kauppinen et al., 2004; Kuusela and Kauppinen, 2007; Uotila and Kauppinen, 2008).

The purpose of this work is to evaluate the use of FTIR DRIFT with a diamond sampling stick and cantilever-enhanced FTIR PAS in the characterization of HMs.

MATERIALS AND EQUIPMENT

FTIR Sampling Techniques

Cantilever-enhanced photoacoustic spectroscopy (PAS) and diffuse reflectance (DRIFT) spectra were taken with PerkinElmer Spectrum One (450–4000 cm^{-1}, resolution of 4 cm^{-1}, 10 scans, aperture 8.94 mm, scan speed: 0.2 cm/s). PAS spectra were taken with Gasera PA301 Photoacoustic FTIR accessory (450–4000 cm^{-1}, resolution of 8 cm^{-1}, 10 scans, aperture 8.94 mm, scan speed: 0.2 cm/s), with the cell being filled with helium gas (flow 0.5 l/min) and a carbon black reference. A unique preparation method was not required for solid, powdered herbals; 0.030–0.040 g of powder was placed in the PAS cell. Each sample was sampled 5 times to reduce the influence of inhomogeneity in the test results.

A modified DRIFT sampling technique was used. Solid, powdered samples were measured directly on the diamond sampling stick. Liquid extracts (60 µl) were placed in the aluminum sampling-cup and evaporated. The raw diffuse reflectance spectra DRIFT will appear different from its transmission equivalent (stronger than expected absorption from weak IR bands). DRIFT spectra were taken in Kubelka-Munk units to compensate for these differences (Greene et al., 2004).

Plant Material
Characterization of Plant Material

Seven medicinal plant species in the form of dried tea samples were measured: chamomile [*Matricaria recutita, 8 different commercial available tea samples* (code – A)], silver birch [*Betula pendula Roth* (B)], hibiscus [*Hibiscus sabdariffa* (C)], peppermint [*Mentha piperita* (D)], cornflower [*Centaurea cyanus* (E)], meadowsweet [*Filipendula ulmaria* (Ms), tansy (*Tanacetum vulgare* (Ta)] (information about samples and measurement conditions in **Table 1**).

The herbs meadowsweet (Ms) and tansy (Ta) (*Filipendula ulmaria, Achillea millefolium, Tanacetum vulgare*) were collected while they were blossoming in meadows in Latvia (Sigulda and Ropaži district). The other herb samples were made from herbals produced by Latvian herbal companies.

Preparation of Plant Material

All dried herbs were grinded to powder and sifted through a 2 mm sieve. Powders were stored at room temperature for further analysis.

TABLE 1 | Used materials and methods.

Herbal material	Code in text and figures	Samples	Measurement methods	Measurement conditions
Chamomile (*Matricaria recutita*)	A	8 different commercial available HM samples, Dried powder, extracts (6 pcs.)	DRIFT FTIR PAS FTIR (10 scans, 5 independent sampling for each sample)	450–4000 cm^{-1} DRIFT – 10 scans, resolution 4 cm^{-1} PAS – 10 scans, resolution 8 cm^{-1}
Silver birch (*Betula pendula Roth*)	B	1 commercial available HM sample, Dried powder	DRIFT FTIR PAS FTIR (10 scans, 5 independent sampling for each sample)	450–4000 cm^{-1} DRIFT – 10 scans, resolution 4 cm^{-1} PAS – 10 scans, resolution 8 cm^{-1}
Hibiscus (*Hibiscus sabdariffa*	C	1 commercial available HM sample, Dried powder	DRIFT FTIR PAS FTIR (10 scans, 5 independent sampling for each sample)	450–4000 cm^{-1} DRIFT – 10 scans, resolution 4 cm^{-1} PAS – 10 scans, resolution 8 cm^{-1}
Peppermint (*Mentha piperita*)	D	1 commercial available HM sample, Dried powder	DRIFT FTIR PAS FTIR (10 scans, 5 independent sampling for each sample)	450–4000 cm^{-1} DRIFT – 10 scans, resolution 4 cm^{-1} PAS – 10 scans, resolution 8 cm^{-1}
Cornflower (*Centaurea cyanus*	E	1 commercial available HM sample, Dried powder	DRIFT FTIR PAS FTIR (10 scans, 5 independent sampling for each sample)	450–4000 cm^{-1} DRIFT – 10 scans, resolution 4 cm^{-1} PAS – 10 scans, resolution 8 cm^{-1}
Meadowsweet (*Filipendula ulmaria*)	Ms	13 different self-collected samples (5 leaves, 8 flowers), Dried powder	DRIFT FTIR (10 scans, 1 independent sampling for each sample)	450–4000 cm^{-1} DRIFT – 10 scans, resolution 4 cm^{-1}
Tansy (*Tanacetum vulgare*)	Ta	2 samples (leaves and flowers), Dried powder, Extracts (13 samples)	DRIFT FTIR (10 scans, 1 independent sampling for each sample)	450–4000 cm^{-1} DRIFT – 10 scans, resolution 4 cm^{-1}

Two extract preparation methods were used:

Method No. 1. The modified extraction method was developed based on literature studies on tannin and phenolic compound extraction methods (Izam, 2012; Nelce et al., 2013). 10 g of dried plant powder was extracted in 100 ml of 30%, 50% or 70% ethanol or acetone in an orbital shaker (180 rpm) for 120 min. at room temperature. The extracts obtained were filtered using Whatman No. 1 filter paper and evaporated to dryness by using a rotary vacuum evaporator to viscose constancy liquid at 60°C and stored at -4°C in an airtight container for further analysis.

Method No. 2. The extraction method was developed based on literature studies on compound extraction for FTIR sampling methods. 5 g of dried plant powder was extracted in 50 ml of ethanol in an occasional shaker for 24 h at room temperature. The extract was filtered using Whatman No. 1 filter paper, and the supernatant was collected and stored at -4°C in an airtight container for further analysis. The residue was collected, dried at room temperature, and stored at room temperature for further analysis (Wulandari et al., 2016).

METHODS

Spectral Pre-processing
The FTIR spectra were investigated, smoothed, and had their baseline correction and normalization performed with the

academic freeware software *SpectraGryph 1.2.14.* The spectra were normalized to the most intense band in the fingerprint region 850–1850 cm^{-1}.

Chemometrics
The PCA and HCA were performed using *SIMCA 14* software. The discrimination was performed in the fingerprint spectral region 850–1850 cm^{-1}.

All spectra were smoothed and denoised by a Savitzky – Golay filter (polynomial order 5 and points 15) and the second derivative of the samples was recorded. PCA was used to identify the dominant clusters in the data set (Davis and Mauer, 2010). For the hierarchical cluster analysis, Ward's algorithm was used (Murtagh and Legendre, 2014). We performed an unsupervised multivariate analysis technique because this method does not require a dependent variable for modeling, it searches for patterns among the independent variables, and groups of samples are formed based on the structure of the variables (Anzanello et al., 2014).

RESULTS

Figure 1 shows the characteristic FTIR PAS and DRIFT spectra of 5 different herbals: chamomile (A), silver birch (B), hibiscus (C), peppermint (D) and cornflower (E). The FTIR spectra of analyzed HMs occur in 2 spectral regions, showing organic matter and bonds in the sample (850–1850 cm^{-1} and 2700–3200 cm^{-1}).

FIGURE 1 | FTIR PAS spectra **(A)** and FTIR DRIFT spectra **(B)** of herbal medicines: chamomile (*Matricaria recutita*), silver birch (*Betula pendula Roth*), hibiscus (*Hibiscus sabdariffa*), peppermint *(Mentha piperita) and* cornflower *(Centaurea cyanus)*. The fingerprint region 800–850 cm^{-1} is colored in blue.
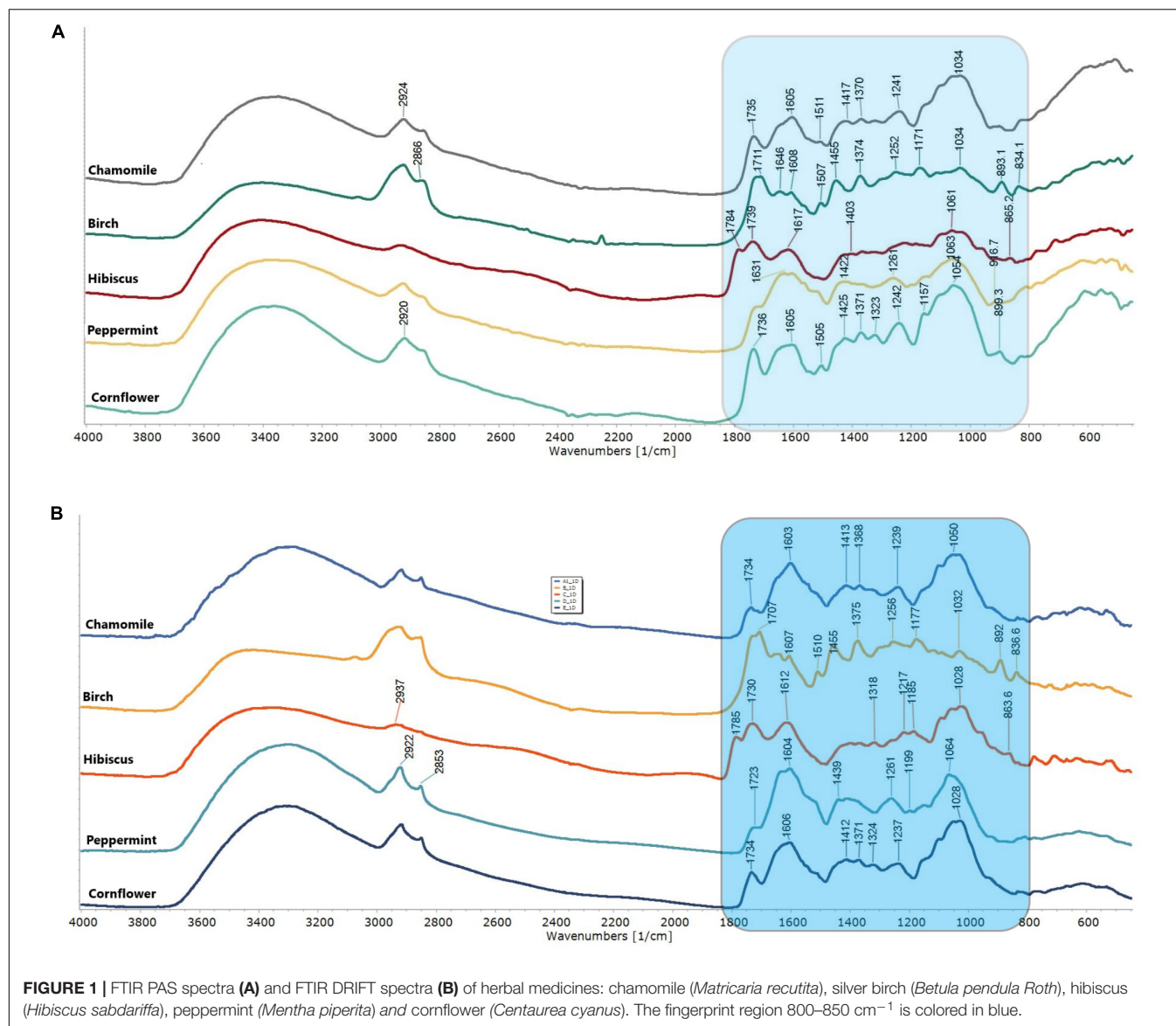
The recorded spectral result shows distinctive spectral patterns in fingerprint region (850—1850 cm^{-1}), well defined water absorption band (-O-H stretch, 3200–3400 cm^{-1}), well defined, but not specific C-H related peaks (2924 and 2853 cm^{-1} –CH- and –CH$_2$-CH$_3$). While birch shows a more pronounced spectral pattern with well-defined lines. Chamomile and cornflower show high similarity and predictable lower distinction by PCA and HCA. Specific spectral lines have been identified and summarized in **Table 2**.

Generally, PAS spectra show higher "spectral noises," specifically in the 1950–2500 cm^{-1} area. However, these "noises" did not affect the FTIR spectra differentiation because they were located outside the fingerprint area and outside the analytically significant area of functional groups. PAS and DRIFT spectra not only show significant differences in spectral line intensities, but also show a similar spectral pattern, and can be directly compared with each other.

Validation and Repeatability of PAS and DRIFT

To validate the repeatability of the two FTIR sampling methods PAS and DRIFT in the field of HM, firstly, five PAS and DRIFT spectra of each sample were recorded for five very different herbals under the same measurement conditions:

- herbal with flowers native to and grown in Latvia:

  - chamomile (*Matricaria recutita*), 8 commercial tea samples;
  - cornflower *(Centaurea cyanus)*, 1 sample;

- herbal with buds native to and grown in Latvia - silver birch (*Betula pendula Roth*, B), 1 sample;
- herbal non-native to Latvia - hibiscus (*Hibiscus sabdariff*), 1 sample;

TABLE 2 | PAS and DRIFT FTIR main absorption bands for fingerprint region of Chamomile **(A)**, Cornflower **(E)**, Birch buds **(B)**, Hibiscus **(C)**, Peppermint **(D)** and assignments.

Sample	DRIFT, cm^{-1}	PAS, cm^{-1}	Reference, cm^{-1}	Bond, stretching	Functional group/compounds	References
Chamomile (A), Cornflower (E)	1734	1735	1737	C = O bond	Luteolin	Zou et al., 2005
	1603	1605	1628	C = O bond	Luteolin	Parlinska-Wojtan et al., 2016
	1374	1370	1375	C–N stretch	Aromatic amines I, II	Dehelean et al., 2007
	1239	1241	1250	C–H stretch and O–H deform. of carboxyl groups and bending of N–H bond	Amide III	Choong et al., 2016
	1102	1105	1109	–C–O	Terpenoids, flavones	Lehto et al., 2018
	1050	1052	1053	C–O–C stretch	Aromatic ethers and polysaccharides	Szymczycha-Madeja et al., 2013
Birch buds (B)	1730	1731	1735 1726–1730	C = O stretching C = O vibration	in unconjugated ketone, carbonyl, and aliphatic groups (xylan); esters, ketones, and aldehydes in hemicelluloses	
	1707	1711	1704 1655–1658	absorbed OH and conjugated CO, HOH, OH bending	in lignin or cellulose absorbed water	
	1455	1455	1453	CH$_2$ deformation stretching; CH deformation asymmetric in plane; HCH and OCH in plane bending vibration.	lignin and xylan, lignin and hemicelluloses,	
	1375	1374	1372 1368–1369	aliphatic CH stretching; CH deformation.	in methyl and phenol in cellulose and hemicelluloses, in plane CH bending;	
	1256	1252	1254 1261–1263	stretching of OCO; CO stretching; CO linkages.	guaiacyl ring; in lignin; in guaiacyl aromatic methoxyl groups	
	892	894	895	C-H COC, CCO, and CCH deformation stretching	out of plane glucose ring in cellulose and hemicelluloses and for guaiacyl rings in lignin	
Hibiscus (C)	1785	1784	1788	R$_1$R$_2$C = CH$_2$	Methylene, overtone of δ' CH (out of plane)	
	1730	1739	1743	vC = O	Ketone	
	1612	1617	1629	Ortho-CO-C$_6$H$_4$-OH	Influenced by I,M, and steric effects of substituent	
	1095	1097	1098	-C-H	In-plane CH pending modes	
	1063	1061	1065	-C-H	In-plane CH pending modes	
	1028	1030	1033	vs SO$_3$		

(Continued)

TABLE 2 | Continued

Sample	DRIFT, cm⁻¹	PAS, cm⁻¹	Reference, cm⁻¹	Bond, stretching	Functional group/compounds	References
Peppermint (D)	1723	1725	1725	C = O vibration	bonded conjugated ketones, aldehydes, quinines and esters	
	1630	1628	1628	C = C vibration of aromatic structures, C = O stretching	amide I and carboxylic acids	
	1439	1422	1441	O–H in plane bend, C–O stretch vibration	carboxylic acids carbonates	
	1261	1261	1250	C–H stretch and O–H deform. of carboxyl groups and bending of N–H bond	Amide III	
	1149	1150	1153	C–O–C stretch	Aromatic ethers and polysaccharides	

- herbal with peppermint leaves (*Mentha piperita*), 1 sample.

The FTIR spectra were taken in the wavenumber range 400–4000 cm^{-1} (**Figure 1**). The validation was performed in the fingerprint region 850–1850 cm^{-1}. The Pearson Product Moment Correlation coefficient r was applied as a support tool to interpret correlation in the fingerprint region at the 850–1850 cm^{-1} using OriginPro2017 software.

The calculated Pearson's correlations value r showed a high correlation for both DRIFT (0.982–0.994) and PAS (0.9992–0.9999) FTIR sampling techniques.

Pearson's correlation (r) was performed for 5 different herbals to obtain correlations between all spectra. The highest Pearson's r values with other herbals showed cornflower (avg 0.745; max 0.936; min 0.473), chamomile (avg 0.742; max 0.936; min 0.534) and peppermint (avg 0.570; max 794; min 0.429). Conversely, the lowest was silver birch (avg 0.505; max 0.584; min 0.429) and hibiscus (avg 0.619; max 0.701; min 0.485). This indicates a greater possibility of differentiation using FTIR sampling techniques, while also proving the high similarity of FTIR fingerprint pattern and highlights the need for a more sophisticated method of discrimination, such as statistical methods of PCA or HCA.

Differentiation of FTIR PAS and DRIFT Spectra

Combining the FTIR methods with unsupervised multidimensional statistical analysis, conclusions about the effect of the sampling method on the obtained results were obtained. To compensate for the differences in differences of the sampling techniques (DRIFT, PAS), second derivatives of fingerprint region spectra were used as input data. Spectral derivatives reduce the impact in differences in spectral sensitivity and peak width. The formation of clusters was depicted in diagrams and dendrograms in **Figure 2**. Three major clusters in PCA can be identified (**Figure 2A**). The analysis shows that the greatest influence on cluster formation is not the choice of the FTIR sampling method, but the specificity of HM spectra in the fingerprint region. The HCA shows that differentiation, according to FTIR sampling methods, is possible as well (**Figure 2B**). PCA1 describes 48.9%, but PCA 9.1%, forming a total of 58% of spectral information. Loadings (**Figure 2C**) shows a clear difference between PCA1 and PCA2, which provides a clear discrimination for a PCA1/PCA2 diagram and HCA.

Narrowing the research area, leaving only one HM, for instance, chamomile, **Figures 3A,B** demonstrate that spectra can be differentiated according to FTIR sampling methods as well.

Differentiation of HMs Leaves vs. Flowers

The next point of interest was the possibility to differentiate parts of the HM (flowers, leaves, stems). This is a very important factor in the production and quality control of HMs because not always can production regulations specify a proportion between flowers and leaves in dried HM.

In our experiment, we tested both FTIR spectra of HM leaf and flower powders and HM leaves and flower ethanol and

FIGURE 2 | The differentiation of FTIR PAS and DRIFT spectra of herbal medicines: chamomile [*Matricaria recutita,* (code – A)], silver birch [*Betula pendula Roth* (B)], hibiscus [*Hibiscus sabdariffa* (C)], peppermint *[Mentha piperita* (D)], cornflower *[Centaurea cyanus* (E)]. **(A)** The PCA clusters for PAS and DRIFT FTIR sampling methods. **(B)** The HCA dendrogram for PAS and DRIFT FTIR sampling methods. **(C)** Loadings for PAS and DRIFT FTIR sampling methods.

acetone extracts. **Figures 4A,B** demonstrate separate clusters for leaves and flowers. **Figure 4A** illustrates a cluster formation for meadowsweet Ms *(Filipendula ulmaria)* leaf and flower powders. **Figure 4B** shows separate clusters for tansy Ta *(Tanacetum vulgare)* leaf and flower extracts. Moreover, a correlation can be

seen between the powder used to make the extract and the leaf or flower extract itself. It demonstrates that the FTIR sampling methods can be used for the differentiation of leaves and flowers in both solid and liquid samples.

Comparison of Dried Herbals vs. Extracts in Ethanol

Traditional extracts are made from dried plant material extracted by the appropriate liquid. Method No. 2 was used to prepare 5 extracts of chamomile Ch (*Matricaria recutita*) sourced from different producers. After extraction, the solid residual was collected and dried. All samples (5 powders before extraction, 5 evaporated extracts, and 5 dried residual powders after extraction) measured with the DRIFT method. The fingerprint pattern for powders shows a great deal of similarities. FTIR spectra for extract shows sharper peaks and higher intensity. For evaluation of FTIR spectra, PCA statistical analyses were applied. The resulting PCA1 (71%)/PCA2 (13%) scatter plot (**Figure 5**) shows characteristics of clusterification according to the sample type. Dispersion across the PCA1 axis could be described by the concentration of the mobile phase in samples. Reducing the concentration in the residual powder after extraction shows a slight shift to the right, and a strong shift in the opposite direction for extracts. The shift intensity correspondent was predicted to be stronger for samples with a higher content of flowers.

DISCUSSION

A comparison between spectra recorded by PAS and DRIFT sampling methods showed high sensitivity and good discrimination of herbal species based on spectral information. The high complexity of chemical composition and similarity of the main structure of herbal materials adds the complexity of FTIR the spectra interpretation. Traditional methods of spectral interpretation, spectral library search, and line position identification, give very limited information or do not have any practical functionality. The direct comparison of spectral patterns indicates reproducible spectral fingerprints, which could be used for a more sophisticated statistical examination by PCA and HCA methods. The results obtained provide information about the spectral behavior of homogenized herbal and herbal extracts and can be used for establishing identification and discrimination criteria. It has been demonstrated that PAS and DRIFT, in chemometrics, can be a useful experimental tool for the characterization and discrimination of herbals. Also, it must be mentioned that there is a high reproducibility of the PAS FTIR method, but it also demands a high-cost sampling cell and a more advanced FTIR spectrometer with a high-intensity beam and secondary detector connection. DRIFT method gives lower reproducibility, while proving to be more versatile and could be used for powder and liquid samples, while not demanding any specific requirements for the FTIR spectrometer. Despite the high similarity of the fingerprint pattern in the region 850–1850 cm^{-1}, the described approach gives promising results for the identification, discrimination, and

FIGURE 3 | The differentiation of FTIR PAS and DRIFT spectra of chamomile Ch (*Matricaria recutita*). **(A)** The PCA clusters for PAS (blue) and DRIFT (green) FTIR sampling methods. **(B)** The HCA dendrogram for PAS (blue) and DRIFT (green) FTIR sampling methods.

FIGURE 4 | The differentiation of HMs leaves vs. flowers. **(A)** The PCA clusters for meadowsweet Ms *(Filipendula ulmaria)* leaf and flower powders. **(B)** The PCA clusters for tansy Ta *(Tanacetum vulgare)* leaf and flower extracts in ethanol and acetone. Green – leaves and stems; blue- flowers; red – leaf extracts and yellow – flower extracts.

FIGURE 5 | Differentiation HMs powder before extraction vs. extracts in ethanol vs. after extraction. The PCA clusters for chamomile powder before and after extraction and ethanol.

CONCLUSION

Research shows a significant potential of the FTIR sampling methods PAS and DRIFT for a fast and sensitive, reproducible, and non-destructive method for the quality control of HM in a form of powder or liquid extracts. The experiments conducted prove a high sensitivity and capabilities of FTIR method in combination with chemometrics to discriminate important characteristics of MH, such as the identification of plant parts, the concentration of extractable compounds in MH, and differentiation of samples by types. PAS sampling gives unmatchable reproducibility (Pearson's *r* value = 0.999) for different herbal material but has lower sensitivity and higher spectral noise. A higher sample area analyzed by PAS also improve sampling reproducibility reducing the influence of natural sample inhomogeneity observed by the DRIFT method. PAS gives a clear advantage of high clusterification showed by PCA and HCA. The DRIFT method shows higher versatility for analyzing powder and liquid sample, but lower reproducibility for sampling and spectral measurements (Pearson's *r* value 0.982–0.994). Both methods show high potential for further research. Additionally, the disadvantages of PAS should be mentioned: higher cost of purchases, use of He gas and a more complex sampling routine.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

characterization of MH. Unfortunately, the complexity of the FTIR pattern and limited information about FTIR possibilities for HM limits FTIR for wider usage in HM research. Also, the traditional Diamond ATR sampling cell's low sensitivity after 1400 cm^{-1} adds another obstacle for broader usage, therefore DRIFT, and PAS prove to be a more appropriate sampling method. The results of the usage of the PAS sampling method gives a high hope for future research: no sampling preparation needed, exceptional reproducibility and comparable spectral pattern with other sampling methods. On the other hand, it also has a few drawbacks: higher costs of PAS cell, lower S/N, and spectral resolution. Also, it should be mentioned that the 8 cm^{-1} resolution used for PAS measurements allows for a better signal to noise.

AUTHOR CONTRIBUTIONS

FUNDING

REFERENCES

Anzanello, M. J., Ortiz, R. S., Limberger, R., and Mariotti, K. (2014). Performance of some supervised and unsupervised multivariate techniques for grouping authentic and unauthentic viagra and cialis. *Egypt. J. Forensic Sci.* 4, 83–89. doi: 10.1016/j.ejfs.2014.03.004

Bostijn, N., Hellings, M., Van Der Veen, M., Vervaet, C., and De Beer, T. (2018). In-line UV spectroscopy for the quantification of low-dose active ingredients during the manufacturing of pharmaceutical semi-solid and liquid formulations. *Anal. Chim. Acta* 12, 54–62. doi: 10.1016/j.aca.2018.02.007

Bunaciu, A. A., Aboul-Enein, H. Y., and Fleschin, S. (2011). Recent applications of Fourier transform infrared spectrophotometry in herbal medicine analysis. *Appl. Spectrosc. Rev.* 46, 251–260. doi: 10.1080/05704928.2011.565532

Choong, Y.-K., Yousof, N., Wasiman, M. I., Jamal, J. A., and Ismail, Z. (2016). Determination of effects of sample srocessing on Hibiscus sabdariffa L. using tri-step infrared spectroscopy. *J. Anal. Bioanal. Tech.* 7:335. doi: 10.4172/2155-9872.1000335

Davis, R., and Mauer, L. (2010). Fourier transform infrared (FT-IR) spectroscopy: a rapid tool for detection and analysis of foodborne pathogenic bacteria. *Curr. Res. Technol. Educ. Top.* 2, 1582–1594.

Dehelean, C., Pinzaru, S. C., Peev, C., Soica, C., and Antal, D. S. (2007). Characterization of birch tree leaves, buds and bark dry extracts with antitumor activity. *J. Optoelectron. Adv. M.* 9, 783–787.

Dias, R. C. E., Valderrama, P., Março, P. H., Scholz, M. B. D. S., Edelmann, M. J., and Yeretzian, C. (2018). Data on roasted coffee with specific defects analyzed by infrared-photoacoustic spectroscopy and chemometrics. *Data Brief.* 20, 242–249. doi: 10.1016/j.dib.2018.08.013

Eniņa, V. (2017). *Veselība Pie Mājas Sliekšņa. 100 Populārākie Ārstniecības Augi Latvijā.* Riga: Zvaigzne ABC.

Gasera (2010). FTIR–PAS Photoacoustic Fourier transform infrared spectroscopy (FTIR–PAS). Available at: http://www.gasera.fi/technology/

Gemperline, P. J. (2006). *Practical Guide to Chemometrics: Ed. by Paul Gemperline*, 2nd Edn. Boca Raton: CRC.

Grdadolnik, J. (2002). ATR-FTIR spectroscopy: its advantages and limitations. *Acta Chim. Slov.* 49, 631–642.

Greene, E. F., Tauch, S., Webb, E., and Amarasiriwardena, D. (2004). Application of diffuse reflectance infrared fourier transform spectroscopy (DRIFTS) for the identification of potential diagenesis and crystallinity changes in teeth. *Microchem. J.* 76, 141–149. doi: 10.1016/j.microc.2003.11.006

Heinrich, M. (2015). Quality and safety of herbal medical products: regulation and the need for quality assurance along the value chains. *Br. J. Clin. Pharmacol.* 80, 62–66. doi: 10.1111/bcp.12586

Izam, A. B. (2012). *Solid-liquid Extraction of Hydrolysable Tannin (galic acid) from Stem Bark of Jatropha Curcas Using Various Type of Extraction.* [dissertation],: University of Malaysia Pahang, Malaysia.

Kadiroğlu, P., Aydemir, L. Y., and Akcakaya, F. G. (2018). Prediction of functional properties of registered chickpea samples using FT-IR spectroscopy and chemometrics. *LWT* 93, 463–469. doi: 10.1016/j.lwt.2018.03.080

Kauppinen, J., Wilcken, K., Kauppinen, I., and Koskinen, V. (2004). High sensitivity in gas analysis with photoacoustic detection. *Microchem. J.* 76, 151–159. doi: 10.1016/j.microc.2003.11.007

Kazarian, S. G., and Chan, K. L. (2006). Applications of ATR-FTIR spectroscopic imaging to biomedical samples. *Biochim. Biophys. Acta.* 1758, 858–867. doi: 10.1016/j.bbamem.2006.02.011

Kitanov, G., Karcheva-Bahchevanska, D., and Lukova, P. (2015). Comparative analysis of monographs on herbal drugs and herbal drug preparations included in the European Pharmacopoeia (Ph. Eur. 8). *Annales UMCS Pharmacia* 62, 20–27.

Kuusela, T., and Kauppinen, J. (2007). Photoacoustic gas analysis using interferometric cantilever microphone. *Appl. Spectrosc. Rev.* 42, 443–474. doi: 10.1080/00102200701421755

Legner, N., Meinen, C., and Rauber, R. R. (2018). Differentiation of agricultural plant cultivars and proveniences using ftir spectroscopy. *Front. Plant Sci.* 9:748. doi: 10.3389/fpls.2018.00748

Lehto, J., Louhelainen, J., Kłosinńska, T., Drożdżek, M., and Alén, R. (2018). Characterization of alkali-extracted wood by FTIR-ATR. Spectroscopy. *Biomass Convers. Biorefin.* 8, 847–855. doi: 10.1007/s13399-018-0327-5

Miaw, C. S. W., Assis, C., Silva, A. R. C. S., Cunha, M. L., Sena, M. M., and de Souza, S. V. C. (2018). Determination of main fruits in adulterated nectars by ATR-FTIR spectroscopy combined with multivariate calibration and variable selection methods. *Food Chem.* 15, 272–280. doi: 10.1016/j.foodchem.2018.02.015

Moros, J., Garrigues, S., and Guardia, M. (2010). Vibrational spectroscopy provides a green tool for multi-component analysis, TRAC-Trend. *Anal. Chem.* 29, 578–591. doi: 10.1016/j.trac.2009.12.012

Murtagh, F., and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion. *J. Classif.* 31, 274–295. doi: 10.1007/s00357-014-9161-z

Nelce, M. M., Mahendradatta, M., Laga, A., and Djide, N. (2013). Tannin extract of guava leaves (*Psidium guajava* L) variation with concentration organic solvents. *Int. J. Sci. Eng. Technol.* 2, 106–110.

Parlinska-Wojtan, M., Kus-Liskiewicz, M., Depciuch, J., and Sadik, O. (2016). Green synthesis and antibacterial effects of aqueous colloidal solutions of silver nanoparticles using camomile terpenoids as a combined reducing and capping agent. *Bioprocess Biosyst. Eng.* 39, 1213–1223. doi: 10.1007/s00449-016-1599-1594

Peerapattana, J., Otsuka, K., Hattori, Y., and Otsuka, M. (2015). Quantitative analysis of α-mangostin in hydrophilic ointment using near-infrared spectroscopy. *Drug Dev. Ind. Pharm.* 41, 515–521. doi: 10.3109/03639045.2014.884115

Rehrl, J., Karttunen, A. P., Nicolai, N., Hörmann, T., Horn, M., Korhonen, O., et al. (2018). Control of three different continuous pharmaceutical manufacturing processes: use of soft sensors. *Int. J. Pharm.* 16, 60–72. doi: 10.1016/j.ijpharm.2018.03.027

Rodriguez-Saona, L. E., and Allendorf, M. E. (2011). Use of FTIR for rapid authentication and detection of adulteration of food. *Annu. Rev. Food Sci. Technol.* 2, 467–483. doi: 10.1146/annurev-food-022510-133750

Rohman, A., Nugroho, A., Lukitaningsih, E., and Sudjadi. (2014). Application of vibrational spectroscopy in combination with chemometrics techniques for authentication of herbal medicine. *Appl. Spectrosc. Rev.* 49, 603–613. doi: 10.1080/05704928.2014.882347

Rohman, A., Windarsih, A., and Hossain, M. A. M. (2019). Application of near- and mid-infrared spectroscopy combined with chemometrics for discrimination and authentication of herbal products: a review. *J. Appl. Pharm. Sci.* 9, 137–147. doi: 10.7324/JAPS.2019.90319

Smith, B. C. (2011). *Fundamentals of Fourier Transform Infrared Spectroscopy*, 2nd Edn. Boca Raton, FL: CRC Press, doi: 10.1201/b10777

Stuart, B. H. (2004). *Infrared Spectroscopy: Fundamentals and Applications.* Hoboken, NJ: Wiley.

Szymczycha-Madeja, A., Welna, M., and Zyrnicki, W. (2013). Multi-Element analysis, bioavailability and fractionation of herbal tea products. *J. Braz. Chem. Soc.* 24, 777–787. doi: 10.5935/0103-5053.20130102

Uotila, J., and Kauppinen, J. (2008). Fourier transform infrared measurement of solid-, liquid-, and gas-phase samples with a single photoacoustic cell. *Appl. Spectrosc.* 62, 655–660. doi: 10.1366/000370208784658048

Wulandari, L., Retnaningtyas, Y., Lukman, N., and Lukman, H. (2016). Analysis of flavonoid in medicinal plant extract using infrared spectroscopy and chemometrics. *J. Anal. Methods Chem.* 2016:4696803. doi: 10.1155/2016/4696803

Yang, I.-C., Tsai, C.-Y., Hsieh, K.-W., Yang, C.-W., Ouyang, F., Lo, Y. M., et al. (2013). Integration of SIMCA and near-infrared spectroscopy for rapid and precise identification of herbal medicines. *J. Food Drug Anal.* 21, 268–278. doi: 10.1016/j.jfda.2013.07.008

Zou, H.-B., Yang, G.-S., Qin, Z.-R., Jiang, W.-Q., Du, A.-Q., and Aboul-Enein, H. Y. (2005). Progress in quality control of herbal medicine with IR fingerprint spectra. *Anal. Lett.* 38, 1457–1475. doi: 10.1081/AL-200062153

Check for
updates

From the Soft to the Hard: Changes in Microchemistry During Cell Wall Maturation of Walnut Shells

Nannan Xiao[1], Peter Bock[1], Sebastian J. Antreich[1], Yannick Marc Staedler[2], Jürg Schönenberger[2] and Notburga Gierlinger[1]*

[1] *Institute of Biophysics, Department of Nanobiotechnology, University of Natural Resources and Life Sciences, Vienna, Austria,* [2] *Division of Structural and Functional Botany, Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria*

The walnut shell is a hard and protective layer that provides an essential barrier between the seed and its environment. The shell is based on only one unit cell type: the polylobate sclerenchyma cell. For a better understanding of the interlocked walnut shell tissue, we investigate the structural and compositional changes during the development of the shell from the soft to the hard state. Structural changes at the macro level are explored by X-ray tomography and on the cell and cell wall level various microscopic techniques are applied. Walnut shell development takes place beneath the outer green husk, which protects and delivers components during the development of the walnut. The cells toward this outer green husk have the thickest and most lignified cell walls. With maturation secondary cell wall thickening takes place and the amount of all cell wall components (cellulose, hemicelluloses and especially lignin) is increased as revealed by FTIR microscopy. Focusing on the cell wall level, Raman imaging showed that lignin is deposited first into the pectin network between the cells and cell corners, at the very beginning of secondary cell wall formation. Furthermore, Raman imaging of fluorescence visualized numerous pits as a network of channels, connecting all the interlocked polylobate walnut shells. In the final mature stage, fluorescence increased throughout the cell wall and a fluorescent layer was detected toward the lumen in the inner part. This accumulation of aromatic components is reminiscent of heartwood formation of trees and is suggested to improve protection properties of the mature walnut shell. Understanding the walnut shell and its development will inspire biomimetic material design and packaging concepts, but is also important for waste valorization, considering that walnuts are the most widespread tree nuts in the world.

Keywords: cell wall, pectin, lignification, FTIR microscopic imaging, Confocal Raman microscopy, microchemistry, sclerenchyma, *Juglans regia*

INTRODUCTION

The nut is commonly defined as a dry, indehiscent, usually one-seed fruit with a hard and tough endocarp (shell) enclosing the seed, which develops from a simple ovary. The hardened endocarp of the nut provides a physical barrier around the seed and protects the embryo against biotic and abiotic factors in the natural environment (Dardick and Callahan, 2014). This remarkable design can be attributed to natural selection in the course of evolution (Sallon et al., 2008).

The walnut (*Juglans regia L.*), also known as English or Persian walnut, is the most widespread tree nut in the world, and is thus an economically important tree species (De Rigo et al., 2016). World production of walnuts exceeds three million tons since 2012 and took over almond and hazelnut (Bernard et al., 2018). Studies on nuts have proven their nutritional importance (Martinez et al., 2010), through antioxidant activities (Jahanbani et al., 2016; Panth et al., 2016). In a sustainable economy we furthermore can benefit from an optimized utilization of the nut waste—the shell and the husk—and for this, in-depth knowledge of these materials is of vital importance.

The rapid progress of genome sequencing of *Juglans* species was reviewed recently and will pave the way for functional genomics research (Chen et al., 2019). Fruit growth and development are a major research interest (Pinney and Polito, 1983; Wu et al., 2009) as well as the composition and nutritional value of the seed (Fukuda et al., 2003; Kornsteiner et al., 2006; Zhang et al., 2009; Martinez et al., 2010). In recent years, the walnut shell has also been studied due to its potential for the production of bioethanol (Yang et al., 2015; Lancefield et al., 2017), pyroligneous acid (Jahanban-Esfahlan and Amarowicz, 2018), charcoal and activated carbon (Xie et al., 2013) and "nutty carbon," which is used for Na-ion battery anodes (Wahid et al., 2017). For novel applications and material development, fundamental knowledge of the structure as well as the chemistry of the shell is needed. Recently, a polylobate cell shape with interlocked packing was found to enable the superior mechanical properties of walnut shells (Antreich et al., 2019). All cells are connected via numerous pits (Reis et al., 1992; Antreich et al., 2019), which maintain symplastic connection by a recess of the cell wall (Reis et al., 1992). Chemical studies of the cell wall have shown that lignin is a main component (above 50%), followed by cellulose (25%), and hemicelluloses (22%) (Demirbas, 2005). During the differentiation of the walnut shell, the lignin content of the endocarp increases gradually, as shown by chemical analysis (Zhao et al., 2016). However, measuring total chemical composition of cellulose, hemicelluloses and lignin content typically requires tissue disruption and pretreatment to separate it from the cell wall matrix (Chen et al., 2015; Lancefield et al., 2017; Shah et al., 2018). Histochemical staining gives information in context with the structure (Li et al., 2018), but often lacks sensitivity among chemically similar components (Simon et al., 2018). For advanced understanding of the distribution of cell wall substances in the nutshell in context with the microstructure we explore the feasibility of vibrational microspectroscopy and imaging.

Vibrational spectroscopy methods such as Fourier-transform infrared spectroscopy (FT-IR) and Raman spectroscopy are increasingly used for chemical analysis in plant research, because they are fast, noninvasive, nondestructive, and require only limited sample preparation (Felten et al., 2015; Gierlinger, 2018). The two techniques often provide complementary information about the molecular vibrations of a given sample due to different energy transfers and thus different selection rules (Smith and Dent, 2005). FT-IR microspectroscopic imaging combines imaging with spectral information in a spatial context, which can provide an overview of all major chemical components of cell walls with a spatial resolution of about $10\,\mu$m (Mccann and Carpita, 2008; Mazurek et al., 2013; Yang et al., 2018). For chemical imaging with high spatial resolution Raman microscopy has shown a high potential via the selective acquisition of spectra from different cell wall regions at the sub-micron level (250 nm) (Gierlinger and Schwanninger, 2006, 2007; Gierlinger et al., 2012). Advances in laser technology, filter efficiency, CCD sensitivity and superior optics currently enable instruments to record signal with a good signal/noise ratio, which, in turn, paved the way for fast and sensitive scanning (Agarwal, 2006; Gierlinger et al., 2012). A Raman image is composed of thousands of pixels, and a complete spectrum can be acquired from each pixel. By selecting spectral positions which are unique to an individual chemical component, their spatial distribution can be visualized by an intensity heat map (Gierlinger, 2018). Microspectroscopy tools are thus powerful for characterizing dynamic chemical changes in the cell wall during development and maturation of fruits and vegetables (Chylinska et al., 2017).

The aim of this study is to understand the structure and composition of the mature walnut shell and its development. Analyzing chemical changes of nutshells in the context of microstructure during maturation (lignification) by FT-IR and Confocal Raman Microscopy will provide valuable insights into composition of the interlocked polylobate walnut cells and their "fabrication." These results will inspire biomimetic material developments and promote the utilization of walnut shells for new products.

MATERIALS AND METHODS

Plant Materials
Walnut (*Juglans regia*, cultivar "Geisenheim") fruits were collected from various positions on the same tree, growing in the "BOKU horticulture Jedlersdorf" in Vienna, Austria once every month from June to October in 2017. At each developmental stage, ten similar sized fruits were selected, harvested and instantly stored at $-20°$C in plastic bags.

Sample Preparation
Hand-cut equatorial sections of fruits from five developmental stages were cut with a razor blade and a saw. Then, $8\,\mu$m-thick consecutive sections from July and October samples were cut with a cryostat microtome (CM3050S, Leica Biosystems Nussloch GmbH, Wetzlar, Germany) and rotary microtome (RM2235, Leica Biosystems Nussloch GmbH, Wetzlar, Germany), respectively. For FT-IR measurements, the sections were transferred onto a standard glass slide and dried in a desiccator for at least 48 h, then placed between two CaF$_2$ windows (22 mm diameter, 0.5 mm thick) for FT-IR micro-spectroscopic imaging. The remaining consecutive sections were placed on a standard glass slide with a drop of distilled water, covered with a glass coverslip (0.17 mm) and sealed with nail polish. To test whether extractives are the reason for autofluorescence of the October samples the sections were treated with distilled water and 50% ethanol at $70°$C for 48 h, respectively (as suggested by Queirós et al., 2019) and then measured by Confocal Raman Microscopy.

Micro-Computed Tomography (Micro-CT)

Frozen walnuts from June to October (whole nuts with green husk) were scanned in an X-ray micro-computed tomograph (MicroXCT-200, Zeiss Carl Zeiss, Jena, Germany). Samples were put in a double-walled glass container and covered with Parafilm to keep the temperature low during the scan. Fast scans were performed (exposure time 1.0–1.3 s with binning 4) to minimize the artifacts from heating during the scan. Scanning parameters were set to 50 kV tube voltage and 160 μA current for the June to September samples, 40 kV and 200 μA for the October samples. Each scan consisted of ∼500 2D radiographs with a voxel size of 89–100 μm^3, depending on the total size of the sample. 3D data construction was performed via the software XMReconstructor 8.1.6599 (Zeiss Carl Zeiss, Jena, Germany).

Histological Analysis

For histochemical staining, phloroglucinol (Wiesner) staining was performed as described in Yeung (1998). Briefly, hand-cut equatorial sections of five developmental stages were left for 20 min in phloroglucinol-HCl staining solution [20 mg/mL phloroglucinol in 20% ethanol and mixed with 12 N HCl (v:v=80:20)], and immediately photographed with a Canon EOS M10 fitted with a macro lens (35 mm, f/2.8 Canon Inc., Tokyo, Japan). The Wiesner reagent (phloroglucinol-HCl) mainly reacts with O-4-linked coniferyl and sinapyl aldehydes in lignifying cell walls (Pomar et al., 2002) and was used to follow the onset of lignification.

Sections from July and October nuts were stained with Fuchsin-Chrysoidine-Astrablue (FCA) solution [0.1 mg/mL of New Fuchsin, 0.143 mg/mL Chrysoidine, 1.25 mg/mL Astra blue and acetic acid (v:v=1:50)], and then washed step by step with distilled water, ethanol (30, 70%) and isopropanol. The samples were immersed in the FCA-solution for 48 h to guarantee a full penetration of the staining into the dense microstructure of the cell walls. Stained sections were embedded in Euparal and photographed under a Labophot-2 microscope (Nikon Corporation, Tokyo, Japan).

Fourier-Transform Infrared (FTIR) Spectroscopy

A Vertex 70 Fourier-transform infrared (FT-IR) spectrometer, coupled to a Hyperion 2000 microscope (15× objective), which was equipped with a liquid nitrogen-cooled MCT-D316-025 (mercury cadmium telluride) detector and KBr beam-splitter (Bruker Optik GmbH, Ettlingen, Germany), was used to perform FT-IR mapping with an automated XY motorized stage. Both visible and spectroscopic imaging data of the sections were acquired at room temperature in transmission mode over the range of 4,000–700 cm^{-1} at an aperture size of 25 × 25 μm. Absorbance spectra were acquired at a spectral resolution of 8 cm^{-1}. A rectangle area was selected to collect the FT-IR images of July (250 × 1500 μm) and October (300 × 1950 μm) samples, covering outer to inner tissues. A background spectrum of the CaF$_2$ surface was collected before measuring the samples. The Opus 6.5 software (Bruker Optik GmbH, Ettlingen, Germany) enabled control of the microscope and collection of the spectra from the samples. Spectral data processing and image acquisition

was performed using ImageLab (EPINA GmbH, Pressbaum, Austria) software, and the chemical mapping was displayed based on the intensity of functional groups. Then all the spectra were truncated to the finger print region (1,800–800 cm^{-1}), baseline corrected and normalized using Opus 7.5 software (Bruker Optik GmbH, Ettlingen, Germany). Principal Component Analysis (PCA) was performed by Unscrambler X 10.3 software (CAMO Software AS, Oslo, Norway).

Confocal Raman Microscopy

Raman spectra were acquired from microsections using a confocal Raman microscope (alpha300RA, WITec Ulm, Germany) equipped with a 100× oil immersion objective (NA 1.4, Carl Zeiss, Jena, Germany) and a motorized XYZ stage. A linear polarized (0°) coherent compass sapphire green laser (λ_{ex} =532 nm, WITec, Ulm, Germany) was passed through a polarization-preserving single-mode optical fiber and focused through the objective with a spatial resolution of 0.3 μm on the sample. The Raman scattering signal passed a multi-mode fiber (50 μm diameter) and was detected by a CCD camera (Andor DV401 BV, Belfast, UK) behind a spectrometer (600 g mm^{-1} grating, UHTS 300 WITec). The laser power used on the July and October section was 34.7 and 10 mW, respectively. The sample was mapped by collecting single spectra at every image pixel, with an integration time of 0.04 s per spectrum, a spectral resolution of 4 cm^{-1} in the range of 3,800–300 cm^{-1}. The monochromator of the spectrometer was calibrated using the Raman scattering line produced by a silicon plate (520 cm^{-1}). For measurement setup the Control Four (WITec, Ulm, Germany) was used.

Raman data analysis was performed with Project Four (WITec, Ulm, Germany) and Opus 7.5 software (Bruker Optik GmbH, Ettlingen, Germany). After applying cosmic ray spike removal, Raman chemical images were generated based on the integration of relevant wavenumber regions (e.g., CH stretching, lignin, cellulose, pectin, and fluorescence), which enabled definition of cell wall areas of interest and calculation of average spectra from these regions. All calculated spectra were exported into Opus 7.5 for comparison and baseline correction.

RESULTS

Morphological Analysis of Shell Development

Walnut fruits completed their development and ripening within six months (May-October 2017) (**Figure 1**). The spatial organization of the walnut at five developmental stages was visualized in 3D scans (**Supplementary Videos 1–5**) based on micro-CT data. The reconstructed 3D models of the walnut fruit allow for the dissection of the tissues into 2D virtual sections (**Figure 1A**). Due to differences in tissue density the developing seed and the different shell layers can be visualized in the different stages. Dense tissues appear light, while dark areas represent air space (Verboven et al., 2012). In June the outer layer appeared homogenous, but in July a distinct light concentric layer between the outer layer and the seed was apparent (see **Figure 1A**). This layer with higher density represented the onset of nutshell
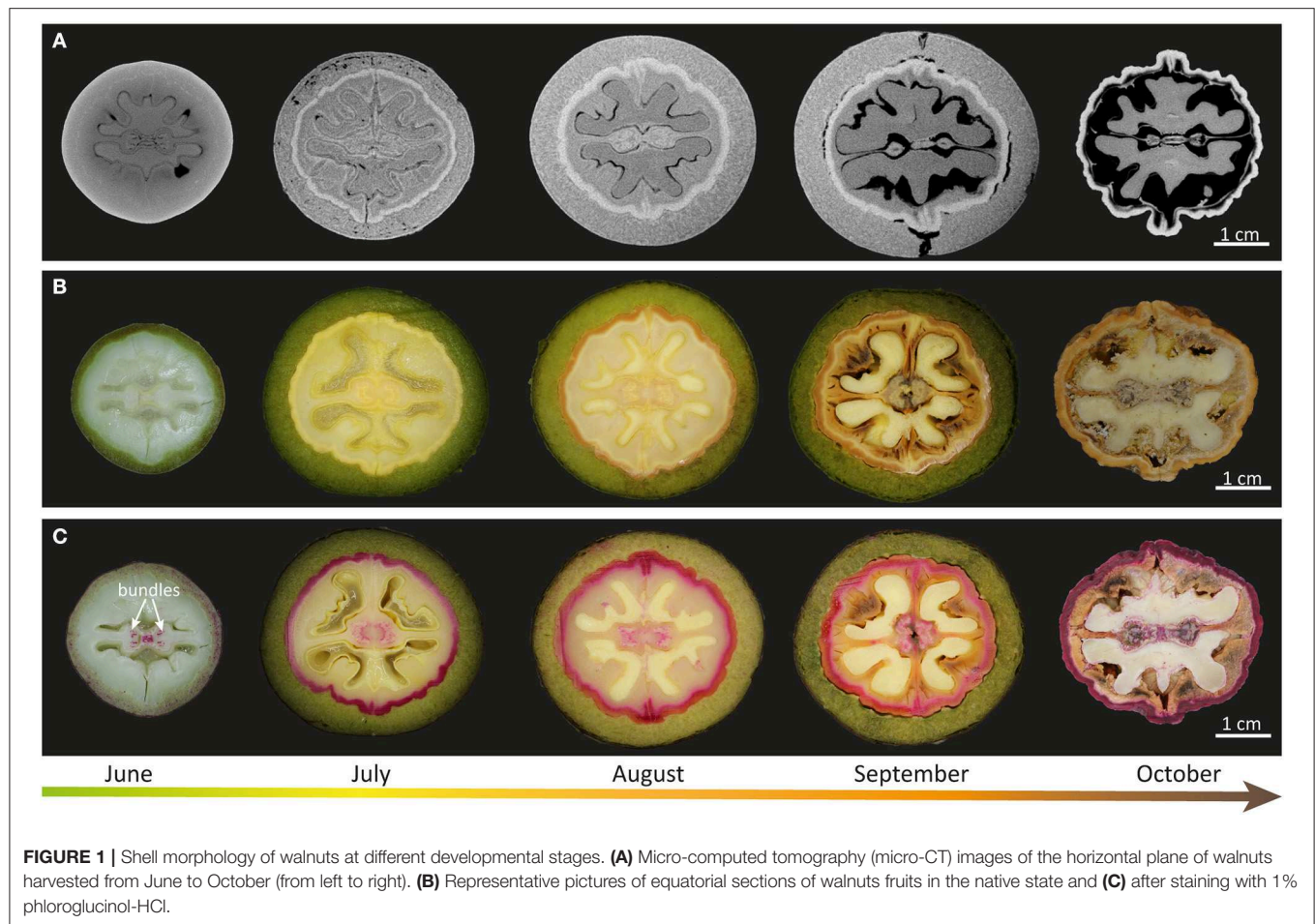
FIGURE 1 | Shell morphology of walnuts at different developmental stages. **(A)** Micro-computed tomography (micro-CT) images of the horizontal plane of walnuts harvested from June to October (from left to right). **(B)** Representative pictures of equatorial sections of walnuts fruits in the native state and **(C)** after staining with 1% phloroglucinol-HCl.

differentiation. The shell region became thicker during tissue maturation and formed a seal around the seed. Upon maturation (in September) more and more air (dark regions) was found around the seed and also between the shell and the green outer husk, which finally was shed in October (**Figures 1A,B**). In the photographic images of the equatorial sections, the onset of shell differentiation was apparent as a light yellow layer (**Figure 1B**). During maturation the shell became brown: first close to the green husk (August) and later uniformly brown (October) (**Figure 1B**). The onset of lignification based on Wiesner staining was detected close to the green husk as a pink layer (**Figure 1C**). In August, this concentric tissue layer underlying the green husk became more red stained, while the inner part was still pink. The strong lignification of the shell was clearly visible until the final stage in October, although the shades of red changed during development (**Figure 1C**).

Fuchsin-Chrysoidin-Astra (FCA) staining of microsections confirmed that lignification began in July in the outer part of the shell (**Figures 2A,B**), while the very thin walls of the inner part were stained only blue and thus not yet lignified (**Figures 2A,C**). The October sample was stained even more reddish and the cell wall thickness increased, especially in the outer part of the shell (**Figures 2D,E**). In the very inner part the cell walls were thin, but stained red and were thus also lignified (**Figure 2F**).

FT-IR Imaging of Walnuts Shell Micro-Sections

Fourier-Transform Infrared (FT-IR) microscopy was applied to probe the chemical composition of the walnut shells based on the absorption at particular infrared light frequencies corresponding to specific chemical bonds and groups of the different cell wall polymers (**Figure 3**). The strongest bands between 1,015 and 1,060 cm^{-1} are assigned to C-O stretching vibrations of cellulose (Fengel and Ludwig, 1991); therefore, the intensity at 1036 cm^{-1} was used to monitor cellulose distribution (**Figures 3A,B**). For hemicelluloses the absorption band at 1,734 cm^{-1} is indicative and assigned to C=O stretching vibration of acetyl groups (Faix, 1991) and here used to follow the hemicelluloses. The band centered at 1,504 cm^{-1} is assigned to aromatic skeletal vibrations of lignin (Faix, 1991) and used for imaging the aromatic components in the shell. The intensity-based color-coded maps of the three cell wall polymers, cellulose, hemicelluloses, and lignin allowed us to visualize the changes across the shell (outer to inner part) for samples collected in July (**Figure 3A**) and October (**Figure 3B**). In the July sample (**Figure 3A**), the top map reveals that the highest cellulose band intensity (yellow) lies in the outer part of the shell (nine-fold higher than that of the inner part). The maps for hemicelluloses (1734 cm^{-1}) and lignin (1,504 cm^{-1}) showed a similar intensity pattern to those obtained for cellulose (**Figure 3A**). In contrast, in October
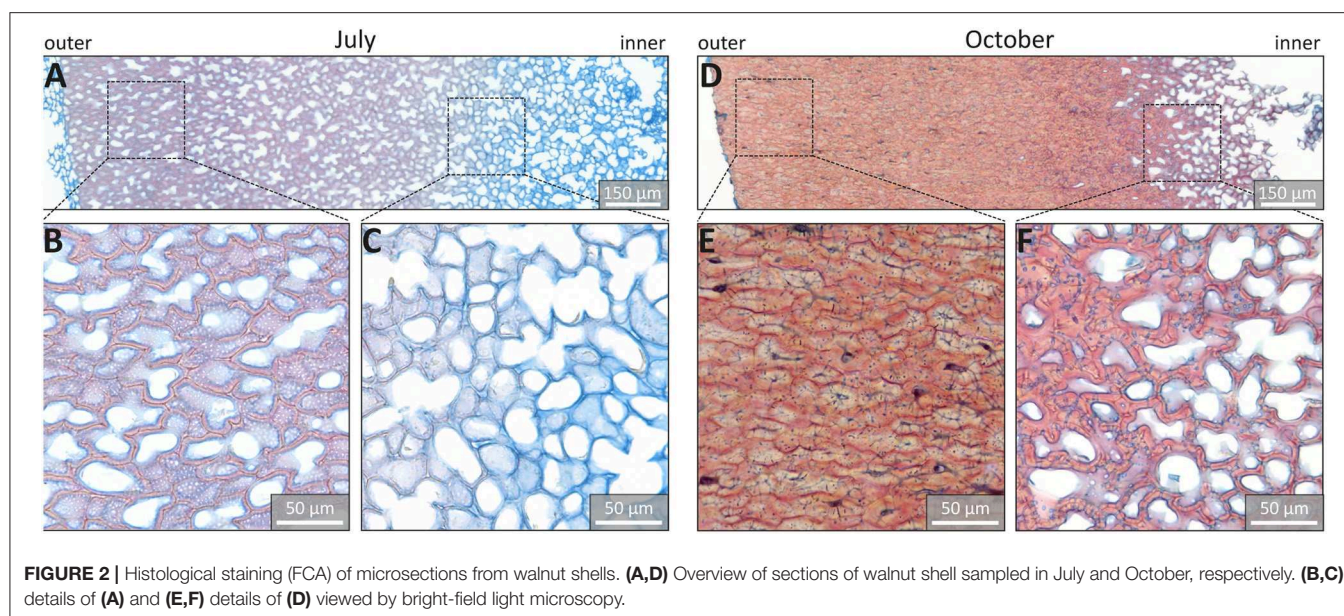
FIGURE 2 | Histological staining (FCA) of microsections from walnut shells. (A,D) Overview of sections of walnut shell sampled in July and October, respectively. (B,C) details of (A) and (E,F) details of (D) viewed by bright-field light microscopy.

maps, intensity distribution was homogenous over a large part of the measured area, with a steep gradient confined to the inner part (**Figure 3B**). The shape of the spectra of the outer and inner regions was similar for the 2 months, but the absolute intensities differed and the lignin marker bands (1,504, 1,592 cm^{-1}) were not detected in the very inner part of the July sample (**Figures 3C,D**). The cellulose intensity in the October sample was roughly twice that of the July sample, whereas hemicelluloses increased by around 1.5 times and the lignin signal was around three times stronger in the October sample compared to the July sample (**Figures 3C,D**).

By normalizing FT-IR spectra on the 1,374 cm^{-1} band, the major intensity differences due to thickening of the cell wall are canceled out and compositional differences can be analyzed using principal component analysis (PCA) (**Figures 3E,F**). Composition along the section in the October sample is less variable than in the July sample as illustrated by the tight grouping of the outer part on the right side along the axis of Principal Component 1 (PC1, red squares). PC1 explains 90% of the spectral variability and spreads the samples from inner to outer part for the October and July samples along the PC1-axis. Spectra from the outer shell (orange round) and middle part (green round) of the July sample are along PC1 in the same region as the middle part of the October sample (**Figure 3E**, orange squares). The PC1-loading reveals bands typical for hemicelluloses (e.g., 1,747 cm^{-1}), lignin (e.g., 1,508 cm^{-1}) as well as cellulose (e.g., 1,167, 1,126, 1,060 cm^{-1}) as observed in the walnut spectra (compare loading 1 **Figure 3F** with spectra of **Figures 3C,D**). However, some bands have proportionally lower intensities (1,591, 1,323, 1,239 cm^{-1}); moreover a strong negative band at 1,006 cm^{-1} (shoulder 983 cm^{-1}) as well as strong medium bands at 1,714, 1,357, 1,307, and 1,203 cm^{-1} are observed (**Figure 3F**, loading 1). PC2 explains another 7% of the variability and separates July samples from October samples and especially the spectra of the inner

parts of the shell (**Figure 3E**, blue markers). The loading 2 has strong contributions at 1,591, 1,502, 1,456, 1,419, 1,323, 1,223, 1,126 cm^{-1}.

Raman Imaging

In the next step, Confocal Raman microscopy was applied to obtain detailed insights into the molecular composition of the nutshell at the microscale. The measurement area covered again the entire shell from the outer to the inner part (**Figures 4A**, **5A** and **Supplementary Figures 1, 2**). To highlight developmental changes, representative areas were selected: one region from the outer part of the shell and a wider region from the inner part, where most of the changes were observed by FT-IR (**Figure 3**). To get an overview of the microstructure, chemical images were generated by integrating over the CH-stretching region (2,745–3,054 cm^{-1}), including signal contributions from all cell wall polymers (Gierlinger et al., 2012) (**Figures 4B**, **5B** and **Supplementary Figures 1A, 2A**). In the July sample, highest CH-stretching intensity (light color) was found in the left image (representative for the outer part), and a substantial decrease in intensity toward the seed—the inner side of the shell (**Figure 4B**, **Supplementary Figure 1A**). In the contrast, in the section of the October sample (**Figure 5B**, **Supplementary Figure 2A**), an overall high intensity was observed in the outer as well as in the inner part.

The integral of the lignin-specific region 1,535–1,704 cm^{-1} (Gierlinger and Schwanninger, 2007) was combined with the integration of the 1,380 cm^{-1} cellulose band, which highlights non-lignified regions (Gierlinger, 2018). In these combination images areas without lignin are clearly visualized in green and at the same time the lignified regions are highlighted in magenta (**Figures 4C**, **5C**). In the July sample, the highest intensity (magenta color) of the lignin band was observed in the outer part, while the cellulose integration (1,380 cm^{-1}) highlighted the thin cell walls of the inner part (**Figure 4C**,
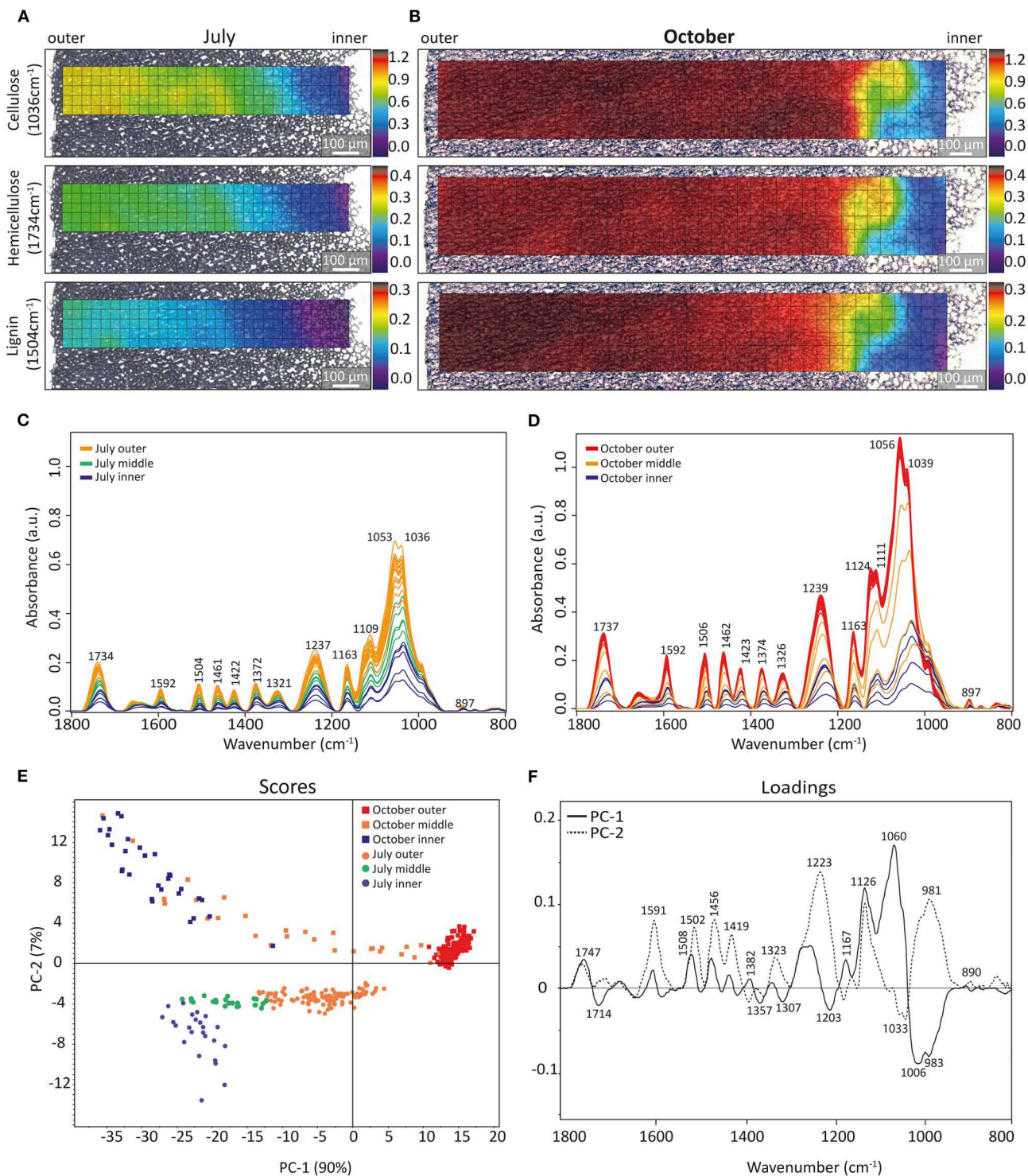
FIGURE 3 | IR image reconstruction of sections of walnut shells from July and October. Chemical mapping overlaid with bright-field images of sections based on the integrated absorbance of cellulose (1,036 cm^{-1}), hemicelluloses (1,734 cm^{-1}), and lignin (1,504 cm^{-1}) of a microsection from July **(A)** and October **(B)**. A rainbow scheme has been used to denote absorbance, with the warmest colors (red) indicating the highest absorbance, and cool colors (blue) representing a low spectral intensity. **(C,D)** Show comparison of baseline corrected spectra of the 1st line from outer to the inner part of sections from July and October, respectively. Score plots **(E)** of principal component analysis (PCA) of infrared spectra obtained from the walnut shell sections of July and October after baseline-correction and normalized at 1,374 cm^{-1} and the corresponding loadings **(F)** of PC1 and PC2.
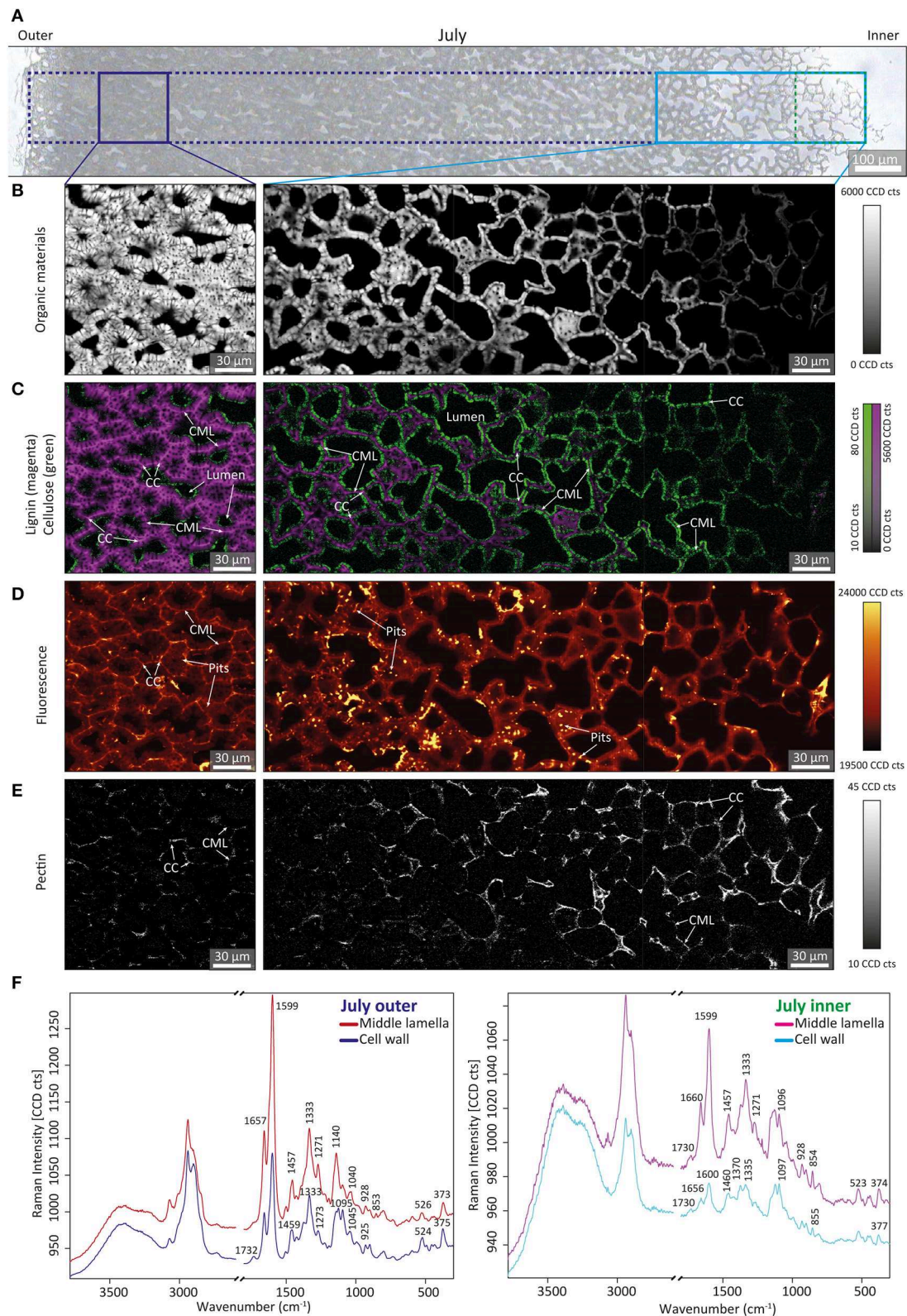
FIGURE 4 | Bright field photomicrograph and representative Raman images of walnut shell section of July. **(A)** Bright field image of a microsection of walnut shell from outer to inner. **(B–E)** Raman images (150 × 150 μm) were calculated by integrating over **(B)** CH -stretching region 2,745–3,054 cm⁻¹. **(C)** Lignin around the region 1,535–1,704 cm⁻¹ (in magenta) overlaid with cellulose bands (1,358–1,401 cm⁻¹, in green). **(D)** Fluorescence region. **(E)** Intensity of pectin (855 cm⁻¹) band. **(F)** Average spectra extracted from the cell wall and compound middle lamella area of the outer (blue) and very inner part (green) of walnut shell section of July, respectively.

FIGURE 5 | Bright field photomicrograph and Raman images of walnut shell section of October. **(A)** Bright field image of a microsection of walnut shell from outer to inner. **(B–E)** Raman images (150 × 150 μm) were calculated by integrating over **(B)** organic materials region 2,745–3,054 cm^{-1}. **(C)** Lignin around the region 1,535–1,704 cm^{-1} (in magenta) overlaid with cellulose bands (1,358–1,401 cm^{-1}, in green). **(D)** Fluorescence region. **(E)** Average spectra extracted from the cell wall and compound middle lamella area of the outer (purple) and very inner part (yellow) of walnut shell section of October, respectively.

Supplementary Figure 1B). At the micro-scale, a higher amount of aromatics was visualized in the cell corners (CC) and compound middle lamella (CML) compared to the cell wall. In the thicker cell walls, an inner, still not yet lignified layer (green) was detected, close to the lumen (**Figure 4C**, left image). In the October samples, aromatic components were visualized throughout the whole shell (in magenta; **Figure 5C**, **Supplementary Figure 2B**). Even in the inner part the thin walls were lignified and high amounts of aromatics were found in the CC and CML (**Figure 5C**, right side).

By plotting the overall intensity change of the background, which stands for all fluorescing components, additional details were visualized (**Figures 4D**, **5D**, **Supplementary Figures 1C, 2C**). Especially in the zone where lignification begins (**Figure 4C**, magenta), highly fluorescing components were found in the pits, the numerous cell to cell connections and partly on the inner surfaces of the cells (**Figure 4D**, middle). In the October samples, the fluorescing layer toward the lumen is even more pronounced in the inner part (**Figure 5D**, right side). In the outer part of the July sample, CML and CC showed higher fluorescence and coincided with the lignin distribution (**Figures 4C,D**, left image). The outer part of the October sample showed the background with the highest fluorescence between the cells and inside pit channels (**Figure 5D**, left image).

Finally, pectin distribution was visualized by integrating over the 857 cm^{-1} band, assigned to pectin (Synytsya et al., 2003). In the July sample, pectin was clearly detected throughout the entire shell (**Figure 4E**). Unlike other cell wall components pectin increased in the inner part (**Figure 4E**, right side). The distribution on the micro-scale showed an accumulation in CC and CML, especially in the unlignified regions (**Figure 4E**, right side). The derived spectra from the cell wall and the middle lamella of the shell of the July sample (**Figure 4F**) clearly confirmed the pectin band at 855 cm^{-1} in the July samples, beside the characteristic bands for cellulose and lignin. The increase of the aromatic stretching vibration at 1,599 cm^{-1} from the inner (right side) to the outer shell part (left side) is clearly seen as well as the higher amount in the middle lamella compared to the cell wall (red vs blue spectra). In the October sample the pectin marker band could neither be observed in the inner part, nor in the outer part as confirmed in the derived average spectra from the cell corners and cell walls (**Figure 5E**). In all spectra lignin bands dominate (1,599, 1,335, 1,140 cm^{-1}) and contrary to what was observed in the shell of the July sample (**Figure 4F**), the fluorescence background in the cell wall is much higher than in the middle lamella (**Figure 5E**). Extraction of the micro-sections (with the aim to remove extractable components) did not result in lower fluorescence background and better spectral quality (**Supplementary Figure 4**).

Zooming into the inner part of the July sample (**Figure 6A**) and applying a multivariate unmixing approach, the onset of lignification was visualized in detail (**Figures 6B–E**). Non negative matrix factorization (NMF) delivers the most pure component (endmember) spectra and their abundance maps (Prats-Mateu et al., 2018). One component was retrieved from CC, CML and pits (**Figure 6B**) and showed clear aromatic bands

at 1,657, 1,598, 1,335, 1,140 cm^{-1} together with the pectin band at 853 cm^{-1} (Synytsya et al., 2003) (**Figure 6D**, magenta). The onset of cell wall thickening goes hand in hand with lignification as proven by the fact that the most pure component from the cell wall (**Figure 6B**, green), clearly included aromatic bands, although less intense (**Figure 6D**, green spectrum). In the innermost thinner walls the aromatic component was restricted to the cell corner and intracellular space (**Figure 6C**, magenta) and an additional band at 649 cm^{-1} was detected (**Figure 6E**, magenta). The endmember spectrum of the thin cell wall in the innermost part (**Figure 6C**) shows a strong band at 856 cm^{-1} [assigned to pectin (Synytsya et al., 2003)] beside cellulose bands [e.g., 1,093, 377 cm^{-1}; (Wiley and Atalla, 1987)], but no aromatic band at 1,600 cm^{-1} (**Figure 6E**, green). In this innermost part of the July nutshell aromatic components are present not in the cell wall, but only in the space between the cells and cell corner (**Figures 6C,E**, magenta).

DISCUSSION

Green Husk: Production and Protection of the Nutshell

X-ray tomography is a non-destructive method gaining popularity for internal quality evaluation in various fields of agriculture and food quality evaluation (Kotwaliwale et al., 2014). In this study, the technique gave a detailed overview of the structure of the whole fruit during development (**Figure 1A** and **Supplementary Videos 1–5**). In the June sample, the shell is not clearly distinguishable from the husk, only the bundles indicate the position. From the July to the September samples, the shell layer is clearly visible and increases in thickness, while around the seed more and more air space is arising (black area). Finally the much denser (brighter on the scans) and dry nutshell protects the seed, which is mostly surrounded by an empty cavity (**Figure 1A**, **Supplementary Movie 5**). Zooming into the nutshell via x-ray tomography showed that the entire shell is composed of polylobate cells, each of which tightly interlocked with on average 14 neighbors (Antreich et al., 2019). While other nutshells, like Macadamia build up their cells with different cell types, i.e., fibers and sclerenchyma cells (Schuler et al., 2014), walnut relies only on the puzzle cell type (Antreich et al., 2019). This explains why no layering within the walnut shell is visible by X-ray-tomography.

The drying process, especially during the last month of development is represented by a strong black contrast and results in a more and more "wrinkled" and dense (white) nutshell (**Figure 1A**). The outer husk, which protects the nut during the shell formation, is gray on the CT scans and is thus a less dense layer. During the first month until July it increases in thickness, but stays constant during shell formation. Also light-microscopic images before and after staining (**Figures 1B,C**) confirm the protecting husk to show major changes from June to July, but staying constant in terms of size and color during shell development (July-September). The color changes in the inner shell from light yellow to brown (**Figure 1B**) point to a change in lignin amount and/or impregnation

FIGURE 6 | Non-negative matrix factorization (NMF) analysis of the inner July shell. **(A)** Bright field image of a microsection of walnut shell of July with the two positions used for analysis marked in the innermost right side. Colored composite images **(B,C)** showing the two representative endmember (EM) abundance maps of cell wall (green) and cell corner (magenta) and the corresponding endmember spectra **(D,E)** after baseline correction.

with other phenolic substances. Staining the nut halves with Wiesner reagent, confirmed the presence of cell wall associated phenolic molecules (phenolic acids and C6-C1 derivatives of hydroxycinnamaldehydes/alcohols) in the nutshell by red coloration from July on (**Figure 1C**). Similar studies on Chinese walnut varieties ("Zanmei" and "Zhenzhuxiang") showed the beginning of coloration and thus lignification on June 16th and

reported the finished shell layer differentiation one month later (July 20th) (Zhao et al., 2016). Although the shell is finally the most stained region, the very first beginning of lignification in the June sample takes place in the very inner part of the diaphragm along the two major bundles supporting the kernel (**Figure 1C**, June). Being the first part to be impregnated with aromatics this walnut septum was recently proven as a rich source

of polyphenols and has been used as a traditional nutraceutical material (Liu et al., 2019). Within the shell the first cells to become sclerified were reported to be those at the micropylar end, proceeding basi-petally with the cells along the suture lines becoming very rapidly sclerified (Pinney and Polito, 1983).

Cross-sections stained with FCA confirmed the start of lignification in July (**Figure 2A**). The outer part (toward the green husk (**Figures 2A,B**) is more reddish than the inner part (**Figure 2C**). This outer part is adjacent to the fleshy outer green husk, which provides nutrition for the seed's growth (Wu et al., 2009). Vascular bundles are found at the border in the surrounding husk and are probably involved in providing and transporting the needed resources for the development of the shell (**Supplementary Figure 3**), although they are not passing through the shell like in Macadamia (Schuler et al., 2014) or coconut (Schmier et al., 2000). The cell walls adjacent to the green husk are the first to be impregnated with aromatic components in the developing nutshell from July on (**Figures 2A**, **3A,C**, **4C,F**). The beginning of lignification in this outer part and the decrease in the amount of aromatics from the green husk toward the inner part points to the role of the green husk in providing components for maturation of the shell.

Nut Shell: Secondary Cell Wall Formation and Lignification

Raman spectroscopy is a very suitable tool to detect the onset of lignification, because of the signal enhancement of conjugated aromatic compounds. Lignin monolignols—the main building blocks of lignin (Vanholme et al., 2010)—are such compounds and give rise to a prominent band at 1,600 cm^{-1}. This band itself is only indicative for an aromatic ring stretch, but its relative intensity to other bands can be interpreted as caused by an aromatic ring participating in a conjugated π-system. With the co-appearance of the bands at 1,657, 1,333, 1,271, and 1,140 cm^{-1} the acquired Raman spectra can with good confidence be assigned to coniferyl alcohol and aldehyde (Bock and Gierlinger, 2019). Based on the signal enhancement of these monolignols the onset of lignification can be precisely monitored by following their distribution during cell wall formation at different developmental stages (**Figures 4–6**). Based on the spectra we can not distinguish if coniferyl alcohol and alcohol are present as monolignols in the developing cell wall or as endgroups in a continuously growing lignin polymer. The fact that neither extraction of young developmental stages nor the mature nut shells changed their aromatic spectral signature, points to detection of endgroups of the lignin polymer.

The inner part of the nutshell sampled in July represents thin primary cell walls, as proven by the derived plant cell wall endmember spectrum (**Figure 6E**, green spectrum) with characteristic marker bands of pectin at 856 cm^{-1} (Synytsya et al., 2003) and cellulose [e.g., 377, 1,093 cm^{-1}, (Wiley and Atalla, 1987)] and the absence of the aromatic 1,600 cm^{-1} vibration. However, coniferyl alcohol and aldehyde bands are detected in the space between two cells, at the cell corner and at the lumen sided surfaces (**Figure 6C**). Two cells toward the green husk the cell thickness almost doubled (**Figure 6B**) and

aromatic components spread clearly in the middle lamella as well as in the pits (**Figure 6B**), which seem to be pathways to transport the monolignols. The numerous pits found in stone cells are reported to maintain symplastic connection between cells by a recess of the cell wall (Reis et al., 1992; Sebaa and Harche, 2014). When the middle lamella is filled up with monolignols, they spread into the secondary cell wall as proven by the aromatic bands in the endmember spectra of the cell wall (**Figure 6D**, green spectrum). How lignin monomers are trafficked from inside the cell to the cell wall is currently debated and possible mechanisms involve transporters, the diffusion of monomers across lipid bilayers and the release of monolignol glucosides stored in vacuoles (Perkins et al., 2019). For the nutshell we could show a clear tracking of monolignols from lumen over pits to first accumulate in the middle lamella and cell corners between the cells, before impregnating the cell wall. In the state of primary cell wall (**Figures 6C,E**) the monolignol dominated endmember spectrum shows an additional band at 648 cm^{-1}, which might be attributed to proteins involved in monolignol synthesis. There is not yet a comprehensive model of the mechanisms of monolignol export and above reported mechanisms are not mutually exclusive, but may predominate in different tissues and in different developmental stages (Perkins et al., 2019).

Another interesting fact is the clear Raman detection of pectin within the primary cell wall on the innermost part of the shell (**Figures 6C,E**, green) and filling together with lignin the space between the cells (CC, CML, and pits in **Figures 6B,D** magenta) (**Figure 4E**). Raman imaging shows a clear decrease in intensity of the pectin band (853 cm^{-1}) toward the green husk in the nutshell sampled in July (**Figure 4E**), while at the same time the aromatic 1600 cm^{-1} stretching band increased (**Figure 4C**). By microautoradiography of polygalcturonan deposition it was shown that pectin formation terminated when secondary cell wall synthesis began (Imai and Terashima, 1992). Using pectin sensitive antibodies the absence of pectin in the secondary cell wall has been verified in pine xylem (Hafren et al., 2000) and stone cells of Norway spruce phloem (Kim and Daniel, 2017). In a very recent study on xylem cell wall formation in pioneer roots and stems of poplar pectins did not co-localize in lignified cell walls, but were found in primary tissues (Marzec-Schmidt et al., 2019). The role of acidic pectin in secondary cell wall formation of the nutshell is confirmed by our intense pectin band at 853–854 cm^{-1} in the July sample (**Figures 4F**, **6D,E**), which corresponds exactly to the same position as Polygalacturonic acid (Synytsya et al., 2003). This band was found to monitor changes in pectin composition, as it decreases with methylation (min. 850 cm^{-1}) and increases with acetylation (max. 862 cm^{-1}) (Synytsya et al., 2003). Although the July sample includes thin unlignified primary cell wall toward the seed (**Figures 6C,E**) and thicker lignified cell wall toward the green husk (**Figure 4D**), the pectin composition based on the Raman band position stays constant while the amount changes (**Figures 4F**, **6D,E**). These results in the nutshell development coincide with the observation that pectin is partly degraded on maturation of secondary cell wall formation in wood (Westermark, 1982; Westermark et al., 1986) and induces lignification (Robertsen, 1986).

Maturation of the Shell by Impregnation and Dehydration

The color changes from July to October in the native (**Figure 1B**) as well as the stained nut halves (**Figure 1C**) indicate an ongoing nut shell maturation in the last three month of walnut development. The thicker cell walls in October are even more reddish, suggesting higher amount of lignin and/or additional impregnation with extractives (**Figures 2B,E,F**).

The score plot of the PCA analysis of the FT-IR spectra (**Figure 3E**) confirmed spectral and thus chemical differences between July and October sample and across the tissue from the inner (toward seed) to the outer part (toward green husk). Along the PC1 axis (explaining 90% of the variability), most of the spectra of the October sample built a cluster on the positive right side (**Figure 3E**, red circles), while the spectra from the outer position of the July sample fall together with the spectra of the middle part of the October sample (**Figure 3E**). Some bands in the PC1 loading (**Figure 3F**) point to a change in aromatic components (e.g., aromatic skeletal vibration 1,508 cm^{-1}, C-H deformation combined with 1,456 cm^{-1}, reported to be common in different lignin samples (Boeriu et al., 2004). Another remarkable band increasing in October outer is the 1,126 cm^{-1} band, reported to be characteristic for S-lignin (Cai et al., 2010), but might together with the 1,060 cm^{-1} band, also come from contributions of tannins (Arshad et al., 1969). PC2 explains another 7% and separates July from October samples and especially the spectra of the inner parts of the shell (blue markers, **Figure 3E**). The loading 2 has strong contributions at 1,591, 1,502, 1,456, 1,419, 1,323, 1,223, and 1,126 cm^{-1}, all typical bands in S-lignin (Faix, 1991) and/or together with 981 cm^{-1} interpreted as tannins (Arshad et al., 1969). A high lignin content around 45% was reported for walnut shell and considered to be SGH type with dominant G-units (86.7%) (Li et al., 2018), while others come up with 30% lignin content with an S/G ratio of 1.6 (Queirós et al., 2019). Extractive content was reported 7–10% (Li et al., 2018; Queirós et al., 2019) with a high amount of phenolics, including flavonoids and tannins (Queirós et al., 2019). Recently hydroxystilbenes, a class of nonflavonoid polyphenolics, have been found to be part of the lignin structure of palm fruit endocarps based on nuclear magnetic resonance spectroscopy (Carlos Del Rio et al., 2017; Rencoret et al., 2018). The incorporation of piceatannol into the lignin polymer was suggested to have a role in seed protection (Rencoret et al., 2018).

Raman microscopy showed the effect of maturation by a tremendously increasing fluorescence background (compare **Figures 4**, **5**). Unfortunately, the higher background of the Raman spectra only provides information on the strongest aromatic stretching vibrations, whereas at some places, especially the pits and the lumen, no bands could be resolved. Raman imaging results suggest a final impregnation of the tissue through the numerous pit channels by additional aromatic components as the fluorescence signal is increasing in the cell wall, but even more in the pits (**Figure 5D** and **Supplementary Figure 2C**). Extraction of the sections did not remove the fluorescence background (**Supplementary Figure 4**), which points to a non-extractable phenolic component or lignin. However, lignin autofluorescence imaging showed the highest intensity in the

cell corner and middle lamella (**Supplementary Figure 5A**) in contrast to the Raman fluorescence lining the cells and the pits (**Supplementary Figure 4**). Together with the SEM images (**Supplementary Figure 5B**), which shows components sticking to the wall, we therefore conclude that aromatic components other than lignin stick to the inner wall of the cells and seal the pits (**Figure 5D**). Although not soluble by the usual extraction procedure additional aromatics are impregnating the walnut tissue in the mature state. Raman imaging of pine wood showed a similar accumulation of aromatic components (pinosylvins) during heartwood formation - with lumen sided surfaces and pits more impregnated than the cell wall itself (Felhofer et al., 2018). This final impregnation step of the cell wall, filling up free spaces, and sealing the pit channels will improve stability and longevity in a similar manner as observed in heartwood of trees.

Although wood and nutshell have different mechanical functions, which are realized by isotropic puzzle cells (Antreich et al., 2019) for high compression forces and anisotropic fiber arrangement for high tensile strength respectively (Gibson, 2012), the underlying secondary cell wall and its maturation show common features.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

NG and NX conceived the experimental design and data analysis. NX and SA performed light microscopy and SEM microscopy. NX and PB conducted the FTIR and Raman spectroscopy. SA, YS, and JS contributed in design of the micro-CT experiment and in the analysis of the data. All authors contributed in writing and reviewing the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00466/full#supplementary-material

Supplementary Video 1 | Micro-CT video of walnut fruits harvested from June.

Supplementary Video 2 | Micro-CT video of walnut fruits harvested from July.

Supplementary Video 3 | Micro-CT video of walnut fruits harvested from August.

Supplementary Video 4 | Micro-CT video of walnut fruits harvested from September.

Supplementary Video 5 | Micro-CT video of walnut fruits harvested from October.

REFERENCES

Agarwal, U. P. (2006). Raman imaging to investigate ultrastructure and composition of plant cell walls: distribution of lignin and cellulose in black spruce wood (*Picea mariana*). *Planta* 224, 1141–1153. doi: 10.1007/s00425-006-0295-z

Antreich, S. J., Xiao, N., Huss, J. C., Horbelt, N., Eder, M., Weinkamer, R., et al. (2019). The puzzle of the walnut shell: a novel cell type with interlocked packing. *Adv. Sci.* 6:1900644. doi: 10.1002/advs.201900644

Arshad, M., Beg, A., and Siddiqui, Z. A. (1969). Infrared spectroscopic investigation of tannins. *Angew. Makromolek. Chem.* 7, 67–78. doi: 10.1002/apmc.1969.050070106

Bernard, A., Barreneche, T., Lheureux, F., and Dirlewanger, E. (2018). Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia L.*) germplasm using SSR markers. *PLos ONE* 13:e0208021. doi: 10.1371/journal.pone.0208021

Bock, P., and Gierlinger, N. (2019). Infrared and Raman spectra of lignin substructures: coniferyl alcohol, abietin, and coniferyl aldehyde. *J. Raman Spectrosc.* 50, 778–792. doi: 10.1002/jrs.5588

Boeriu, C. G., Bravo, D., Gosselink, R. J. A., and Van Dam, J. E. G. (2004). Characterisation of structure-dependent functional properties of lignin with infrared spectroscopy. *Industr. Crops Products* 20, 205–218. doi: 10.1016/j.indcrop.2004.04.022

Cai, Y. P., Li, G. Q., Nie, J. Q., Lin, Y., Nie, F., Zhang, J. Y., et al. (2010). Study of the structure and biosynthetic pathway of lignin in stone cells of pear. *Sci. Horticult.* 125, 374–379. doi: 10.1016/j.scienta.2010.04.029

Carlos Del Rio, J., Rencoret, J., Gutierrez, A., Kim, H., and Ralph, J. (2017). Hydroxystilbenes are monomers in palm fruit endocarp lignins. *Plant Physiol* 174, 2072–2082. doi: 10.1104/pp.17.00362

Chen, F., Chen, J. H., Wang, Z. J., Zhang, J. W., Li, X. J., Lin, M. G., et al. (2019). Genomics: cracking the mysteries of walnuts. *J. Genet.* 98:33. doi: 10.1007/s12041-019-1084-3

Chen, L., Wang, X. H., Yang, H. P., Lu, Q., Li, D., Yang, Q., et al. (2015). Study on pyrolysis behaviors of non-woody lignins with TG-FTIR and Py-GC/MS. *J. Analyt. Appl. Pyrolysis* 113, 499–507. doi: 10.1016/j.jaap.2015.03.018

Chylinska, M., Szymanska-Chargot, M., Derylo, K., Tchorzewska, D., and Zdunek, A. (2017). Changing of biochemical parameters and cell wall polysaccharides distribution during physiological development of tomato fruit. *Plant Physiol. Biochem.* 119, 328–337. doi: 10.1016/j.plaphy.2017.09.010

Dardick, C., and Callahan, A. M. (2014). Evolution of the fruit endocarp: molecular mechanisms underlying adaptations in seed protection and dispersal strategies. *Front. Plant Sci.* 5:284. doi: 10.3389/fpls.2014.00284

De Rigo, D., Enescu, C. M., Houston Durrant, T., Tinner, W., and Caudullo, G. (2016). *Juglans regia* in Europe: distribution, habitat, usage and threats. *Eur. Atlas Forest Tree Species*, eds J. San-Miguel-Ayanz, D. de Rigo, G. Caudullo, T. Houston Durrant, and A. Mauri (Luxembourg, EU), e01977c+. doi: 10.7892/boris.80796

Demirbas, A. (2005). Estimating of structural composition of wood and non-wood biomass samples. *Energy Sources* 27, 761–767. doi: 10.1080/00908310490450971

Faix, O. (1991). Classification of lignins from different botanical origins by FT-IR Spectroscopy. *Holzforschung* 45, 21–27. doi: 10.1515/hfsg.1991.45.s1.21

Felhofer, M., Prats-Mateu, B., Bock, P., and Gierlinger, N. (2018). Antifungal stilbene impregnation: transport and distribution on the micron-level. *Tree Physiol.* 38, 1526–1537. doi: 10.1093/treephys/tpy073

Felten, J., Hall, H., Jaumot, J., Tauler, R., De Juan, A., and Gorzsas, A. (2015). Vibrational spectroscopic image analysis of biological material using multivariate curve resolution-alternating least squares (MCR-ALS). *Nature Protocols* 10, 217–240. doi: 10.1038/nprot.2015.008

Fengel, D., and Ludwig, M. (1991). Möglichkeiten und Grenzen der FTIR-Spektroskopie bei der Charakterisierung von Cellulose. *Das Papier* 45, 45–51.

Fukuda, T., Ito, H., and Yoshida, T. (2003). Antioxidative polyphenols from walnuts (*Juglans regia L.*). *Phytochemistry* 63, 795–801. doi: 10.1016/S0031-9422(03)00333-9

Gibson, L. J. (2012). The hierarchical structure and mechanics of plant materials. *J. R Soc. Interface* 9, 2749–2766. doi: 10.1098/rsif.2012.0341

Gierlinger, N. (2018). New insights into plant cell walls by vibrational microspectroscopy. *Appl. Spectroscopy Rev.* 53, 517–551. doi: 10.1080/05704928.2017.1363052

Gierlinger, N., Keplinger, T., and Harrington, M. (2012). Imaging of plant cell walls by confocal Raman microscopy. *Nat. Protocols* 7, 1694–1708. doi: 10.1038/nprot.2012.092

Gierlinger, N., and Schwanninger, M. (2006). Chemical imaging of poplar wood cell walls by confocal Raman microscopy. *Plant Physiol.* 140, 1246–1254. doi: 10.1104/pp.105.066993

Gierlinger, N., and Schwanninger, M. (2007). The potential of Raman microscopy and Raman imaging in plant research. *Spectrosc. Int. J.* 21, 69–89. doi: 10.1155/2007/498206

Hafren, J., Daniel, G., and Westermark, U. (2000). The distribution of acidic and esterified pectin in cambium, developing xylem and mature xylem of Pinus sylvestris. *Iawa J.* 21, 157–168. doi: 10.1163/22941932-90000242

Imai, T., and Terashima, N. (1992). Determination of the distribution and reaction of polysaccharides in wood cell-walls by the isotope tracer technique.3. Visualization of the deposition and distribution of galacturonan in the cell-walls of magnolia (Magnolia-Kobus Dc) xylem by microautoradiography. *Mokuzai Gakkaishi* 38, 475–481.

Jahanban-Esfahlan, A., and Amarowicz, R. (2018). Walnut (*Juglans regia L.*) shell pyroligneous acid: chemical constituents and functional applications. *Rsc Adv.* 8, 22376–22391. doi: 10.1039/C8RA03684E

Jahanbani, R., Ghaffari, S. M., Salami, M., Vahdati, K., Sepehri, H., Sarvestani, N. N., et al. (2016). Antioxidant and anticancer activities of walnut (*Juglans regia L.*) protein hydrolysates using different proteases. *Plant Foods Human Nutrition* 71, 402–409. doi: 10.1007/s11130-016-0576-z

Kim, J. S., and Daniel, G. (2017). Immunolocalization of pectin and hemicellulose epitopes in the phloem of Norway spruce and Scots pine. *Trees-Struct. Function* 31, 1335–1353. doi: 10.1007/s00468-017-1552-4

Kornsteiner, M., Wagner, K. H., and Elmadfa, I. (2006). Tocopherols and total phenolics in 10 different nut types. *Food Chem.* 98, 381–387. doi: 10.1016/j.foodchem.2005.07.033

Kotwaliwale, N., Singh, K., Kalne, A., Jha, S. N., Seth, N., and Kar, A. (2014). X-ray imaging methods for internal quality evaluation of agricultural produce. *J. Food Sci. Technol. Mysore* 51, 1–15. doi: 10.1007/s13197-011-0485-y

Lancefield, C. S., Panovic, I., Deuss, P. J., Barta, K., and Westwood, N. J. (2017). Pre-treatment of lignocellulosic feedstocks using biorenewable alcohols: towards complete biomass valorisation. *Green Chem.* 19, 202–214. doi: 10.1039/C6GC02739C

Li, W. Q., Amos, K., Li, M., Pu, Y. Q., Debolt, S., Ragauskas, A. J., et al. (2018). Fractionation and characterization of lignin streams from unique high-lignin content endocarp feedstocks. *Biotechnol. Biofuels* 11:304. doi: 10.1186/s13068-018-1305-7

Liu, P. Z., Li, L. L., Song, L. J., Sun, X. T., Yan, S. J., and Huang, W. J. (2019). Characterisation of phenolics in fruit septum of *Juglans regia* Linn. by ultra performance liquid chromatography coupled with Orbitrap mass spectrometer. *Food Chem.* 286, 669–677. doi: 10.1016/j.foodchem.2019.02.054

Martinez, M. L., Labuckas, D. O., Lamarque, A. L., and Maestri, D. M. (2010). Walnut (*Juglans regia L.*): genetic resources, chemistry, by-products. *J. Sci. Food Agriculture* 90, 1959–1967. doi: 10.1002/jsfa.4059

Marzec-Schmidt, K., Ludwikow, A., Wojciechowska, N., Kasprowicz-Maluski, A., Mucha, J., and Bagniewska-Zadworna, A. (2019). Xylem cell wall formation in

pioneer roots and stems of *Populus trichocarpa* (Torr. and Gray). *Front. Plant Sci.* 10:1419. doi: 10.3389/fpls.2019.01419

Mazurek, S., Mucciolo, A., Humbel, B. M., and Nawrath, C. (2013). Transmission Fourier transform infrared microspectroscopy allows simultaneous assessment of cutin and cell-wall polysaccharides of *Arabidopsis* petals. *Plant J.* 74, 880–891. doi: 10.1111/tpj.12164

Mccann, M. C., and Carpita, N. C. (2008). Designing the deconstruction of plant cell walls. *Curr. Opin. Plant Biol.* 11, 314–320. doi: 10.1016/j.pbi.2008.04.001

Panth, N., Paudel, K. R., and Karki, R. (2016). Phytochemical profile and biological activity of *Juglans regia*. *J. Integr. Med.* 14, 359–373. doi: 10.1016/S2095-4964(16)60274-1

Perkins, M., Smith, R. A., and Samuels, L. (2019). The transport of monomers during lignification in plants: anything goes but how? *Curr. Opin. Biotechnol.* 56, 69–74. doi: 10.1016/j.copbio.2018.09.011

Pinney, K., and Polito, V. S. (1983). English walnut fruit-growth and development. *Sci. Horticult.* 21, 19–28. doi: 10.1016/0304-4238(83)90182-6

Pomar, F., Merino, F., and Barcelo, A. R. (2002). O-4-linked coniferyl and sinapyl aldehydes in lignifying cell walls are the main targets of the Wiesner (phloroglucinol-HCl) reaction. *Protoplasma* 220, 17–28. doi: 10.1007/s00709-002-0030-y

Prats-Mateu, B., Felhofer, M., De Juan, A., and Gierlinger, N. (2018). Multivariate unmixing approaches on Raman images of plant cell walls: new insights or overinterpretation of results? *Plant Methods* 14:52. doi: 10.1186/s13007-018-0320-9

Queirós, S. G. P. C., Cardoso, S., Lourenço, A., Ferreira, J., Miranda, I., Lourenço, M. J., et al. (2019). Characterization of walnut, almond, and pine nut shells regarding chemical composition and extract composition. *Biomass Conversion Biorefinery* 1–14. doi: 10.1007/s13399-019-00424-2

Reis, D., Roland, J. C., Mosiniak, M., Darzens, D., and Vian, B. (1992). The sustained and warped helicoidal pattern of a xylan-cellulose composite - the stony endocarp model. *Protoplasma* 166, 21–34. doi: 10.1007/BF01320139

Rencoret, J., Kim, H., Evaristo, A. B., Gutierrez, A., Ralph, J., and Del Rio, J. C. (2018). Variability in lignin composition and structure in cell walls of different parts of Macauba (*Acrocomia aculeata*) Palm fruit. *J. Agric. Food Chem.* 66, 138–153. doi: 10.1021/acs.jafc.7b04638

Robertsen, B. (1986). Elicitors of the production of lignin-like compounds in cucumber hypocotyls. *Physiol. Mol. Plant Pathol.* 28, 137–148. doi: 10.1016/S0048-4059(86)80015-7

Sallon, S., Solowey, E., Cohen, Y., Korchinsky, R., Egli, M., Woodhatch, I., et al. (2008). Germination, genetics, and growth of an ancient date seed. *Science* 320, 1464–1464. doi: 10.1126/science.1153600

Schmier, S., Hosoda, N., and Speck, T. (2000). Hierarchical structure of the *Cocos nucifera* (Coconut) endocarp: functional morphology and its influence on fracture toughness. *Molecules* 25:223. doi: 10.3390/molecules25010223

Schuler, P., Speck, T., Buhrig-Polaczek, A., and Fleck, C. (2014). Structure-function relationships in *Macadamia integrifolia* seed coats - fundamentals of the hierarchical microstructure. *PLoS ONE* 9:2913. doi: 10.1371/journal.pone.0102913

Sebaa, H. S., and Harche, M. K. (2014). Anatomical structure and ultrastructure of the endocarp cell walls of *Argania spinosa* (L.) Skeels (Sapotaceae). *Micron* 67, 100–106. doi: 10.1016/j.micron.2014.07.001

Shah, M. A., Khan, M. N. S., and Kumar, V. (2018). Biomass residue characterization for their potential application as biofuels. *J. Thermal Anal. Calorimetry* 134, 2137–2145. doi: 10.1007/s10973-018-7560-9

Simon, C., Lion, C., Biot, C., Gierlinger, N., and Hawkins, S. (2018). Lignification and advances in lignin imaging in plant cell walls. *Annual Plant Rev. Online* 1, 1–32. doi: 10.1002/9781119312994.apr0656

Smith, E., and Dent, G. (2005). *Modern Raman Spectroscopy—A Practical Approach*. Chichester: John Wiley & Sons, Ltd. doi: 10.1002/0470011831

Synytsya, A., Copikova, J., Matejka, P., and Machovic, V. (2003). Fourier transform Raman and infrared spectroscopy of pectins. *Carbohydrate Polym.* 54, 97–106. doi: 10.1016/S0144-8617(03)00158-9

Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. (2010). Lignin biosynthesis and structure. *Plant Physiol.* 153, 895–905. doi: 10.1104/pp.110.155119

Verboven, P., Pedersen, O., Herremans, E., Ho, Q. T., Nicolai, B. M., Colmer, T. D., et al. (2012). Root aeration via aerenchymatous phellem: three-dimensional micro-imaging and radial O_2 profiles in *Melilotus siculus*. *N. Phytol.* 193, 420–431. doi: 10.1111/j.1469-8137.2011.03934.x

Wahid, M., Gawli, Y., Puthusseri, D., Kumar, A., Shelke, M. V., and Ogale, S. (2017). Nutty carbon: morphology replicating hard carbon from walnut shell for Na Ion battery anode. *Acs Omega* 2, 3601–3609. doi: 10.1021/acsomega.7b00633

Westermark, U. (1982). Calcium promoted phenolic coupling by superoxide radical - a possible lignification reaction in wood. *Wood Sci. Technol.* 16, 71–78. doi: 10.1007/BF00351376

Westermark, U., Hardell, H. L., and Iversen, T. (1986). The content of protein and pectin in the lignified middle lamella primary wall from spruce fibers. *Holzforschung* 40, 65–68. doi: 10.1515/hfsg.1986.40.2.65

Wiley, J. H., and Atalla, R. H. (1987). Band assignments in the Raman-spectra of celluloses. *Carbohydrate Res.* 160, 113–129. doi: 10.1016/0008-6215(87)80306-3

Wu, G. L., Liu, Q. L., and Da Silva, J. (2009). Ultrastructure of pericarp and seed capsule cells in the developing walnut (*Juglans regia* L.) fruit. *South African J. Botany* 75, 128–136. doi: 10.1016/j.sajb.2008.09.001

Xie, R. Z., Wang, H., Chen, Y., and Jiang, W. J. (2013). Walnut shell-based activated carbon with excellent copper (II) adsorption and lower chromium (VI) removal prepared by acid-base modification. *Environ. Prog. Sustain. Energy* 32, 688–696. doi: 10.1002/ep.11686

Yang, W., Shimizu, I., Ono, T., and Kimura, Y. (2015). Preparation of biodegradable foam from walnut shells treated by subcritical water. *J. Chem. Technol. Biotechnol.* 90, 44–49. doi: 10.1002/jctb.4451

Yang, Z. L., Mei, J. Q., Liu, Z. Q., Huang, G. Q., Huang, G., and Han, L. J. (2018). Visualization and semiquantitative study of the distribution of major components in wheat straw in mesoscopic scale using Fourier transform infrared microspectroscopic imaging. *Anal. Chem.* 90, 7332–7340. doi: 10.1021/acs.analchem.8b00614

Yeung, E. C. (1998). "A beginner's guide to the study of plant structure," in *Tested Studies for Laboratory Teaching, Volume 19.*, eds S. J. Karcher, Editor. *Proceedings of the 19th Workshop/Conference of the Association for Biology Laboratory Education* (ABLE), 125–142.

Zhang, Z. J., Liao, L. P., Moore, J., Wu, T., and Wang, Z. T. (2009). Antioxidant phenolic compounds from walnut kernels (*Juglans regia* L.). *Food Chem.* 113, 160–165. doi: 10.1016/j.foodchem.2008.07.061

Zhao, S. G., Wen, J., Wang, H. X., Zhang, Z. H., and Li, C. B. (2016). Changes in lignin content and activity of related enzymes in the endocarp during the walnut shell development period. *Horticultural Plant J.* 2, 141–146. doi: 10.1016/j.hpj.2016.08.003

Check for
updates

Flavor-Related Quality Attributes of Ripe Tomatoes Are Not Significantly Affected Under Two Common Household Conditions

Larissa Kanski*, Marcel Naumann and Elke Pawelzik

Division of Quality of Plant Products, Department of Crop Sciences, Faculty of Agriculture, University of Göttingen, Göttingen, Germany

Consumer complaints about the flavor of fresh tomato fruits (*Solanum lycopersicum* L.) have increased in the past few decades, and numerous studies have been done on the flavor of tomatoes and how it is influenced. However, it has not yet been taken into account how consumer handling affects the flavor when considering the complete post-harvest chain—from retailer (distributor) to retail to consumer. In this study, the impact of two household storage regimes on the volatile profile and important flavor-related compounds were examined, considering the entire post-harvest handling. New breeding lines ($n = 2$) and their parental cultivars ($n = 3$) were evaluated. Fruits were harvested ripe and stored at 12.5°C for 1 day, at 20°C for 2 days, and afterward at either 20 or 7°C for another 4 days. The aroma volatile profile was measured using GC-MS and GC-FID. A trained panel was used to characterize the sensory attributes of the fruits. In both storage regimes, the relative amount of hexanal increased during the storage period in three of the five cultivars/breeding lines while benzaldehyde was the only volatile compound that decreased significantly in four cultivars/breeding lines. The relative concentration of the precursors of lipid-derived volatiles—linoleic (C18:2) and linolenic (C18:3) acid—did not change in both storage regimes. The lycopene and β-carotene contents increased slightly during storage (20 and 7°C), as the carotenoid-derived volatile 6-methyl-5-hepten-2-one did. The fructose and glucose concentrations did not vary significantly, while the content of total soluble solids increased during both storage regimes. No significant difference could be found between the fruits stored at 20 or 7°C for 4 days after the post-harvest handling in all the parameters analyzed, including the sensory analysis, considering all cultivars/breeding lines. A storage temperature of 7°C is not detrimental for the flavor of ripe fruits under the experimental conditions used. The genetic background of the studied cultivars/breeding lines have a higher impact on the flavor variation than the two common household storage conditions when storing ripe fruits and taking the entire post-harvest handling into account.

Keywords: *Solanum lycopersicum* L., aroma volatiles, flavor, post-harvest chain, household storage, quality, sensory analysis

INTRODUCTION

Tomato (*Solanum lycopersicum* L.) is one of the most widely consumed vegetables in the world (Díaz de León-Sánchez et al., 2009), with 182 million tons produced worldwide in 2017 (FAOSTAT, 2019). However, consumers have been increasingly complaining about the flavor of the fresh fruits (Tieman et al., 2012; Klee and Tieman, 2013; Zhang et al., 2016). Reduced consumer acceptance of the fruits is caused by breeding programs that in particular focus on extended shelf-life, size, and yield in combination with inappropriate post-harvest handling conditions (Tieman et al., 2017). Nevertheless, tomato flavor is acquiring more and more importance for the consumers and, consequently, for the producers as well (Piombino et al., 2013). In this context, taste and flavor are widely used terms and very often used as synonyms. However, taste refers to the gustatory receptors on the tongue (sweet, sour, salty, bitter, and umami) while flavor is the result of taste, retronasal olfaction (perception of aroma volatiles via the mouth), and trigeminal inputs, and can be further enhanced by orthonasal olfaction ("sniffing") (Spence, 2013, 2015). Tomato flavor is a complex interaction of aroma volatiles and taste, which is influenced also by visual and textural signals in the brain (Baldwin et al., 2008; Vogel et al., 2010). So far, over 400 different volatile organic compounds have been found in tomatoes, though only a small number contribute to the characteristic tomato flavor of the fruits (Baldwin et al., 2000; Krumbein et al., 2004). These volatiles are derived from several precursors, including fatty acids, amino acids, and carotenoids (Zanor et al., 2009; Vogel et al., 2010). Vogel et al. (2010) revealed that the reduction of apocarotenoids in fruits leads to reduced flavor acceptability and negatively correlates with sweetness perception. Other important contributors to flavor perception in tomatoes are sugars (mainly fructose and glucose) and acids (mainly citric and malic acid) (Ruiz et al., 2005; Tieman et al., 2012). However, the genetic background plays an important role in the chemical and sensory composition of the fruit (Tieman et al., 2012; Baldwin et al., 2015). The exact pathways of all aroma volatiles have not yet been identified and further studies are necessary to gain more insight in the complex interaction of aroma and the pathway of the volatiles and its precursors. The huge influence of the cultivar, affecting flavor perception was revealed by Tieman et al. (2012) as well. They observed volatile content variations with up to 3,000-fold differences across 152 studied heirloom varieties. Another study of Tieman et al. (2017) identified genetic loci and chemical compounds associated with consumer liking. They determined various flavor-associated chemical compounds of different heirloom varieties and compared them with modern cultivars. The results demonstrated that modern cultivars are not well liked, even when growing under commercial conditions and harvested fully ripe. Consequently, breeding new cultivars with enhanced flavor perception could be an approach toward flavor improvement in general, in addition to appropriate post-harvest handling (Kader, 2008). Many studies have dealt with tomato flavor (e.g., Selli et al., 2014; Klee and Tieman, 2018) and it changes during storage (e.g., Krumbein et al., 2004), especially the reduction in flavor due to the exposure of fruits to low temperatures (e.g., Maul et al., 2000; Renard et al., 2013; Ponce-Valadez et al., 2016). Household storage conditions can also affect tomato fruit quality (Renard et al., 2013). There are two commonly used handling conditions for storing tomato fruits in households, either in the refrigerator (4–8°C) or at room temperature (about 20°C). Renard et al. (2013) compared two household storage regimes that included room temperature (20°C) and refrigerated storage (4°C) for different durations. The authors reported a strong negative effect on the volatile profile of red-ripe fruits stored at 4°C, but up to 1 week (6 days) the aroma could be restored by reconditioning the tomatoes at room temperature for 24 h. In the study of Díaz de León-Sánchez et al. (2009), sensory analysis, volatiles, and alcohol dehydrogenase activity in light red tomato fruits stored at either 10 or 20°C were compared and the main changes in the aroma profile were detected after 6 days of storage, regardless of the temperature. In addition to volatile compounds, non-volatile compounds are investigated in various studies, as they are main contributors to flavor (e.g., total soluble solids, titratable acidity, carotenoids, fatty acids, and firmness) (Díaz de León-Sánchez et al., 2009; Renard et al., 2013; Selli et al., 2014). Some studies examined the postharvest storage durations of up to 3 weeks (Slimestad and Verheul, 2005; Verheul et al., 2015; Dew et al., 2016). Nevertheless, these studies only focused on parts of the postharvest process and did not consider the whole transportation route from harvest via retailing to the consumer. The demand for sustainable post-harvest handling, on the other hand, exists on both the producer and consumer sides to maintain fruit quality (Kader, 2008). New approaches have also been developed measuring taste (electronic tongue) and aroma (electronic nose) of food and beverages for objective high-throughput profiling and can be complementary to existing methods (Beullens et al., 2008). However, further research is still necessary. The objectives of this study were, therefore: (i) to compare freshly harvested fruits with fruits after short-term storage in two different household storage regimes in terms of important flavor-related quality attributes (e.g., total soluble solids, titratable acids, fructose and glucose, citric acid, volatiles, carotenoids, and fatty acids) and sensory perception, taking into account the entire transportation route, (ii) to compare two new breeding lines with their parental cultivars, focusing on the quality attributes, and (iii) to investigate the suitability of the e-tongue and determine whether the results correspond to the sensory attributes of the sensory panel evaluation.

MATERIALS AND METHODS

Plant Material and Growth Conditions

The study was carried out at the experimental research station of the University of Göttingen, Germany (51°30′17.6″N, 9°55′16.2″E). Indeterminate tomato plants were grown under organic low-input conditions in the field under a shelter during summer 2018. Low-input conditions are characterized by a moderate irrigation regime and no, or only low, fertilization (European Council, 2008). Approximately 250 L/m^2 water were irrigated throughout the complete growing season. An organic

NK-fertilizer was applied once (1% concentration, Aminofert® Vinasse, Beckmann & Brehm GmbH, Beckeln, Germany). Two F_4-breeding lines (Black Cherry × Primabella and Black Cherry × Roterno F1) as well as their parental cultivars were used in this study. Black Cherry (Culinaris, Saatgut für Lebensmittel, Göttingen, Germany) and Primabella (Gärtnerei LohmannsHof, Germany) are cocktail tomatoes, while Roterno F1 (Rijk Zwaan, Netherlands) is a salad cultivar. Seeds were germinated in Bio-Traysubstrat (Klasmann-Deilmann GmbH, Geeste, Germany) under greenhouse conditions (22°C day, 18°C night, 16 h/8 h) in April 2018 and were potted 17 days after seeding in Bio-Kräutersubstrat (Klasmann-Deilmann GmbH, Geeste, Germany). After potting, plants grew at 20°C day and 15°C night (16 h/8 h) conditions in a greenhouse before planting in the field. The experimental design was completely randomized with 4 biological replications per cultivar/breeding line and 12 plants per replica, resulting in 240 plants in total. We minimized border effects by planting three plants at either end of each row and one extra row at each side of the experiment. All plants were cultivated with a distance of 0.5 m within the rows and 1 m distance between the rows.

Harvest and Postharvest Treatments and Processing of the Samples Prior to Analyses

For each biological replication, ripe fruits were harvested based on the color measurements at August 20th (Black Cherry, Primabella, Black Cherry × Primabella) and 22nd (Roterno F1, Black Cherry × Roterno F1) and the samples were divided for the different analyses containing the fruit quality analyses, the chemical analyses, and the sensory evaluation. The fruits were first stored for 1 day at 12.5°C (80% humidity). They were subsequently stored for 2 days at 20°C (55% humidity) and finally separated and either stored at room temperature (20°C, 55% humidity) or at refrigerator temperature (7°C, 85% humidity) for 4 days. All postharvest treatments were conducted in darkness and analyses were performed at the fruits directly after harvest as well as at the stored fruits (7 and 20°C) (**Figure 1**). 3 to 12 fruits per biological replication were used for each of the analyses described in the following chapters. For the color, texture, and aroma analyses as well as for panel sensory, fresh fruits were used. For the measurement of total soluble solids, dry matter, and titratable acidity, fruits were sliced, frozen and stored at −20°C until analysis. Soluble sugars, acids, minerals, and fatty acids contents were measured on freeze-dried material (freeze-dryer, EPSILON 2-40, Christ, Osterode am Harz, Germany), which was ground afterward with a ball mill. For the carotenoid and e-tongue analyses, fresh material were frozen in liquid nitrogen and subsequently stored at −20°C until analysis.

Fruit Quality Parameters and Chemical Analyses

Color Measurement and Texture Analysis

The color of the fruits was measured in accordance with the CIEL*a*b* system [Commission Internationale de l'Éclairage (CIE), L = lightness, *a*-value = green to red, *b*-value = blue to yellow] using a Minolta Chroma meter CR-400 (Konica Minolta, Inc., Marunouchi, Japan). The hue angle [H°] was calculated according to Pék et al. (2010). Two opposite equatorial sites on each fruit were measured, with each plot containing 12 fruits. The firmness of the whole fruit was measured in Newton [N] with a texture analyzer on the equatorial site of 8 fruits (5 mm staple micro cylinder, speed: 6 mm/s, distance: 6 mm, TA.XT.plus, Stable Micro Systems Ltd, TA.XT.plus, Godalming, United Kingdom).

Total Soluble Solids (TSS), Dry Matter (DM), and Titratable Acidity (TA)

Frozen samples were thawed and homogenized with a hand blender for 30 s (MQ 5000 Soup, Braun, Kronberg/Taunus, Germany). 10 g of the homogenized sample were weighed in a petri dish, dried in an oven at 105°C for 1 day, and the dry matter was calculated afterward. The remaining sample was filled in 50 ml centrifuge tubes and centrifuged for 20 min at 4,696 *g* at 20°C (Centrifuge 5804 R, Eppendorf, Hamburg, Germany). A few drops of the supernatant were placed on a refractometer (handheld refractometer, A. Krüss Optronic GmbH, Hamburg, Germany) to determine TSS in °Brix. For the evaluation of TA 20 ml of deionized water and 3 ml of the supernatant were pipetted in a glass beaker and the solution was titrated with 0.1 N NaOH solution to pH 8.1 (pH-titrator Titroline 96, SCHOTT AG, Mainz, Germany). The following formula was used to calculate the titratable acidity (TA) as the percentage of citric acid, the main acid in tomato fruits:

$$TA = \frac{usage\ ml\ 0.1\ mol/l\ NaOH\ \times\ 0.1\ mol/l\ \times\ milliequivalent\ of\ citric\ acid\ (0.064\ mVal)\ \times\ 100}{ml\ sample\ (3\ ml)}$$
$$(\%)$$

$$(1)$$

Soluble Sugar, Acid, and Mineral Concentrations

300 mg of freeze-dried material were weighed in 15 ml centrifuge tubes and 8 ml of deionized water were added and shaken horizontally on a shaker for 1 h at room temperature. Afterward, 0.5 ml Carrez I [3.6 g $K_4Fe(CN)_6$ in 100 ml deionized water] and 0.5 ml Carrez II (7.2 g $H_{14}O_{11}SZn$ in 100 ml deionized water) solutions were added, then mixed and centrifuged for 20 min at 4,696 *g* (Heraeus Megafuge 16R Centrifuge, Thermo Fisher Scientific, Waltham, United States). The supernatant was transferred in a 25 ml flask. The pellet was dissolved again in deionized water (8 ml), vortexed, and centrifuged; the supernatants were combined. In total, the procedure was performed three times and the combined supernatants were filled up to 25 ml with deionized water. The samples were filtered with filter paper (Type 615 1/4, Macherey-Nagel, Düren, Germany) in screw cap bottles and stored at −20°C until measurement. Prior to measurement, the samples were thawed, and 1 ml of the solution was filtered through a 0.45 μm PTFE syringe filter (VWR International, Radnor, PA, United States) in 1.5 ml vials. The extracts were quantified by using HPLC (Jasco, Pfungstadt, Germany) for fructose and glucose with following settings: injection volume = 20 μl; eluent = 80% acetonitrile and 20% water; flow rate = 1 ml/min; column = LiChrospher 100 NH2 (5 μm); column temperature = 22°C; refractive index detector.

FIGURE 1 | The post-harvest handling, simulating the post-harvest chain of the fruits from the harvest to the retailer (distributor), to the retail to the consumer.
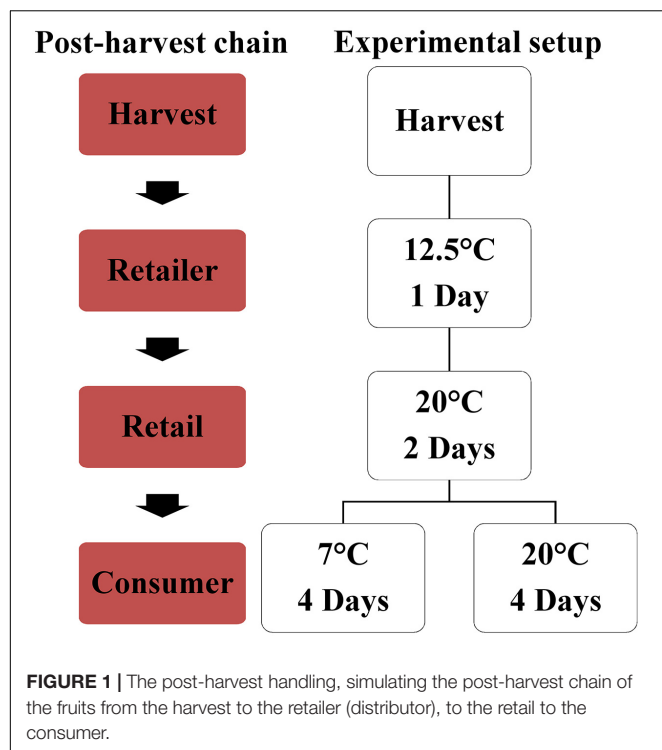
The settings for citric and malic acid quantification was as followed: injection volume = 10 μl; eluent = KH_2PO_4 (0.025 M; pH = 2.5); flow rate = 0.2 ml/min; column = ReproSil-XR 120 C8 5 μm; UV-detector = 200 nm. For the mineral concentrations, 100 mg of the freeze-dried sample were used to analyze and evaluate the minerals, as described previously in Koch et al. (2019).

E-Tongue

Fruits were thawed and blended with a hand blender for 30 s, filled in 50 ml centrifuge tubes and centrifuged at 4,696 *g* for 30 min at 20°C (Heraeus Megafuge 16R Centrifuge, Thermo Fisher Scientific, Waltham, United States). Afterward, the supernatant was filtered through a 615 1/4 filter (Macherey-Nagel, Düren, Germany) and stored at −80°C until analyses. A total quantity of 13.33 ml of the sample was pipetted in a glass beaker from the e-tongue and 66.67 ml deionized water were added. Subsequently, the samples were analyzed with the ASTREE Electronic Tongue (Alpha M.O.S., Toulouse, France) containing the 7-sensor array #6 (Ref. 803-0175; AHS, PKS, CTS, NMS, CPS, ANS, SCS). It is composed of a 16-position autosampler and a reference electrode. The inquisition time was 120 s and the cleaning time for the sensors was conducted after each measurement in deionized water for 10 s.

Analyses of Volatiles and Important Precursors
Aroma Compounds

The determination of volatiles was based on the method of Olbricht et al. (2007) with some modifications. Tomatoes were washed with deionized water and cut into quarters. 50 g were weighed in 1 L beakers, 89.5 ml of a 3.18 M NaCl-solution were added, and the mixture was homogenized for 30 s with a hand blender. The mixture was filled in 50 ml centrifuge tubes and centrifuged at 4°C for 30 min at 1,690 *g* (Heraeus Megafuge 16R Centrifuge, Thermo Fisher Scientific, Waltham, MA, United States). The supernatant was transferred through a sieve and subsequently 8 ml were pipetted in a 20 ml glass vial already filled with 4 g of NaCl for saturation. 16 μl internal standard (0.16 μM 1-octanol dissolved in ethanol) was added. The sample was sealed with a magnetic crimp cap, vortexed for 10 s, and stored at −20°C until analysis. The volatiles were extracted by headspace solid-phase-micro-extraction (SPME) with a 100-μm polydimethylsiloxane (PDMS) fiber (PAL System, CTC Analytics, Zwingen, Switzerland). The incubation time before sampling was 15 min at 35°C, with an agitator speed of 500 rpm. The sample extraction time was 30 min at 35°C and in the same shaking mode. Thermal desorption in the injector was performed for 1 min at 250°C (splitless mode), followed by 9 min in split mode (split ratio 1:10). The analysis was conducted with a GC-2010 Plus (Shimadzu Deutschland GmbH, Duisburg, Germany) equipped with a flame ionization detector (FID). The FID temperature was set to 250°C. Helium was used as carrier gas with a column flow rate of 1.24 ml/min. The temperature was set to 35°C (hold for 5 min) and went up to 210°C (5°C/min). The final temperature was held for 20 min. For compound separation, a SH-Stabilwax, 0.25 mm ID × 30 m length × 0.25 μm film thickness was selected. For compound identification, a gas chromatograph coupled to a mass spectrometer (GCMS-TQ8040, Shimadzu Deutschland GmbH, Duisburg, Germany) was used with identical GC settings. Eighteen compounds were identified with NIST 14 library (National Institute of Standards and Technology, MD, United States) and confirmed with analytical standards. The mass detector was run in the electron impact ionization mode (70 eV).

Carotenoids Analysis

The samples were ground (30 s at 30 Hz; Retsch, model: MM 400, Haan, Germany) using liquid nitrogen for cooling. 600 mg of frozen and ground sample were weighed in 50 ml centrifuge tubes. The analysis was carried out in accordance to Serio et al. (2007) with some modifications. Instead of nitrogen flux, the *n*-hexane/carotenoids mixture was vaporized (12.45 h) using a rotary vacuum concentrator (RVC 2-25 CD plus, Christ, Osterode am Harz, Germany). Subsequently, samples were dissolved in 1250 μl ethyl acetate/dichloromethane/n-hexane (80:16:4, v:v:v) and filtrated (0.45 μm PTFE syringe filter, VWR International, Radnor, PA, United States). The samples were stored at −80°C prior to analysis with HPLC (Jasco Labor- und Datentechnik GmbH, Gross-Umstadt, Germany). A calibration curve was prepared using lycopene (CAS-Nr. 502-65-8, Roth, Karlsruhe, Germany) and ß-carotene (CAS-Nr. 7235-40-7, Roth, Karlsruhe, Germany). The carotenoids were measured with an UV-detector using spectra 454 nm for ß-carotene and 474 nm for lycopene. For the separation of the carotenoids, a Chromolith® Performance RP-18e (100 × 4.6 mm) column was used. The flow rate was 1 ml/min with an oven

temperature of 28°C. As mobile phase acetonitrile/water/ethyl acetate (51:7:42, v:v:v) was used and the injection volume of the sample was 10 µl.

Fatty Acids

The fatty acids were analyzed in respect of the method of Thies (1971) with some modifications. 350 mg of freeze-dried material were weighed in 15 ml centrifuge tubes. 2 ml of 0.5 M Na-Methylat were added and vortexed. Subsequently, 400 µl 2,2,4-Trimethylpentane and 300 µl of 5% (w/v) $NaHSO_4$ were added and vortexed as well. The upper phase was pipetted in a 200 µl glass vial and stored at −20°C until analysis. For analysis, 0.6 µl of the sample was injected in the GC-FID (Thermo Electron Corporation, Trace GC Ultra; autosampler: A.L.S. 104). Injector and Detector were held at 250°C with a constant oven temperature of 205°C during analysis. Hydrogen was used as carrier gas and a Permabond FFAP-0.25 µm, 25 m × 0.25 mm column for separation. The amount of each fatty acid in the sample was expressed as a relative percentage to all determined fatty acids.

Conventional Profiling by a Trained Sensory Panel

The sensory analyses were performed in a sensory lab with separated booths set in daylight conditions. The sensory panel consisted of 12 experienced assessors who were selected in accordance with international ISO 8586 (ISO, 2012) guidelines. The assessors were trained twice a week for 4 weeks prior to the evaluations. During the first sessions, attributes were developed to describe tomato flavor (appearance, odor, taste, and aftertaste). A list of descriptors was compiled from terms proposed by the panelists. The following sessions were used to present different references to the assessors and reach a consensus. During the training, the following eight attributes were elaborated for tomato fruits: green-grassy odor, tomato-typical odor, tomato-typical flavor, sweetness, sourness, juiciness, firmness of the fruit peel, and aftertaste. During the tests, the panelists received four quarters of tomatoes served in small bowels. Fruits stored at 7°C were adjusted to room temperature before being presented to the panelists. All fruits were cut shortly before serving to preserve the aroma. All samples were served in two replications to the panelists. The assessors evaluated the products on an unstructured line with not perceptible (0%) to highly perceptible (100%). The panelists were provided with water and unsalted cracker (P. Heumann's Matzen, Germany) to neutralize the palate as well as coffee beans to neutralize the olfactory sense. During the test evaluation, each panelist sat in a separate booth in the sensory lab, which was designed according to the specifications of ISO 8589 (ISO, 2007).

Statistics

Statistical analysis was performed using SPSS statistical software (IBM statistics Version 25.0, Armond, NY, United States). The results of the chemical and sensory analyses were carried out using one-way and two-way analysis of variance (ANOVA), followed by Tukey's *Post Hoc* test ($p \leq 0.05$). The PCA (principal component analysis) was performed for the sensory evaluation using FactoMineR package. Pearson correlations and graphs were performed within ggplot2. R version 3.6.1 was used.

RESULTS

Fruit Quality Parameters

The five cultivars/breeding lines investigated are shown in **Figure 2**. Black Cherry (BC) and Primabella (P) are red-brown and red cocktail cultivars, while Roterno F1 (R) is a red salad cultivar. The breeding line Black Cherry × Primabella (BCxP) has a red-brown appearance and the size of a cocktail tomato fruit. On the other hand, the breeding line Black Cherry × Roterno F1 (BCxR) is red-pink and has the size of a salad fruit (**Figure 2**).

The cultivars/breeding lines showed significantly different contents in most quality parameters (**Supplementary Table S2**). With exception for cultivar R at 7°C, the lycopene and ß-carotene content increased after 20 and 7°C household storage although the difference was not significant (**Table 1**). The fructose and glucose concentrations did only increase significantly in one cultivar (P) after 20°C household storage (**Table 1**). The a-value (red color) increased significantly after the storages while the hue angle [H°] and the texture decreased significantly (**Table 1**). No significant change could be found in the citric and malic acid concentrations between the different post-harvest treatments (20 and 7°C) (**Table 1**). The fatty acid contents did not vary compared in fresh harvested fruits and fruits stored at 20 and 7°C (**Table 1** and **Supplementary Table S1**). The magnesium and phosphorus concentrations depended on the cultivar/breeding line and were positively correlated (**Supplementary Figure S1**). Bar plots of important quality parameters were shown in **Figure 3**. They display the cultivar/breeding line and storage regime effects. The ß-carotene content was significantly the highest in P regardless of the storage conditions (**Table 1** and **Figure 3A**). In BC, the ß-carotene content increased significantly after both storage regimes. P also showed a significant higher relative ß-ionone concentration than the other cultivars/breeding lines (**Figure 3B** and **Supplementary Table S2**). BC and BCxP contained the highest concentrations of fructose and glucose (**Figure 3D**). All quality parameters displayed in **Table 1** showed a significant cultivar/breeding line effect, and only fructose, glucose, malic acid, potassium, phosphorus, and magnesium content did not show a significant storage effect.

Volatiles and Important Precursors

The volatile profile was analyzed, identifying 18 different aroma compounds considered to contribute to the tomato flavor in fruits, as shown in **Table 2**. Significant changes of the aroma compounds were always found in single cultivars/breeding lines, depending on the cultivar/breeding line and the compound, but not in all cultivars/breeding lines (**Supplementary Table S1**). The aroma compounds hexanal, 6-methyl-5-hepten-2-one, (Z)-3-hexenol, 2-isobutylthiazole, benzaldehyde, and (E)-geranylacetone showed both cultivar/breeding line and storage effects (**Supplementary Table S1**). Benzaldehyde was the only aroma compound, which decreased in all cultivars/breeding lines after both household storage treatments (20 and 7°C)
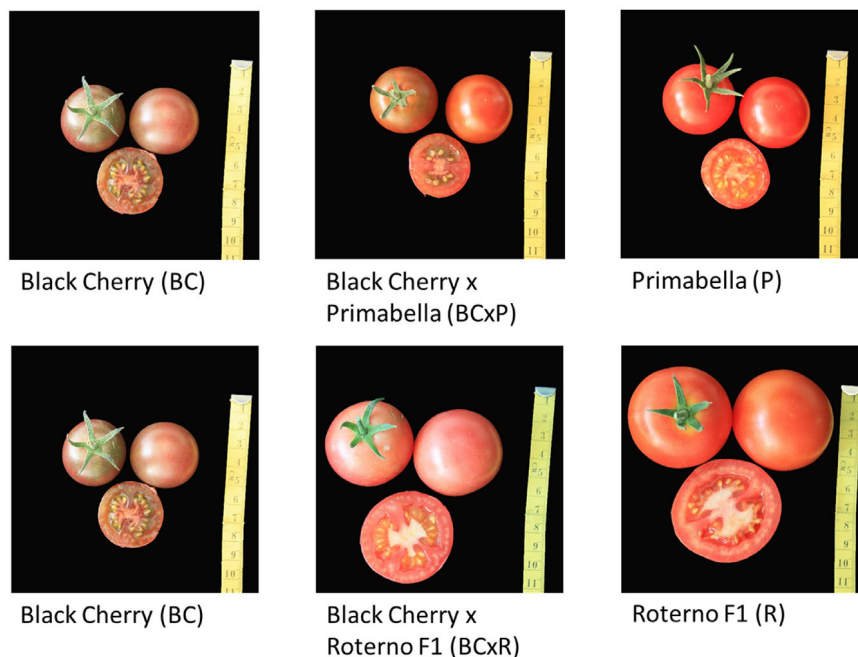
FIGURE 2 | The studied breeding lines and parental cultivars.

(**Table 2**). Nevertheless, a significant reduction could only been seen in P, BC, and BCxR. Hexanal increased significantly after both treatments, except in R, BCxR and BC at 7°C (**Table 2** and **Figure 3C**). The fatty acids-derived volatiles (*Z*)-3-hexenal and (*E*)-2-hexenal correlated positively (**Supplementary Table S4**), but the behavior during storage depended on the cultivar/breeding line as well (**Table 2**). The relative concentrations of the two aroma volatiles increased in BCxP but decreased in the other cultivars/breeding lines compared with the fresh fruits, except in BCxR at 7°C, where they also increased (**Table 2**). The relative 2-isobutylthiazole concentration was significantly the highest in the cultivar R (**Supplementary Table S1**). The relative concentration of the carotenoid-derived volatile 6-methyl-5-hepten-2-one increased after harvest at both storage conditions (20 and 7°C) (**Table 2**). 6-methyl-5-hepten-2-one showed a significantly positive correlation with lycopene (**Supplementary Table S3**), but not with ß-carotene (**Supplementary Table S3**). On the other hand, ß-ionone was positively correlated with ß-carotene but not with lycopene (**Supplementary Table S3**).

Sensory Evaluation

To show the relations between the cultivars/breeding lines and the different storage conditions, we performed a Principal Component Analysis (PCA) (**Figure 4**).

The biplot in **Figure 4A** illustrates the cultivars/breeding lines and explanatory variables. BC and BCxP correlated positively with Dimension 1 (Dim1), regardless of whether the fruits were stored at 20 or 7°C. R and BCxR correlated with the negative values of Dim1, while P correlated with the positive values of Dimension 2 (Dim2) (**Figure 4A**). We could separate the

cultivars/breeding lines into four groups (**Figure 4B**). BC_fresh, BC_7°C, BC_20°C, BCxP_fresh, BCxP_7°C and BCxP_20°C are relatively close together. The same becomes clear for the samples BCxR_fresh, BCxR_7°C, BCxR 20°C, R_7°C, and R_20°C on the opposite side of the first dimension. R_fresh, on the other hand, stands more by itself, and the last group consists of P_fresh, P_7°C, and P_20°C (**Figure 4B**). The cultivars/breeding lines that are on the same side of the given variable have a high value for those variables. **Figure 4C** shows the plotted sensory variables, while the different color shades represent the contribution of the variables in percentage terms to the principal components. Dim1 positively correlates with sweetness, aftertaste, tomato_taste, firmness, green_odor, and sourness. Tomato_odor is the variable more represented on Dim2 and correlated positively with its values (**Figure 4C**). The quality of representation of the cultivars and the breeding lines are plotted in **Figure 4B**. A high cos2 (square cosine) indicates a good representation of the individuals on the principal components. A comparison of sweetness and sourness with laboratory analyses showed significant correlations between the results of the human senses and the instrumental measurements (**Figures 5A,B**). The sensory analyses did not reveal significant storage effects but cultivar/breeding line effects (**Supplementary Table S5** and **Supplementary Figure S2**). BC and BCxP were significantly higher rated in sweetness, regardless of the post-harvest conditions (**Figures 5A, 6**). Measured sweetness and aftertaste by the panelists showed a positive correlation with the tomato-typical flavor (**Figures 5C,D**). BC and BCxR were rated significantly highest in tomato-like flavor (**Figure 6**) and R and BCxR were rated the lowest in the attribute aftertaste in the panel evaluation (**Figure 6**).

TABLE 1 | Fruit quality parameters (mean ± standard deviation) of the five cultivars/breeding lines shown for fresh harvested fruits (fresh) and after 20 and 7°C household storage with $n = 4$.

C/Bl	SR	Lycopene [μg/g FW]	ß-Carotene [μg/g FW]	Dry matter [%]	TSS [°Brix]	TA [%]	TSS/TA-Ratio	Fructose [mg/g FW]	Glucose [mg/g FW]	Citric acid [mg/g FW]	Malic acid [mg/g FW]	K [mg/g FW]	P [mg/g FW]	Mg [mg/g FW]	Fatty acid (18:2) [rel. conc. %]	Fatty acid (18:3) [rel. conc. %]	Texture [N]	a*-value	Hue-angle [°]
P	fresh	100.39 ± 5.5a	11.18 ± 3.2a	7.09 ± 0.1b	5.60 ± 0.2b	0.43 ± 0.1a	13.10 ± 1.3b	19.42 ± 2.0b	15.17 ± 1.7a	4.95 ± 1.0a	0.94 ± 0.2a	1.95 ± 0.2a	0.24 ± 0.0a	0.11 ± 0.0a	50.72 ± 0.5a	6.02 ± 0.5a	9.96 ± 0.4a	13.96 ± 0.5b	61.82 ± 1.1a
	20°C	160.45 ± 24.3a	14.66 ± 1.3a	7.79 ± 0.1a	6.28 ± 0.1a	0.37 ± 0.0a	17.21 ± 1.4a	22.75 ± 1.5a	17.63 ± 1.9a	3.89 ± 0.9a	1.08 ± 0.4a	2.12 ± 0.2a	0.25 ± 0.0a	0.11 ± 0.0a	50.14 ± 1.0a	6.05 ± 0.6a	5.30 ± 0.3c	16.74 ± 0.5a	55.67 ± 0.7b
	7°C	145.14 ± 50.3a	14.34 ± 0.9a	7.46 ± 0.3a	6.03 ± 0.2a	0.37 ± 0.0a	16.52 ± 1.5a	21.93 ± 1.3ab	16.37 ± 1.2a	4.29 ± 0.8a	1.01 ± 0.2a	1.91 ± 0.2a	0.19 ± 0.1a	0.10 ± 0.0a	50.65 ± 0.4a	6.06 ± 0.3a	6.14 ± 0.4b	15.80 ± 1.1a	57.13 ± 0.3b
BCxP	fresh	96.76 ± 23.8b	8.79 ± 2.3a	8.29 ± 0.3b	6.90 ± 0.3b	0.49 ± 0.1a	14.27 ± 2.1a	27.27 ± 4.1a	22.28 ± 3.5a	5.72 ± 0.8a	1.09 ± 0.2a	2.26 ± 0.3a	0.26 ± 0.0a	0.12 ± 0.0a	46.90 ± 1.2a	7.14 ± 0.8a	14.55 ± 0.3a	4.61 ± 1.2b	78.34 ± 2.9a
	20°C	148.73 ± 8.7a	9.35 ± 0.6a	8.80 ± 0.2a	7.60 ± 0.3a	0.47 ± 0.0a	16.01 ± 0.5a	28.13 ± 2.1a	22.22 ± 2.0a	5.00 ± 0.5a	1.01 ± 0.1a	2.39 ± 0.2a	0.29 ± 0.2a	0.13 ± 0.0a	47.82 ± 1.0a	6.17 ± 0.1a	8.22 ± 0.3c	6.96 ± 1.2a	71.87 ± 2.2b
	7°C	125.86 ± 33.1ab	9.79 ± 1.1a	8.69 ± 0.1ab	7.40 ± 0.2ab	0.45 ± 0.0a	16.47 ± 0.6a	27.51 ± 1.6a	21.83 ± 1.8a	6.54 ± 1.1a	1.19 ± 0.4a	2.31 ± 0.2a	0.26 ± 0.0a	0.12 ± 0.0a	48.71 ± 1.8a	6.36 ± 0.4a	10.06 ± 0.9b	6.15 ± 1.0ab	74.09 ± 2.3ab
BC	fresh	75.01 ± 16.1b	6.75 ± 0.6b	8.92 ± 0.2b	7.28 ± 0.1b	0.50 ± 0.0a	14.67 ± 0.2c	27.28 ± 1.9a	23.44 ± 1.5a	4.84 ± 0.3a	0.86 ± 0.2a	2.39 ± 0.1a	0.27 ± 0.0a	0.12 ± 0.0a	46.59 ± 0.8a	4.49 ± 0.3a	13.51 ± 0.9a	0.29 ± 0.4b	88.67 ± 1.5a
	20°C	116.95 ± 23.0a	8.03 ± 0.4a	9.10 ± 0.1ab	7.73 ± 0.1a	0.43 ± 0.0b	18.00 ± 0.3b	27.21 ± 1.5a	23.06 ± 1.6a	4.13 ± 0.4a	0.68 ± 0.1a	2.64 ± 0.3a	0.30 ± 0.0a	0.12 ± 0.0a	46.81 ± 0.8a	4.80 ± 0.3a	8.65 ± 0.7b	2.02 ± 0.4a	82.18 ± 1.5b
	7°C	81.00 ± 19.3ab	8.12 ± 0.6a	9.26 ± 0.1a	7.60 ± 0.2a	0.40 ± 0.0c	18.93 ± 0.5a	29.36 ± 0.9a	25.28 ± 1.0a	4.65 ± 0.8a	0.69 ± 0.3a	2.53 ± 0.2a	0.31 ± 0.0a	0.13 ± 0.0a	46.52 ± 0.8ab	4.44 ± 0.2a	9.26 ± 0.8b	1.35 ± 0.4a	84.43 ± 1.7b
BCxR	Fresh	114.98 ± 17.3a	5.99 ± 1.3a	6.84 ± 0.2a	5.65 ± 0.2a	0.33 ± 0.0a	16.94 ± 1.3a	21.53 ± 2.1a	17.96 ± 1.8a	3.57 ± 0.4a	0.82 ± 0.1a	1.82 ± 0.2a	0.20 ± 0.0a	0.09 ± 0.0a	44.17 ± 1.0a	6.96 ± 0.4a	12.29 ± 0.5a	7.29 ± 1.3b	68.68 ± 3.4a
	20°C	155.25 ± 41.9a	6.02 ± 0.6a	6.41 ± 0.2b	5.53 ± 0.3a	0.28 ± 0.0b	19.92 ± 1.4a	20.87 ± 1.7a	16.44 ± 1.6a	3.08 ± 0.4a	1.03 ± 0.0a	1.78 ± 0.3a	0.19 ± 0.0a	0.09 ± 0.0a	45.15 ± 1.0a	6.34 ± 0.8ab	8.09 ± 0.3b	11.11 ± 1.2a	57.51 ± 2.3b
	7°C	126.99 ± 55.6a	6.95 ± 0.1a	6.67 ± 0.2ab	5.58 ± 0.3a	0.32 ± 0.0ab	17.77 ± 2.1a	20.38 ± 2.2a	16.31 ± 1.1a	4.20 ± 0.9a	1.03 ± 0.3a	1.84 ± 0.2a	0.20 ± 0.0a	0.09 ± 0.0a	44.68 ± 1.0a	5.87 ± 0.3b	8.87 ± 1.4b	10.18 ± 1.0a	59.88 ± 1.6b
R	Fresh	106.26 ± 14.0a	5.77 ± 1.3a	5.90 ± 0.2b	5.10 ± 0.1b	0.30 ± 0.0a	17.00 ± 1.3a	17.69 ± 2.4a	15.43 ± 2.0a	3.89 ± 1.3a	1.33 ± 0.4a	1.65 ± 0.1a	0.19 ± 0.0a	0.09 ± 0.0a	46.25 ± 1.3b	6.77 ± 1.5a	18.54 ± 0.8a	15.85 ± 0.4b	61.27 ± 0.5a
	20°C	127.92 ± 26.0a	7.04 ± 1.0a	6.58 ± 0.1a	5.48 ± 0.2a	0.31 ± 0.0a	17.71 ± 1.1a	19.37 ± 1.0a	16.01 ± 1.3a	3.52 ± 0.4a	1.18 ± 0.3a	1.85 ± 0.2a	0.20 ± 0.2a	0.10 ± 0.0a	48.69 ± 1.2a	5.85 ± 0.7a	12.06 ± 1.0b	17.55 ± 0.8a	56.93 ± 0.9b
	7°C	105.94 ± 9.9a	7.43 ± 0.9a	6.50 ± 0.3a	5.20 ± 0.2ab	0.28 ± 0.0a	18.57 ± 1.6a	17.70 ± 2.4a	15.03 ± 2.1a	4.47 ± 1.5a	1.65 ± 0.7a	1.67 ± 0.1a	0.20 ± 0.0a	0.09 ± 0.0a	47.98 ± 0.7ab	5.72 ± 0.2a	12.60 ± 1.0b	17.44 ± 0.8a	58.04 ± 1.5b
C/Bl		**	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***
SR		***	**	***	***	***	***	ns	ns	**	ns	ns	ns	ns	*	**	***	***	***
C/Bl × SR		ns	ns	***	*	*	*	ns	ns	ns	ns	ns	ns	ns	ns	ns	*	ns	*

Different letters indicate significant differences between fresh harvested fruits, after 20 and 7°C household storage for each cultivar/breeding line (Tukey-Test $p \leq 0.05$). TSS, total soluble solids; TA, titratable acidity; C, cultivar; Bl, breeding line; SR, storage regime; K, potassium; P, phosphorus; Mg, magnesium. ns, not significant, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.
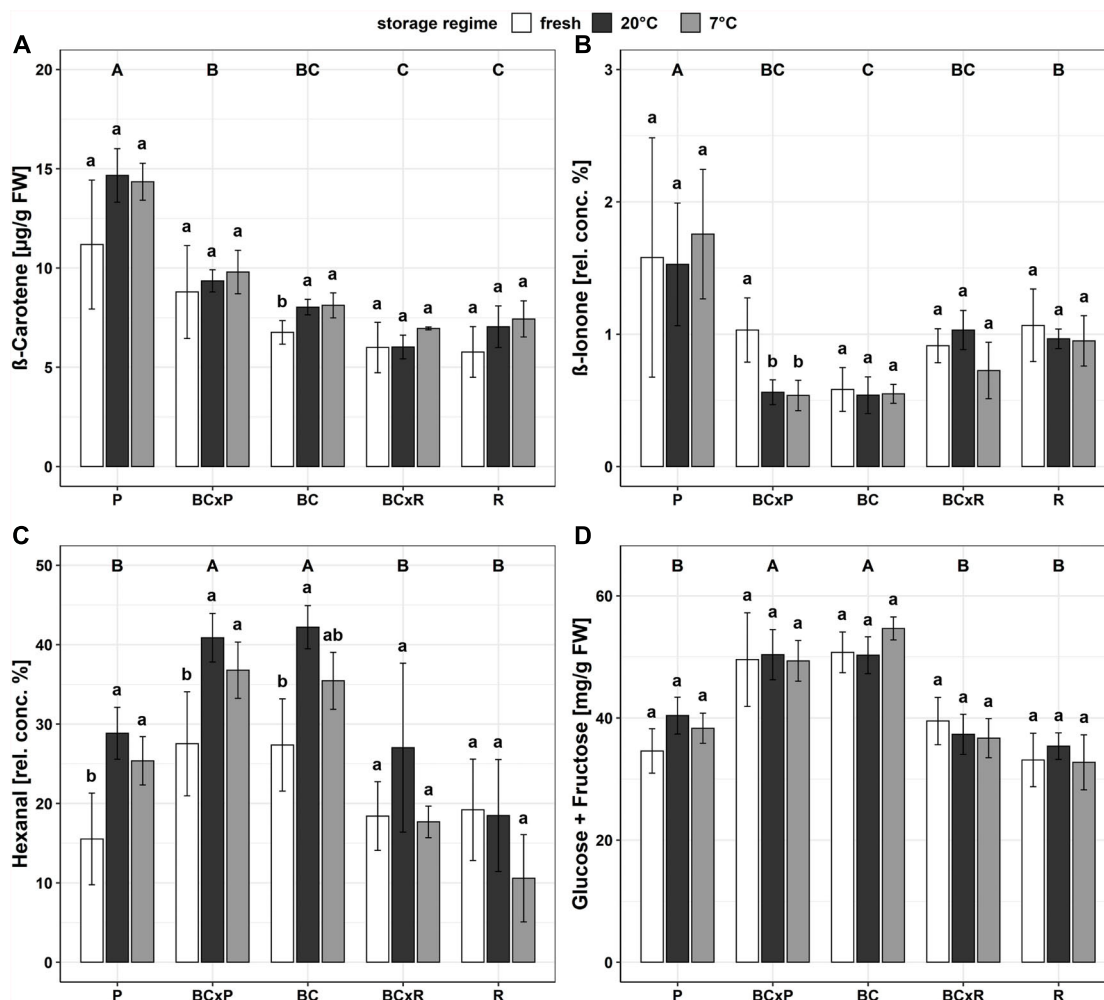
FIGURE 3 | Bar plots of the different breeding lines and parental cultivars for different quality parameters, ß-carotene content **(A)**, relative concentration of ß-ionone **(B)** relative concentration of hexanal **(C)**, and the sum of glucose and fructose content **(D)**. Values are means of n = 4 ± standard deviation; Tukey-Test $p \leq 0.05$.

Sensory Evaluation and E-Tongue Results

The electronic tongue (e-tongue) has been applied to measure taste with regard to the five basic tastes of the human senses (sweetness, sourness, bitterness, saltiness, and umami). We compared the output of the e-tongue results for sweetness and sourness to the sweetness and sourness perception of a trained sensory panel. The results from the sensory panel and the measured data from the electronic tongue showed a significant positive correlation for sweetness ($r = 0.82$) and sourness ($r = 0.52$) (**Figure 7**).

DISCUSSION

In the present study, the entire transportation route of tomatoes from harvest via retailer (distributor) to retail to the consumer was evaluated. We focused on the short-term fruit storage, considering the entire post-harvest chain and studied the impact

of two typical household storage conditions (20 and 7°C) in this context. Harvesting in practice and following a typical post-harvest chain, which includes one day at the distributor (12.5°C) and two days at the retail (20°C) before reaching the consumer. Two new breeding lines and their parental cultivars were evaluated, because little is still known about the fruit flavor behavior when fruits are harvested ripe and undergo the whole transportation route.

Influence of Storage Conditions on Important Fruit Quality Parameters of New Breeding Lines and Their Parental Cultivars

The influence of the cultivars/breeding lines was higher than that of the different postharvest handling during simulated household storage. We found high correlations between the measured total soluble solids and the analyzed fructose and glucose concentrations in the fruits with regard to all cultivars/breeding

TABLE 2 | Eighteen aroma compounds in tomato fruits and their relative fold changes after storage at 20 and 7°C compared to fresh harvested fruits are shown.

Aroma compounds	20°C					7°C				
	P	BC × P	BC	BC × R	R	P	BC × P	BC	BC × R	R
1-Penten-3-one	0.67	0.78	0.64	0.57	0.67	0.56	1.33	0.79	1.36	0.67
Hexanal	1.86	1.49	1.54	1.47	0.96	1.63	1.34	1.3	0.96	0.55
(Z)-3-Hexenal	0.65	1.64	0.63	0.78	0.6	0.27	3.03	0.77	1.9	0.52
(E)-2-Hexenal	0.66	1.11	0.63	0.6	0.54	0.26	1.84	0.77	1.36	0.45
(E)-2-Heptenal	0.96	1.52	0.76	0.64	1.04	0.59	1.56	1.06	1.03	0.77
6-Methyl-5-hepten-3-one	2.16	1.16	1.89	1.26	1.12	2.69	0.91	1.95	1.18	1.38
1-Hexanol	1.27	1.1	0.94	1.76	1.95	1.3	0.94	0.88	0.76	1.56
(Z)-3-Hexenol	0.52	0.49	0.53	0.99	0.98	0.67	0.55	0.68	0.71	1.42
2-Isobutylthiazole	1.36	1.07	0.79	0.77	1.35	1.43	0.95	0.83	1.17	1.46
Benzaldehyde	0.38	0.31	0.31	0.43	0.6	0.52	0.31	0.34	0.57	1
Phenylacetaldehyde	1.19	2.71	1.56	0.62	0.69	1.25	1.59	2.19	0.48	0.65
Neral	1.6	1.11	1.44	0.92	0.91	1.64	0.92	1.44	0.92	0.85
Geranial	2	1.16	1.7	1.09	1.03	1.86	0.87	1.59	0.85	0.81
Methyl salicylate	1.05	0.54	0.59	1.46	1	1.6	0.64	0.48	0.76	1.25
ß-Damascenone	0.56	0.52	0.81	0.99	0.71	0.71	0.5	0.82	1	0.96
(E)-Geranylacetone	1.16	1.16	1.27	1.35	1.08	1.22	0.9	1.14	0.89	0.89
2-Phenylethanol	0.66	1.16	0.86	1.12	1.45	0.68	0.81	0.57	0.87	1.3
ß-Ionone	0.97	0.54	0.93	1.13	0.91	1.11	0.52	0.95	0.79	0.89

In total, 71 detected aroma compounds were normalized to 100% and the relative concentrations were used for the calculation of the relative fold changes (FC) in each aroma compound (n = 4); the color scale ranges from 3 (orange) to 0.2 (dark blue) and corresponds to the FC.

lines as well as significant correlations between titratable acidity as well as citric acid and malic acid (data not shown). These results also clearly showed the variation in these compounds with regard to the cultivars/breeding lines and the importance of these taste-related compounds, which are well-shown in various studies (e.g., Malundo et al., 1995; Piombino et al., 2013; Baldwin et al., 2015). The dry matter content positively correlated with total soluble solids and these were positively correlated to fructose and glucose concentrations and the sweetness perception recorded by the trained panel (data not shown). Enhancing the dry matter content of the fruits could be an interesting approach toward flavor enhancement. Tieman et al. (2017) showed in their study that the negative correlation of fruit weight and sugar content could be linked to the reduction of the high-sugar alleles caused by enhancing the fruit size during breeding. Consistent with our study, Zhang et al. (2016) found no alteration in sugar or acid concentrations after cold storage (5°C). In the present study, the total soluble solids content and the *a*-value (red color) tend to be higher in the stored fruits compared to the fruits analyzed directly after harvest. These results are consistent with Verheul et al. (2015). We found differences in the analyzed parameters between fresh and stored fruits, but no difference between the two short-term household storage regimes, when all fruits pass the same transport route before being stored by the consumer at different temperatures – e.g., at room temperature or in the refrigerator.

Volatiles and Important Precursors

Important volatiles contributing to the tomato flavor of fruits are derived from carotenoids and fatty acids (Farneti et al., 2015) though the pathway of the exact biosynthesis of all aroma volatiles has not been clarified yet (Mathieu et al., 2009). The carotenoid-derived volatiles are produced by cleavage of the carotenoids present in the fruits, whereas the most abundant carotenoids are lycopene and ß-carotene (Lewinsohn et al., 2005; Klee and Tieman, 2018). In our study, we analyzed these main carotenoids, which are precursors of ß-ionone, geranylacetone, geranial, and 6-methyl-5-hepten-2-one (Farneti et al., 2015). Apocarotenoid volatiles can be separated into linear apocarotenoids—such as geranylacetone and 6-methyl-5-hepten-2-one—and cyclic apocarotenoids—namely ß-ionone—and are positively linked to flavor acceptability, having fruity and floral perceptions (Vogel et al., 2010). We found the highest concentrations of both ß-carotene and ß-ionone in the cultivar P. That can be explained by the observation that apocarotenoid volatiles and carotenoid precursors are shown to be proportional to each other (Vogel et al., 2010). The volatile ß-ionone is a direct breakdown product of ß-carotene (Lewinsohn et al., 2005) and so, is related to it. Our results are consistent with studies from Baldwin et al. (2000) and Vogel et al. (2010). With respect to the entire transportation route in our study, most carotenoid-derived volatiles increased slightly during both household conditions viz refrigeration (7°C) and room temperature (20°). A significant increase was only found for the volatile 6-methyl-5-hepten-2-one and only in cultivar P, compared to its content in fresh fruits analyzed directly after harvest. These results were similar with the results from Farneti et al. (2015), who outlined that the carotenoid-derived aroma compounds, e.g., geranylacetone and ß-ionone, responded less severely and were in red ripe tomatoes more cultivar-dependent during cold storage conditions. During the storage at 16°C, these volatiles increased, while they constantly decreased
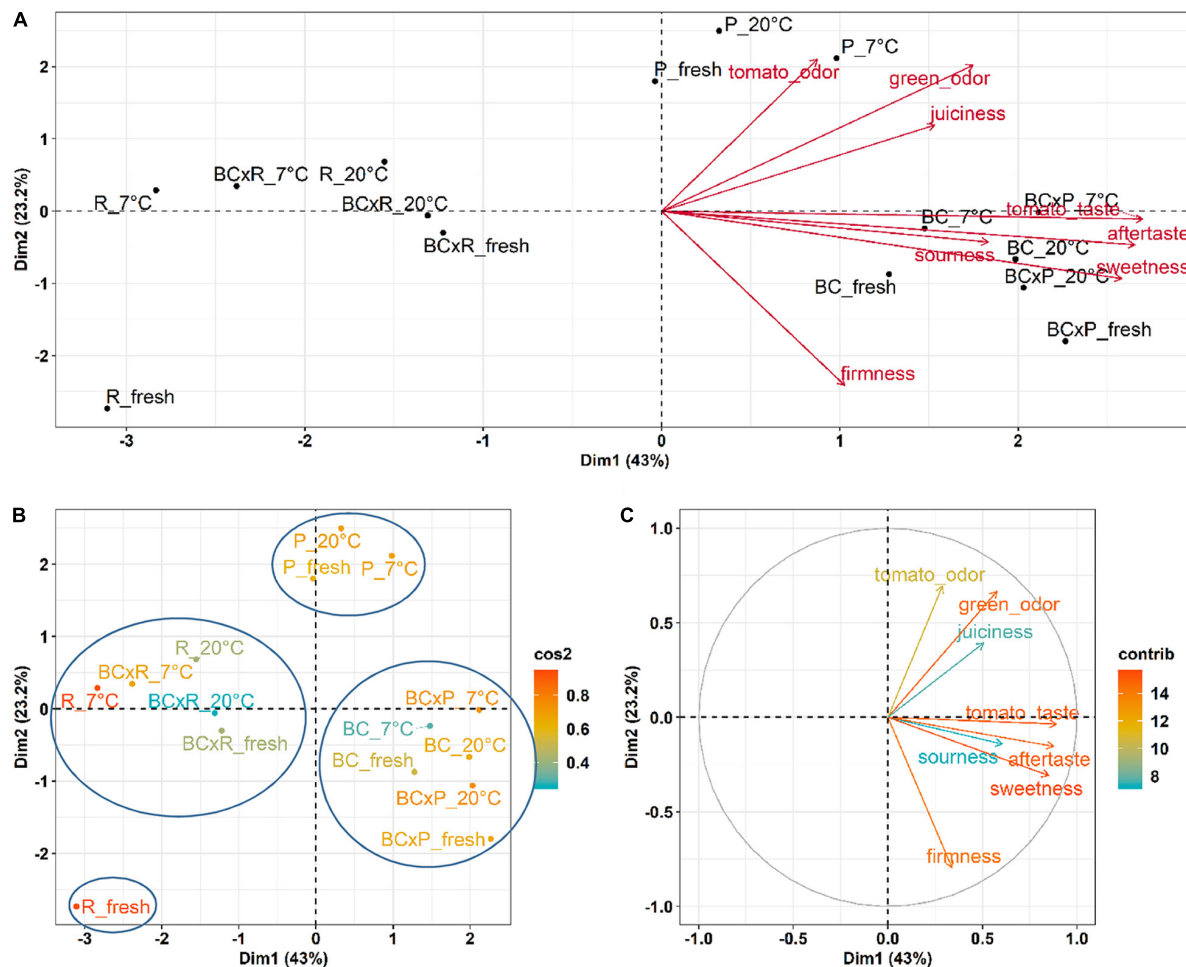
FIGURE 4 | PCA results for the sensory evaluation of the different cultivars/breeding lines compared fresh fruits and fruits stored at 20 and 7°C. PCA show the cultivars/breeding lines and the explanatory variables **(A)**, the quality of representation (cos2) of the cultivars/breeding lines on the two dimensions **(B)**, and the contribution of the sensory variables to the principal components in percentage **(C)**.

during 4°C and Farneti et al. (2015) discussed the observed accumulation of carotenoid-derived volatiles during 16°C storage as a consequence of postharvest ripening. Our results did not show a significant increase in lycopene and ß-carotene content but tend to increase. We found a significant higher coloration (increased *a*-value), which is directly linked to the lycopene content in the fruits (Arias et al., 2000). Nevertheless, we could not find significant differences in the relative content of 6-methyl-5-hepten-2-one, geranial, ß-ionone, and ß-damascenone in the cultivars/breeding lines after both storage regimes. In general, the behavior of the studied aroma compounds was not consistent during the two household storages, with one possible reason being that the cultivars/breeding lines respond differently to the treatments. Another important group of precursors are fatty acids. Whereas, the fatty acid-derived volatiles are formed during cleavage of linoleic (C18:2) and linolenic acid (C18:3) by lipoxygenase, which catalyzes the first step of the fatty acid degradation (Renard et al., 2013). This pathway is the origin of the C6 volatiles in tomato fruits, that include e.g., (*Z*)-3-hexenal,

(*E*)-2-hexenal, hexanal (Tieman et al., 2012; Renard et al., 2013; Klee and Tieman, 2018). Hexanal is the most abundant volatile in tomato fruits and has been described as "green, grassy" (Ruiz et al., 2005). During our storage study with five cultivars/breeding lines, we found a significant increase in the relative amount of hexanal in the fruits of P, BCxP after either 20 or 7°C household storage and in the fruits of BC at 20°C. In the fruits of BCxR and R no significant change could be observed, which might be a cultivar/breeding line effect. (*Z*)-3-hexenal and (*E*)-2-hexenal did not vary significantly after both storage regimes. Renard et al. (2013) observed an increase in hexanal as well during the 20°C storage regime of red-ripe tomatoes, while (*Z*)-3-hexenal and (*E*)-2-hexenal did not change. In contrast to our study, hexanal decreased during cold storage (4°C) (Renard et al., 2013). In the study of Farneti et al. (2015), commercially grown red ripe tomatoes were stored at 16°C up to 20 days. The level of hexanal thereby did not show a significant change but (*E*)-2-hexenal constantly decreased during the storage period. In a study from Ruiz et al. (2005) with one
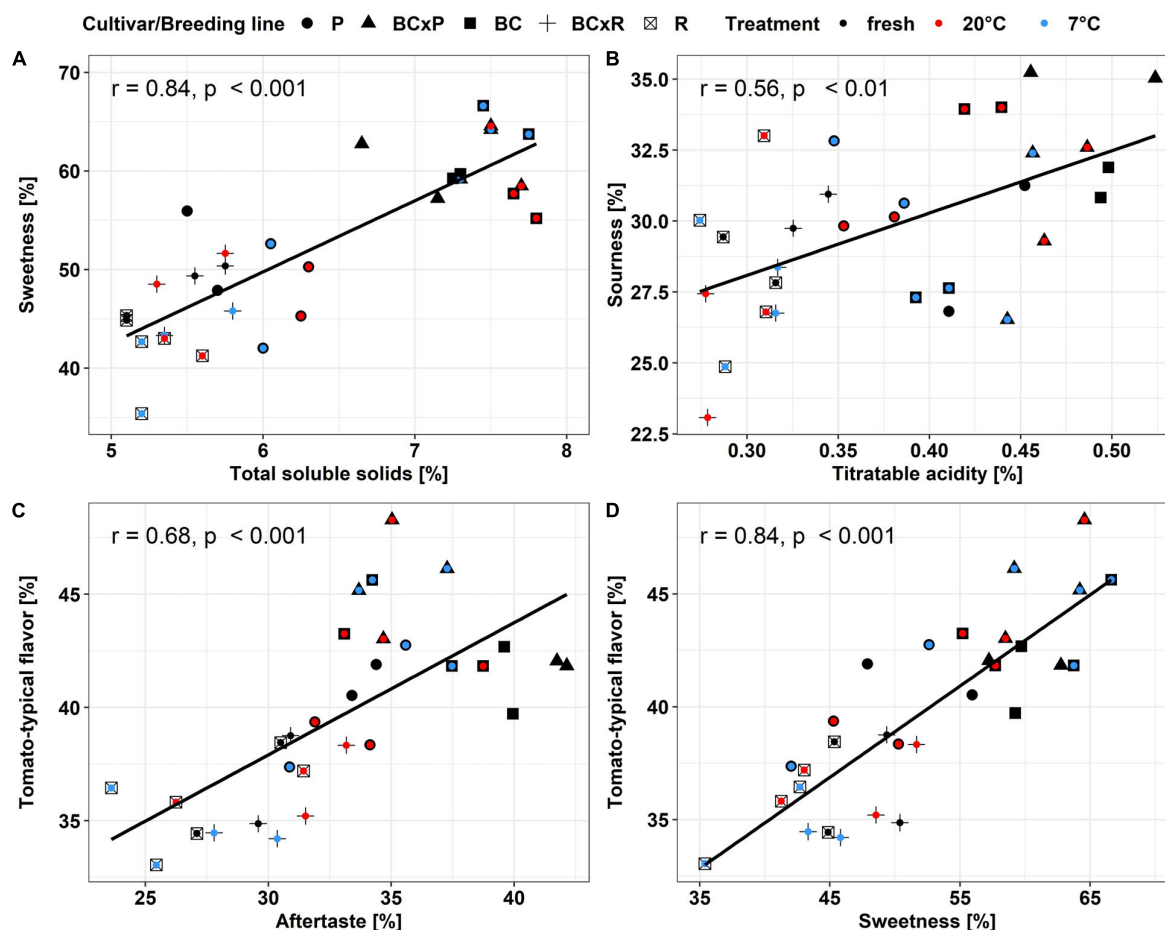
FIGURE 5 | Pearson-correlation of the evaluated sensory attributes, total soluble solids and sweetness **(A)**, titratable acidity and sourness **(B)**, aftertaste and tomato-typical flavor **(C)** and sweetness and tomato-typical flavor **(D)** in the different breeding lines and parental cultivars. Values are means of 10–12 panelists; correlations were significant with $p < 0.05$.

commercial and four traditional varieties they observed that the variety with the highest hexanal and (Z)-3-hexenal content got the highest rankings for "flavor" and "overall acceptability" during a sensory test with untrained tasters. In an additional study by Baldwin et al. (2015), hexanal enhanced the overall flavor in combination with the sweet/sour, TSS/TA ratio. In the present study, the highest rated cultivars/breeding lines in sweetness and tomato-like flavor also contained the highest relative amount of hexanal. The fatty acid contents from the fresh and stored fruits (20 and 7°C), on the other hand, did not show a notable shift in the composition, whereas another study from Renard et al. (2013) observed an increase in the linoleic (C18:2) and linolenic (C18:3) acid concentrations during storage at 4°C. Our results indicate that the behavior of the fruit during cold storage (7°C) is also strongly dependent on the cultivar/breeding line. For example, the up- or down-regulation and restoration of volatiles, namely carotenoid or fatty acid-derived volatiles, underlines the great impact of the cultivar on the flavor of the fruits and the acceptance by the consumer. We did not observe the severe negative effect of

cold storage compared to some other studies, which could be linked to the studied cultivars/breeding lines as well as to the chosen short-term storage regime and the fact that the fruits were harvested ripe.

Sensory Evaluation of the Breeding Lines and Their Parental Cultivars

In our study, the results from instrumental analyses of total soluble solids and titratable acidity, reflecting the sugar and acid content of tomato fruits, showed high correlations to the sweetness and sourness perception of the fruits elevated by the sensory panel. These results are confirmed by Baldwin et al. (2015), who analyzed 38 tomato genotypes over 7 years. Their study also emphasized that higher sweetness and sugar contents correlated positively to overall flavor, which is comparable to our data within the positive correlation of sweetness and tomato-typical flavor ratings. Principal component analyses of the five studied cultivars/breeding lines analyzed directly after harvesting ripe fruits, as well as stored at room temperature or
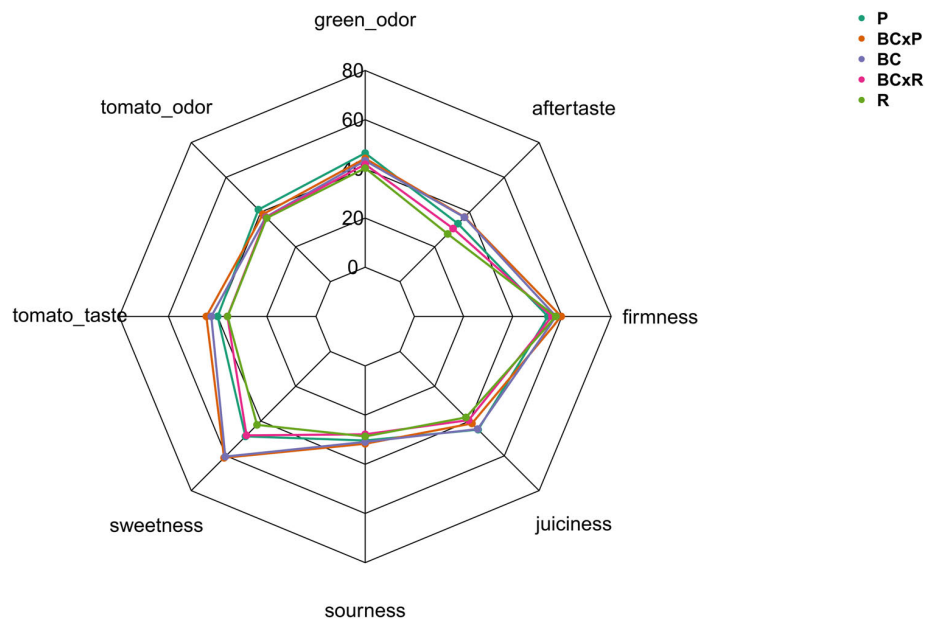
FIGURE 6 | Sensory evaluation of five different breeding lines/cultivars. Evaluation was conducted with 10–12 panelists.

in cold storage, following the post-harvest chain, showed the discrimination between the cultivars/breeding lines. The results from the sensory evaluation showed no significant differences between the two storage regimes as well as between stored and fresh fruits. In contrast to the cultivars and breeding lines that were significantly different. Krumbein et al. (2004), who investigated the effects of a household condition (20°C) on aroma and sensory attributes of three different tomato cultivars up to 21 days, found changes in the aroma volatiles, namely an increase in hexanal and 2-isobutylthiazole, and in the investigated sensory attributes, including odor, flavor, and aftertaste. In contrast,

FIGURE 7 | Comparison of sensory measured results and e-tongue values for sweetness and sourness (Pearson-correlation) within fresh fruits and fruits stored at either 20°C or 7°C. E-tongue results are mapped with six out of seven sensors (without sensor AHS). Results of fresh fruits, after 20 and 7°C storage are included in the correlation; correlations were significant with $p < 0.05$.

Auerswald et al. (1999) revealed no change in the characteristic flavor, mouthfeel, and aftertaste after four and seven days at 20°C in a consumer evaluation, which showed that differences were not perceived from the human senses. A consumer panel test, which evaluated fruits refrigerated at 5°C for seven days, followed by a one day recovery period at 20°C, showed significant lower ratings in the overall liking and illustrated the adverse impact of cold storage (Zhang et al., 2016). Furthermore, the fruits analyzed in the aforementioned study were evaluated already after one and three days of cold treatment as well, which showed no significant effect in loss of volatile compounds. This is comparable to our observed results. We could not find significant differences between the two post-harvest conditions in respect of their sensory attributes, but differences between the cultivars/breeding lines. The variation is visualized in the PCA, showing that the cultivar R and the breeding line BCxR were less associated with the attributes sweetness, tomato-like flavor, and aftertaste and the third group with regard to P was more associated with the attributes green/grassy odor and tomato-like odor. The results of the fruits from the R_fresh deviate from this. They were rated differently compared to the stored fruits of this cultivar. This could be caused by after-ripening effects. Therefore, cultivars with improved flavor composition are a target for breeders, as the strong impact of the cultivar on flavor could be outlined in the present study.

Comparison of Sensory Evaluation and E-Tongue Results

The electronic tongue (e-tongue) is used to evaluate the five basic tastes—sweet, sour, salty, bitter, and umami—in food and beverages and meant to mimic taste perceptions of humans

(Baldwin et al., 2011; Xu et al., 2018). Xu et al. (2018) evaluated four different tomato cultivars at six maturity stages and after refrigeration and blanching, looking at the possibilities of discrimination via the e-tongue. The utilized sensor set comprised the following sensors: ZZ, JE, BB, CA, GA, HA, and JB and they successfully predicted the TSS levels in tomatoes. However, with regard to the correlation to TA, the sensors seemed less reliable. We found both, significant correlations for the sensory attributes sweetness and sourness, obtained with a trained sensory panel, compared to the e-tongue results. Nevertheless, the strength of the correlation with the e-tongue sensors was stronger for sweetness than for sourness. Beullens et al. (2008) predicted individual taste compounds (glucose, fructose, citric acid, malic acid, glutamic acid, sodium, and potassium), which did not show satisfactory results, except for glutamic acid and sodium, while the correlations for the tomato taste-related attributes to sensory panel evaluation showed a better result. The results in the present study show that the classification of the tested tomato cultivars/breeding lines and the prediction of tomato taste of at least sweetness and sourness is possible, which was also revealed in similar studies (Beullens et al., 2008; Baldwin et al., 2011). The e-tongue, therefore, could be an interesting tool for the evaluation and discrimination of these two important quality attributes in tomato fruits.

In summary, considering the numerous, diverse discussions about tomato flavor, we see the difficulty of this complex topic and that many factors influence this sensitive quality parameter. Taking the whole transportation route into account, the difference between fruits stored for four days at 20 or 7°C during household storage does not have a notably influence on the human perception when fruits were harvested ripe. We showed that flavor is severely dependent on the cultivar and that crossing cultivars with enhanced flavor perception is a valuable step to improve flavor perception. The next step is to look on the entire transportation route from the producer to the consumer, finding a way to preserve the flavor of the tomato fruits. We could show that harvesting ripe fruits and storing them only for a short duration, even at 7°C, can preserve tomato flavor. The e-tongue could be used to generate taste contributors and function as a supporter for flavor improvement.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

REFERENCES

Arias, R., Lee, T.-C., Logendra, L., and Janes, H. (2000). Correlation of lycopene measured by HPLC with the L*, a*, b* color readings of a hydroponic tomato and the relationship of maturity with color and lycopene content. *J. Agric. Food Chem.* 48, 1697–1702. doi: 10.1021/jf990974e

Auerswald, H., Peters, P., Brückner, B., Krumbein, A., and Kuchenbuch, R. (1999). Sensory analysis and instrumental measurements of short-term stored tomatoes (Lycopersicon esculentum Mill.). *Postharvest Biol. Technol.* 15, 323–334. doi: 10.1016/s0925-5214(98)00094-5

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethikkommission, University of Göttingen, P.O. Box 37 44, 37027 Göttingen, Chair: Prof. Dr. Hans Michael Heinig, Office: Dr. Michael Müller Bahns, Research Department. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00472/full#supplementary-material

Baldwin, E. A., Bai, J., Plotto, A., and Dea, S. (2011). Electronic Noses and Tongues: applications for the Food and Pharmaceutical Industries. *Sensors* 11, 4744–4766. doi: 10.3390/s110504744

Baldwin, E. A., Goodner, K., and Plotto, A. (2008). Interaction of volatiles, sugars, and acids on perception of tomato aroma and flavor descriptors. *J. Food Sci.* 73, S294–S307. doi: 10.1111/j.1750-3841.2008.00825.x

Baldwin, E. A., Scott, J. W., and Bai, J. (2015). Sensory and chemical flavor analyses of tomato genotypes grown in florida during three different growing seasons in multiple years. *J. Am. Soc. Hortic. Sci.* 140, 490–503. doi: 10.21273/jashs.140.5.490

Baldwin, E. A., Scott, J. W., Shewmaker, C. K., and Schuch, W. (2000). Flavor trivia and tomato aroma: biochemistry and possible mechanisms for control of important aroma components. *HortScience* 35, 1013–1021.

Beullens, K., Mészáros, P., Vermeir, S., Kirsanov, D., Legin, A., Buysens, S., et al. (2008). Analysis of tomato taste using two types of electronic tongues. *Sens. Actuators B Chem.* 131, 10–17. doi: 10.1016/j.snb.2007.12.024

Dew, R., Seal, C. J., and Brandt, K. (2016). Effects of temperature conditions during transport and storage on tomato fruit quality. *Acta Hortic.* 10, 317–322. doi: 10.17660/ActaHortic.2016.1120.48

Díaz de León-Sánchez, F., Pelayo-Zaldívar, C., Rivera-Cabrera, F., Ponce-Valadez, M., Ávila-Alejandre, X., Fernández, F. J., et al. (2009). Effect of refrigerated storage on aroma and alcohol dehydrogenase activity in tomato fruit. *Postharvest Biol. Technol.* 54, 93–100. doi: 10.1016/j.postharvbio.2009.07.003

European Council, (2008). *Commission Regulation (EC) No 889/2008 of 5 September 2008 Laying Down Detailed Rules for the Implementation of Council Regulation (EC) No 834/2007 on Organic Production and Labelling of Organic Products with Regard to Organic Production, Labelling and Control OJ L 250.* Brussels: European Council, 1–84.

FAOSTAT, (2019). *FAOSTAT, Food and Agriculture Organization of the United Nations, Statistics Division. [Online].* Rome: FAO.

Farneti, B., Alarcón, A. A., Papasotiriou, F. G., Samudrala, D., Cristescu, S. M., Costa, G., et al. (2015). Chilling-induced changes in aroma volatile profiles in tomato. *Food Bioprocess Technol.* 8, 1442–1454. doi: 10.1007/s11947-015-1504-1501

ISO, (2007). *ISO 8589:2007 + A1:2014: Sensory analysis – General guidance for the design of test rooms; German version EN ISO 8589:2010 + A1:2014.* Geneva: ISO.

ISO, (2012). *ISO 8586:2012: Sensory analysis – General Guidelines for the Selection, Training and Monitoring of Selected Assessors and Expert Sensory Assessors; German Version EN ISO 8586, 2014.* Geneva: ISO.

Kader, A. A. (2008). Flavor quality of fruits and vegetables. *J. Sci. Food Agric.* 88, 1863–1868. doi: 10.1002/jsfa.3293

Klee, H. J., and Tieman, D. M. (2013). Genetic challenges of flavor improvement in tomato. *Trends Genet.* 29, 257–262. doi: 10.1016/j.tig.2012.12.003

Klee, H. J., and Tieman, D. M. (2018). The genetics of fruit flavour preferences. *Nat. Rev. Genet.* 19, 347–356. doi: 10.1038/s41576-018-0002-5

Koch, M., Busse, M., Naumann, M., Jákli, B., Smit, I., Cakmak, I., et al. (2019). Differential effects of varied potassium and magnesium nutrition on production and partitioning of photoassimilates in potato plants. *Physiol. Plant.* 166, 921–935. doi: 10.1111/ppl.12846

Krumbein, A., Peters, P., Brückner, B. (2004). Flavour compounds and a quantitative descriptive analysis of tomatoes (Lycopersicon esculentum Mill) of different cultivars in short-term storage. *Postharvest Biol. Technol.* 32, 15–28. doi: 10.1016/j.postharvbio.2003.10.004

Lewinsohn, E., Sitrit, Y., Bar, E., Azulay, Y., Meir, A., Zamir, D., et al. (2005). Carotenoid pigmentation affects the volatile composition of tomato and watermelon fruits, as revealed by comparative genetic analyses. *J. Agric. Food Chem.* 53, 3142–3148. doi: 10.1021/jf047927t

Malundo, T. M. M., Shewfelt, R. L., and Scott, J. W. (1995). Flavor quality of fresh tomato (Lycopersicon esculentum Mill.) as affected by sugar and acid levels. *Postharvest Biol. Technol.* 6, 103–110. doi: 10.1016/0925-5214(94)00052-t

Mathieu, S., Cin, V. D., Fei, Z., Li, H., Bliss, P., Taylor, M. G., et al. (2009). Flavour compounds in tomato fruits: identification of loci and potential pathways affecting volatile composition. *J. Exp. Bot.* 60, 325–337. doi: 10.1093/jxb/ern294

Maul, F., Sargent, S. A., Sims, C. A., Baldwin, E. A., Balaban, M. O., and Huber, D. J. (2000). Tomato flavor and aroma quality as affected by storage temperature. *J. Food Sci.* 65, 1228–1237. doi: 10.1111/j.1365-2621.2000.tb10270.x

Olbricht, K., Grafe, C., Weiss, K., and Ulrich, D. (2007). Inheritance of aroma compounds in a model population of Fragaria x ananassa Duch. *Plant Breed.* 127, 87–93. doi: 10.1111/j.1439-0523.2007.01422.x

Pék, Z., Helyes, L., and Lugasi, A. (2010). Color changes and antioxidant content of vine and postharvest-ripened tomato fruits. *HortScience* 45, 466–468. doi: 10.21273/hortsci.45.3.466

Piombino, P., Sinesio, F., Moneta, E., Cammareri, M., Genovese, A., Lisanti, M. T., et al. (2013). Investigating physicochemical, volatile and sensory parameters playing a positive or a negative role on tomato liking. *Food Res. Int.* 50, 409–419. doi: 10.1016/j.foodres.2012.10.033

Ponce-Valadez, M., Escalona-Buendía, H. B., Villa-Hernández, J. M., de León-Sánchez, F. D., Rivera-Cabrera, F., Alia-Tejacal, I., et al. (2016). Effect of refrigerated storage (12.5°C) on tomato (*Solanum lycopersicum*) fruit flavor: a biochemical and sensory analysis. *Postharvest Biol. Technol.* 111, 6–14. doi: 10.1016/j.postharvbio.2015.07.010

Renard, C. M. G. C., Ginies, C., Gouble, B., Bureau, S., and Causse, M. (2013). Home conservation strategies for tomato (*Solanum lycopersicum*): storage temperature vs. duration – Is there a compromise for better aroma preservation?. *Food Chem.* 139, 825–836. doi: 10.1016/j.foodchem.2013.01.038

Ruiz, J. J., Alonso, A., García-Martínez, S., Valero, M., Blasco, P., and Ruiz-Bevia, F. (2005). Quantitative analysis of flavour volatiles detects differences among closely related traditional cultivars of tomato: quantitative analysis of flavour volatiles in traditional tomato cultivars. *J. Sci. Food Agric.* 85, 54–60. doi: 10.1002/jsfa.1879

Selli, S., Kelebek, H., Ayseli, M. T., and Tokbas, H. (2014). Characterization of the most aroma-active compounds in cherry tomato by application of the aroma extract dilution analysis. *Food Chem.* 165, 540–546. doi: 10.1016/j.foodchem.2014.05.147

Serio, F., Leo, J. J., Parente, A., and Santamaria, P. (2007). Potassium nutrition increases the lycopene content of tomato fruit. *J. Hortic. Sci. Biotechnol.* 82, 941–945. doi: 10.1080/14620316.2007.11512330

Slimestad, R., and Verheul, M. J. (2005). Content of chalconaringenin and chlorogenic acid in cherry tomatoes is strongly reduced during postharvest ripening. *J. Agric. Food Chem.* 53, 7251–7256. doi: 10.1021/jf050737d

Spence, C. (2013). Multisensory flavour perception. *Curr. Biol.* 23, R365–R369. doi: 10.1016/j.cub.2013.01.028

Spence, C. (2015). Multisensory flavor perception. *Cell* 161, 24–35. doi: 10.1016/j.cell.2015.03.007

Thies, W. (1971). Schnelle und einfache Analysen der Fettsäurezusammensetzung in einzelnen Raps-Kotyledonen. I. Gaschromatographische und papierchromatographische methoden. *Z. Pflanzenzüchtung* 65, 181–202.

Tieman, D., Bliss, P., McIntyre, L. M., Blandon-Ubeda, A., Bies, D., Odabasi, A. Z., et al. (2012). The chemical interactions underlying tomato flavor preferences. *Curr. Biol.* 22, 1035–1039. doi: 10.1016/j.cub.2012.04.016

Tieman, D., Zhu, G., Resende, M. F., Lin, T., Nguyen, C., Bies, D., et al. (2017). A chemical genetic roadmap to improved tomato flavor. *Science* 355, 391–394. doi: 10.1126/science.aal1556

Verheul, M. J., Slimestad, R., and Tjøstheim, I. H. (2015). From producer to consumer: greenhouse tomato quality as affected by variety, maturity stage at harvest, transport conditions, and supermarket storage. *J. Agric. Food Chem.* 63, 5026–5034. doi: 10.1021/jf505450j

Vogel, J. T., Tieman, D. M., Sims, C. A., Odabasi, A. Z., Clark, D. G., and Klee, H. J. (2010). Carotenoid content impacts flavor acceptability in tomato (*Solanum lycopersicum*). *J. Sci. Food Agric.* 90, 2233–2240. doi: 10.1002/jsfa.4076

Xu, S., Li, J., Baldwin, E. A., Plotto, A., Rosskopf, E., Hong, J. C., et al. (2018). Electronic tongue discrimination of four tomato cultivars harvested at six maturities and exposed to blanching and refrigeration treatments. *Postharvest Biol. Technol.* 136, 42–49. doi: 10.1016/j.postharvbio.2017.10.004

Zanor, M. I., Rambla, J.-L., Chaïb, J., Steppa, A., Medina, A., Granell, A., et al. (2009). Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents. *J. Exp. Bot.* 60, 2139–2154. doi: 10.1093/jxb/erp086

Zhang, B., Tieman, D. M., Jiao, C., Xu, Y., Chen, K., Fe, Z., et al. (2016). Chilling-induced tomato flavor loss is associated with altered volatile synthesis and transient changes in DNA methylation. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12580–12585. doi: 10.1073/pnas.1613910113

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Quantification of Plant Root Species Composition in Peatlands Using FTIR Spectroscopy

Petra Straková[1,2]*, Tuula Larmola[1], Javier Andrés[3], Noora Ilola[4], Piia Launiainen[2],
Keith Edwards[5], Kari Minkkinen[2] and Raija Laiho[1]

[1] Natural Resources Institute Finland (LUKE), Helsinki, Finland, [2] Department of Forest Sciences, University of Helsinki,
Helsinki, Finland, [3] Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland, [4] Financial and Administrative
Services, Education Department, City of Vantaa, Vantaa, Finland, [5] Department of Ecosystem Biology, University of South
Bohemia, České Budějovice, Czechia

Evidence of plant root biomass and production in peatlands at the level of species
or plant functional type (PFT) is needed for defining ecosystem functioning and
predicting its future development. However, such data are limited due to methodological
difficulties and the toilsomeness of separating roots from peat. We developed Fourier
transform infrared (FTIR) spectroscopy based calibration models for quantifying the
mass proportions of several common peatland species, and alternatively, the PFTs
that these species represented, in composite root samples. We further tested whether
woody roots could be classified into diameter classes, and whether dead and living
roots could be separated. We aimed to solve whether general models applicable in
different studies can be developed, and what would be the best way to build such
models. FTIR spectra were measured from dried and powdered roots: both "pure
roots", original samples of 25 species collected in the field, and "root mixtures", artificial
composite samples prepared by mixing known amounts of pure roots of different
species. Partial least squares regression was used to build the calibration models.
The general applicability of the models was tested using roots collected in different
sites or times. Our main finding is that pure roots can replace complex mixtures as
calibration data. Using pure roots, we constructed generally applicable models for
quantification of roots of the main PFTs of northern peatlands. The models provided
accurate estimates even for far distant sites, with root mean square error (RMSE)
1.4–6.6% for graminoids, forbs and ferns. For shrubs and trees the estimates were
less accurate due to higher within-species heterogeneity, partly related to variation in
root diameter. Still, we obtained RMSE 3.9–10.8% for total woody roots, but up to
20.1% for different woody-root types. Species-level and dead-root models performed
well within the calibration dataset but provided unacceptable estimates for independent
samples, limiting their routine application in field conditions. Our PFT-level models can
be applied on roots separated from soil for biomass determination or from ingrowth
cores for estimating root production. We present possibilities for further development of
species-level or dead-root models using the pure-root approach.

**Keywords: FTIR, calibration model, dead roots, fine roots, peatland, plant root composition, plant functional type
(PFT), root chemistry**

INTRODUCTION

Root-mediated carbon (C) fluxes represent the major information gap when estimating C stocks and C transformations of any ecosystem that supports plant communities. Plants are the main drivers of the whole ecosystem productivity, and plant roots, "the hidden part" of the plant community, may comprise an equal or even greater part of the biomass or the annual biomass production compared to "the obvious part" aboveground (Persson, 1983; Jackson et al., 1996; Gower et al., 2001). Remains from roots and root-associated microorganisms then may form 50–70% of sequestered soil C (Clemmensen et al., 2013). Still, root systems in most plant communities are poorly understood.

In peatlands, the wet C hotspots of our planet (e.g., Page et al., 2011; Nichols and Peteet, 2019), root studies are especially rare (e.g., Iversen et al., 2018). Yet, root production is likely a major C flux in sites that are characterized by graminoids, such as sedge fens (Saarinen, 1996), and sites with a tree stand (Bhuiyan et al., 2017; Minkkinen et al., 2018), especially when there is an abundant shrub understory (Finér and Laine, 1998). Naturally forested sites are quite common in many peatland regions (Rydin and Jeglum, 2013). There are further nearly 150 000 km^2 of peatlands drained for forestry purposes, mostly in northern Europe (Paavilainen and Päivänen, 1995; Joosten and Clarke, 2002), where the vascular plant composition is shrub and tree dominated (Laine et al., 1995; Laiho et al., 2003).

Graminoids are adapted to produce large root biomass even in waterlogged conditions and very deep anoxic soil layers (Bernard et al., 1988; Saarinen, 1996; Proctor and He, 2019), compared to which shrubs or trees require drier soil conditions and are thus more shallow-rooting, but still producing significant root mass (Finér and Laine, 2000; Iversen et al., 2018). In addition to producing C inputs into the soil as root litter and root exudates, active plant roots may affect ecosystem functioning by, e.g., shaping the soil microbial community and its functions (e.g., Robroek et al., 2015). Such effects typically depend on plant species or plant functional type (PFT) (Straková et al., 2011; Peltoniemi et al., 2012; Robroek et al., 2015; Kaštovská et al., 2018), even though studies on specific root impacts are still rare. Changes in root biomass, rooting depth and/or species/PFT composition as a response to environmental or global changes may therefore strongly influence responses of the whole ecosystem. The lack of root-related data shows up as high uncertainty in ecosystem models predicting current and future functioning of peatlands (Frolking et al., 2010, 2011), and uncertainty in soil organic C monitoring in general (Lorenz and Lal, 2010; Jandl et al., 2014).

The shortage of knowledge on root-mediated C fluxes is mainly due to prevailing methodological difficulties. Root production may be related to the above-ground vegetation characteristics (Murphy et al., 2009a,b; Murphy and Moore, 2010), but unfortunately, general models do not exist and thus direct root measurements are still needed. Separating roots from soil and live roots from dead roots is extremely laborious; especially so when it comes to peat soils that consist of plant remains, including roots, at various stages of decay (Sjörs, 1991).

The same holds for species identification, which is typically carried out by hand-sorting and visual classification using morphological criteria and anatomical microscopic inspections (e.g., Bhuiyan et al., 2017). Such identification is time consuming, and subjective even for personnel with high level of expertise, and thus largely constrained in mixed stands or species-rich systems. Consequently, considerable effort has lately been invested to developing methods for root species identification and quantification, including DNA-based techniques (e.g., Mommer et al., 2011), pyrolysis (White et al., 2011) or plant wax markers (Dawson et al., 2000; Roumet et al., 2006). Each of these methods has its own limitations, commonly high cost, which largely reduces the number of samples that can be analyzed. Spectroscopy methods, near infrared (NIR) and recently Fourier transform infrared (FTIR), have been identified as the most promising (review by Rewald and Meinen, 2013, and references therein).

Near infrared and FTIR spectroscopy are non-destructive physical methods. The spectrum of light absorbed by a sample in the near-infrared (1100–2500 nm) or infrared (4000–400 cm^{-1}) region gives a chemical signature of the sample, providing information about the presence, character and abundance of chemical bonds or functional groups. Generally, infrared spectroscopy provides such advantages as the possibility to analyze large sample sets in relatively short time at low cost, with little sample preparation needed and no chemicals used. In addition, FTIR-ATR technique requires only small amounts of sample material, which is in many cases limited in root studies. Chemometrics then enable quantification of the required component (e.g., percentage of one plant species in the multispecies root mixture) using multivariate calibration of the spectroscopy data (Esbensen et al., 2002). The process of calibration is similar for both NIR and FTIR, with FTIR shown to be somewhat more precise (Bellon-Maurel and McBratney, 2011) and in principle enabling also direct interpretation of the spectral properties.

For root species identification or quantification, NIR or FTIR spectroscopy has already been applied on agricultural crop and weed species (Rumbaugh et al., 1988; Roumet et al., 2006; Picon-Cochard et al., 2009; Naumann et al., 2010; Kusumo et al., 2011; White et al., 2011; Meinen and Rauber, 2015; Legner et al., 2018; Streit et al., 2019) and forest tree species (Lei and Bauhus, 2010; Domisch et al., 2015; Tong et al., 2016; Finér et al., 2017), and to separate live roots from the dead for forage species (Picon-Cochard et al., 2009). The number of plant species contained in the root mixtures has ranged from two to five. These studies have shown the potential of NIR/FTIR to relate roots to plant species in specific studies using local root materials.

To our knowledge, NIR or FTIR spectroscopy has not been used by others to determine the root composition of peatland plant species. Neither have there been attempts to create identification methods for PFTs instead of individual species, even though the PFT-level might be sufficient and in some cases more applicable in studies covering a wide range of conditions and species. Also, in previous studies, it was not tested whether the models developed for their specific purposes could be applicable more generally for the species studied. Preparing

root mixtures for calibration data for each study site specifically is the most laborious part of utilizing the spectroscopy methods.

We aimed to develop FTIR-based calibration models for predicting the mass proportions of (i) several common peatland species, and alternatively, (ii) the PFTs that these species represented, in composite root samples. We further tested whether (iii) roots of woody plants could be classified into different diameter classes with such models, and (iv) the possibility to estimate proportions of dead and living roots with 4 plant species. Furthermore, we tested (v) whether pure roots (single-species and/or single diameter samples) could replace complex mixtures as calibration data. We strove for general applicability of the models for different peatland sites and samples that would represent, e.g., roots separated from soil samples for biomass determination or roots separated from ingrowth cores for estimating root production. We hypothesized that in root composite samples:

(H1): Roots of all peatland plant species or PFTs can be distinguished (principal component analysis, cluster analysis) and quantified (partial least squares regression calibration models) using FTIR, assuming that the between-species (interspecific) chemical variation captured by FTIR is higher than the within-species (intraspecific) heterogeneity.

(H2): Proportions of very fine (diameter ≤ 0.5 mm), fine (diameter < 2 mm) and coarser (diameter 2–10 mm) roots of shrubs and trees can be distinguished and quantified using FTIR, assuming that roots of the defined diameter classes have different FTIR signatures but the differences do not override between-species chemical variation.

(H3): Proportions of dead and living roots can be quantified using FTIR, assuming that dead and living roots have different FTIR signatures but the differences do not override between-species chemical variation.

MATERIALS AND METHODS

The Root Materials

We gathered an extensive set of plant roots for building FTIR based calibration models and their validation. The roots were altogether of 25 plant species and included both herbaceous (graminoids, forbs and ferns) and woody (shrubs, a broadleaf tree, coniferous trees) plants (**Table 1**). The samples were organized as several specific sample sets used for calibration and validation as outlined in **Table 1** and **Figure 1**. Since we aimed to create generally applicable calibration models, we included naturally occurring variation in root chemistry extensively in the calibration sample sets. To include within-site variation, roots of each species were collected from several plants at different locations of each site. To include between-site variation, roots were collected from several study sites for each species. Only the roots of graminoid *Deschampsia flexuosa* and forb *Epilobium angustifolium* in the *calibration sample set I*, which are not typical peatland species but still may appear in disturbed

drained peatlands, were collected at one study site only. The sites varied in soil type, water-level regime, nutrient regime, and climatic conditions. The roots were collected at different times of the growing season from May to October, in different years. Collection is described in the **Supplementary Material**.

Our initial focus was on fine roots that we defined as roots of diameter < 2 mm for all plant species. Coarser woody roots of diameter 2–10 mm were included into the datasets as substantial part of the belowground biomass of woody plants is formed by this fraction (e.g., Murphy et al., 2009b; Weishampel et al., 2009), and they are not covered by allometric equations used for estimating coarse root biomass of woody species (e.g., Laiho and Finér, 1996). However, when we started to apply the modified ingrowth core method (Laiho et al., 2014) for estimating root production, the results revealed that 92% of the ingrown roots in peatland forests within 3 year incubation were of diameter ≤ 1 mm (Bhuiyan et al., 2017), of which the majority was as thin as ≤ 0.5 mm. Thus a separate class of very fine roots of diameter ≤ 0.5 mm was furthermore added for woody plants.

The study included two types of dead roots, field-dead roots that were collected from living plants (*Carex rostrata*, *Eriophorum vaginatum*, *Pinus sylvestris* and *Vaccinium myrtillus*) in the field and separated from living roots using morphological criteria (color, structure and strength), as well as artificially dead roots of *Pinus sylvestris* and *Vaccinium myrtillus* produced in a root mortality treatment where root death was induced by desiccation. The treatment is described in the **Supplementary Material**. The aim of the artificial killing was to obtain dead, but still largely undecomposed roots.

Root Mixtures

Dried root samples were powdered with an oscillating ball-mill. Root mixtures containing known mass proportions of roots of different plant species (and diameter class for woody roots) were prepared by weighing and mixing pure root powders within the given sample set (**Table 1** and **Figure 1**), similarly as in, e.g., Lei and Bauhus (2010), Domisch et al. (2015), Meinen and Rauber (2015), Tong et al. (2016) and Finér et al. (2017). The proportion of each plant species in the mixtures ranged from 0 to 100%, on a dry mass basis. For calibration and external validation purposes root mixtures were prepared with 2–7 species components, the different plant species thus had from tens to hundreds occurrences in the mixtures (**Table 1**). For distant validation purposes root mixtures were prepared with 2–3 species components. Altogether we utilized about 1500 samples.

FTIR Measurements

Measurements were done on both the "pure roots," original samples of each species and diameter class available, and the "root mixtures" prepared. FTIR spectra of all samples, except for roots from the sites in Canada, Sweden and United Kingdom in the *distant validation sample set VII*, were obtained with a Bruker VERTEX 70 FTIR spectrometer (Bruker Optics, Germany) with a horizontal diamond ATR sampling accessory. Pulverized samples were placed directly on the diamond crystal (diameter 1.8 mm) and a MIRacle high-pressure digital clamp was used to achieve even distribution and contact of the sample and crystal. Each

TABLE 1 | Plant species and frequency of their occurrence in different subsets of the root samples, with number of pure root substrates in parentheses.

Species	Species abbreviation	Number of sites	Calibration sets:		External validation sets:				Distant validation sets:			
			I	II	III	IV	V	VI	VII	VII	VII	VII
			FI	FI	FI	FI	FI	FI	CA	CZ	SE	UK
			$n = 127\,(28)$	$n = 518\,(72)$	$n = 27\,(27)$	$n = 367\,(16)$	$n = 73\,(73)$	$n = 106$	$n = 57\,(21)$	$n = 90\,(90)$	$n = 44\,(8)$	$n = 73\,(39)$
Herbaceous plants:												
Graminoids:												
Carex lasiocarpa	CL	2	29 (4)		9 (9)			37				
Carex rostrata	CR	2	35 (4)		6 (6)			39				
Carex rostrata FD		2	8 (2)									
Deschampsia flexuosa	DF	1	23 (1)									
Eriophorum vaginatum	EV	8	40 (4)						22 (4)	5 (5)	23 (5)	17 (1)
Eriophorum vaginatum FD		5	11 (4)							30 (30)		
Trichophorum cespitosum	TC	2			4 (4)			38				
Forbs:												
Epilobium angustifolium	EA	1	23 (1)									
Menyanthes trifoliata	MT	2	28 (2)		4 (4)			31				
Rubus chamaemorus	RC	3	28 (4)									
Ferns:												
Dryopteris carthusiana	DC	2	25 (2)									
Equisetum fluviatile	EF	2			4 (4)			26				
Woody plants:												
Shrubs and broadleaf trees:												
Andromeda polifolia	AP	3					11 (11)	50			38 (2)	
Betula nana	BN	3		115 (6)		147 (2)	9 (9)	42				
Betula pubescens	BP	2		124 (4)		139 (2)	6 (6)					
Calluna vulgaris	CV	3		99 (3)		155 (2)	4 (4)					58 (20)
Chamaedaphne calyculata	CC	1							42 (6)			
Empetrum nigrum	EN	2		124 (3)		150 (2)						
Erica tetralix	ET	1										52 (18)
Rhododendron groenlandicum	RG	1							44 (8)			
Rhododendron tomentosum	RT	2		131 (6)		116 (2)	6 (6)					
Vaccinium myrtillus	VM	5		160 (10)		152 (2)	13 (13)			28 (28)		
Vaccinium myrtillus FD, AD		3		16 (2)						27 (27)		
Vaccinium oxycoccos	VO	3					10 (10)	44	3 (3)		45 (1)	
Vaccinium uliginosum	VU	3		122 (4)		147 (2)						
Vaccinium vitis-idaea	VV	3		87 (2)								
Coniferous trees:												
Picea abies	PA	5		159 (7)			8 (8)					
Pinus sylvestris	PS	7		233 (21)		150 (2)	6 (6)	32				
Pinus sylvestris FD, AD		4		62 (4)								

Roman numerals refer to the sample sets as in **Figure 1**. *Dead roots abbreviations: AD, artificially dead roots; FD, field-dead roots. Sample origins: CA, Canada; CZ, Czechia; SE, Sweden; UK, United Kingdom.*
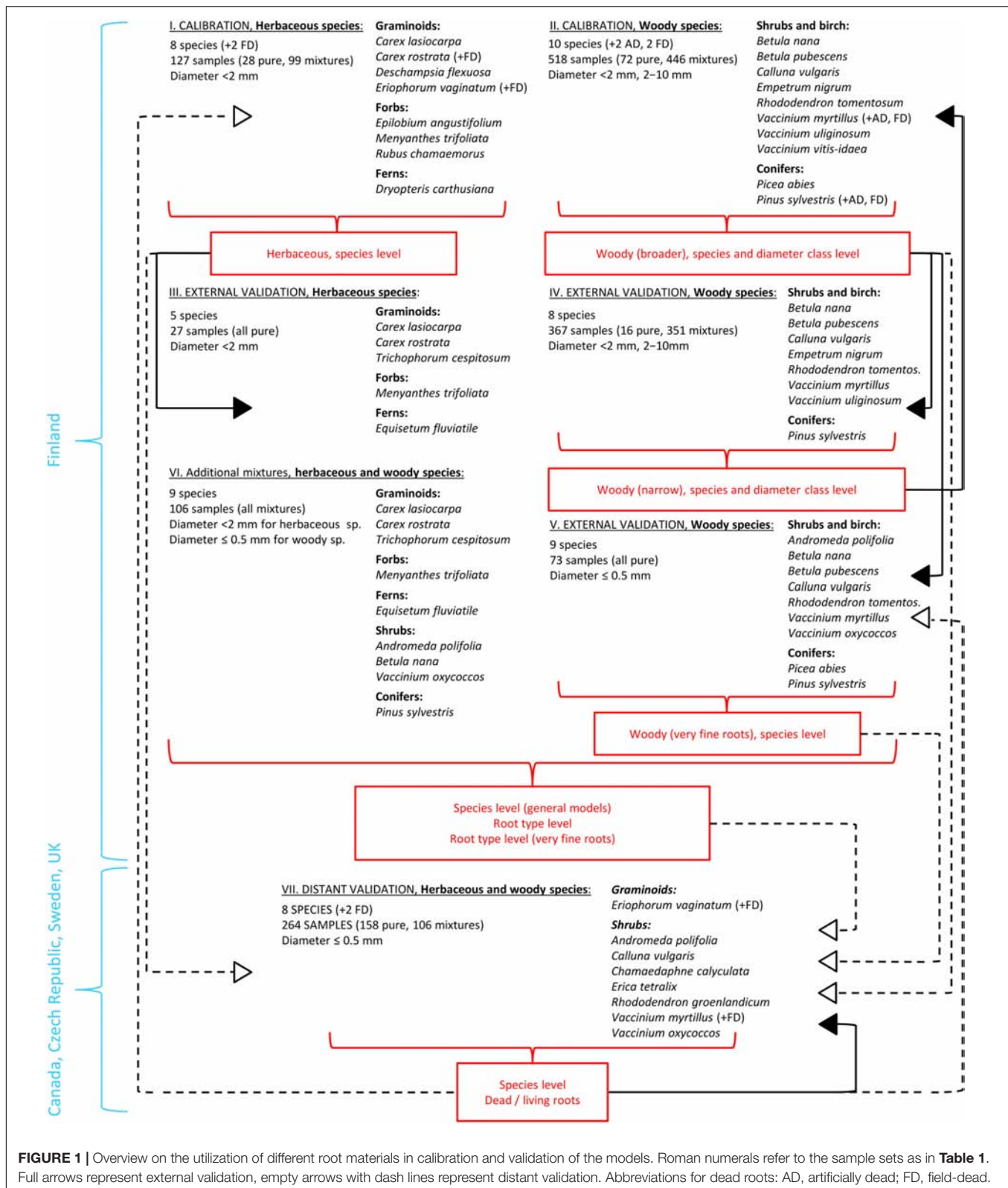
FIGURE 1 | Overview on the utilization of different root materials in calibration and validation of the models. Roman numerals refer to the sample sets as in **Table 1**. Full arrows represent external validation, empty arrows with dash lines represent distant validation. Abbreviations for dead roots: AD, artificially dead; FD, field-dead.

spectrum consisted of 65 averaged absorbance measurements between 4000 and 650 cm^{-1}, with 2 cm^{-1} resolution. Opus software was used to collect the measured data.

For the remaining samples, FTIR spectra were obtained with Shimadzu IRPrestige-21 FTIR spectrometer (Shimadzu Corporation, Japan) with a horizontal diamond ATR sampling

accessory and the same measurement settings as described above for the Bruker spectrometer. IRsolution software was used to collect the measured data. Compatibility of spectra from the two different instruments was ensured by measuring several standard samples with both instruments. The only adjustment needed before merging the data was interpolation of the FTIR spectra measured with the Shimadzu spectrometer to the exact wavenumber range of the Bruker spectrometer. This was done using the Unscrambler software.

Multivariate Data Analyses

FTIR data were smoothed (Savitzky-Golay smoothing with second polynomial order and 11 smoothing points), baseline corrected, mean normalized and transformed by second derivative (Savitzky-Golay derivative with second polynomial order and 15 smoothing points). This combination of pre-treatments was selected after testing different pre-treatments, including first derivative, standard normal variate (SNV), de-trending, multiplicative scatter correction (MSC), attenuated total reflectance (ATR) correction (Esbensen et al., 2002). The spectral parts 4000–2978, 2828–1752 and 772–650 cm^{-1} were excluded from further analyses due to lack of relevant information in these regions after the transformation. FTIR spectra used in the analyses thus consisted of transformed absorbance values at 2980-2830 and 1750-770 cm^{-1}, with 2 cm^{-1} resolution.

Cluster analysis, with Ward's algorithm and Euclidian distance, and principal component analysis (PCA) were used to compare the roots and define their grouping into chemically distinct root types using the FTIR data of pure (non-mixed) samples as the independent variables (X-matrix).

Partial least squares regression (PLSR) was used to build the calibration models, with known percentage of each plant species (and/or diameter class, root type, or dead root variant) in composite samples as the dependent variable (Y) and FTIR data as the independent variables (X-matrix). Three methods were used for the model validation: (1) internal leave-one-out cross validation (Esbensen et al., 2002), (2) external validation by using newly collected local (Finland) samples and (3) distant validation by using newly collected distant (Canada, Czechia, Sweden, United Kingdom) samples. Further notes on the validation are provided in **Supplementary Material**.

For both the calibration models and their validations, the values of the dependent variable predicted as less than 0% were set to 0% while those predicted as greater than 100% were set to 100%. Overall, this did not affect the results considerably. However, in cases where there were no roots of the predicted species or root type in the composite samples and the estimates were all negative (which is in fact a very good outcome), this resulted in root mean square error (RMSE) of the estimates equal to 0 indicating perfect fit of the model to the data, almost never achieved in practise. The r^2 values, slope and offset of the regression line, root mean square error (RMSE) of calibration and validation were used to evaluate the models. A good calibration model should have high r^2 value, slope close to 1, offset close to 0, low RMSE of both calibration and validation, and a relatively low number of factors (PCs) in order to avoid inclusion of

signal noise in the models. For evaluating the reliability of the predictions on new samples, sample deviation and inliner statistic (minimum Mahalanobis distance to the calibration samples) against Hotelling's T^2 statistic were used (Esbensen et al., 2002).

The Unscrambler 10.3 (Camo Process AS; Oslo, Norway) and Canoco 5 (Microcomputer Power, United States) software packages were used for the data analyses.

The Progress for Developing Different Sets of Models and Their Validation

Our data set allowed us to test and compare different ways of constructing the models. Unlike in previous studies, we were also able to validate the models on fully independent local (Finland; external validation) and distant (Canada, Czechia, Sweden, United Kingdom; distant validation) samples.

Grouping Roots of Different Species Into More Universal Root Types: PCA and Cluster Analysis (H1, H2)

We started by PCA and cluster analysis on pure root samples from Finland, the *calibration* and *external validation sample sets I–V* (**Table 1** and **Figure 1**), based on which we defined grouping of roots into different root types. The root types should provide more universal grouping of roots than species level, and generally represented different PFTs: graminoids, forbs, ferns, shrubs and trees, as well as different diameter classes for woody roots: very fine (diameter ≤ 0.5 mm), fine (diameter < 2 mm) and coarser (diameter 2–10 mm).

The PCA and cluster analysis were validated both externally (Finland) and distantly (Canada, Czechia, Sweden and United Kingdom). This was done by projection of the *distant validation sample set VII* into the PCA defined by the *calibration* and *external validation sample sets I–V*, or using them all together in a new cluster analysis. External validation was also done by re-constructing the PCA and cluster analysis for only subset of the Finnish samples (*calibration sample sets I, II*) and then using the other subset, collected in different years (*external validation sample sets III–V*) for the external validation (not shown).

External and distant validation of the outcomes ensured not only a correct grouping of roots of several species and/or diameters into more universal root types, but also tested the possibility to use PCA or cluster analysis as a simple tool for rough identification of newly collected root samples.

PLSR Calibration Models: Mixtures and Pure Root Approach

We continued by constructing PLSR calibration models for quantification of the mass proportions of the defined root types or individual species. We first used the approach of constructing the calibration models by creating artificial mixtures with varying proportions of the different plant species or root types in the mixtures, similarly to the earlier studies (Rumbaugh et al., 1988; Roumet et al., 2006; Picon-Cochard et al., 2009; Lei and Bauhus, 2010; Naumann et al., 2010; Kusumo et al., 2011; White et al., 2011; Domisch et al., 2015; Meinen and Rauber, 2015; Tong et al., 2016; Finér et al., 2017; Legner et al., 2018; Streit et al., 2019). Then we invented an alternative approach of constructing the

calibration models using only pure root substrates (single species and/or diameter class), thus skipping the need of creating the artificial mixtures and having only 0% and 100% values on the calibration curve.

PLSR calibration models at the level of root type

Calibration models at the level of root type were constructed using all available roots samples from Finland, the *calibration* and *external validation sample sets I–VI*. These models thus included woody roots of different diameter classes (≤ 0.5 mm, < 2 mm, 2–10 mm) and we made attempts to quantify their proportions for woody plants in composite samples. Another set of root type level models was constructed using only the *calibration* and *external validation sample sets I, III, V and VI*. These calibration models thus included only the very fine (diameter ≤ 0.5 mm) roots of woody plants, while keeping all the herbaceous plant roots.

Additionally, both of these model sets were also constructed using only the pure roots (no mixtures) of the specified sample sets.

PLSR calibration models at the level of plant species and diameter class (H1, H2)

Calibration models at the level of plant species were first constructed separately for herbaceous (*calibration sample set I*) and woody (*calibration* and *external validation sample sets II, IV, V*) plants. The separation of herbaceous and woody plants reduced the number of species included in the models, and allowed more detailed testing and investigation of the models when they were not affected by the presence of the other plant type. As the herbaceous and woody plants coexist in real peatland sites, however, they were in the end merged in the general species level calibration models (*calibration* and *external validation sample sets I–VI*). This allowed us to test if the model performance is negatively affected by the presence of both herbaceous and woody plants and the increased number of species in the models.

Next, we again constructed alternative calibration models using pure roots of selected herbaceous and woody plants. Three species (*Eriophorum vaginatum, Andromeda polifolia, Vaccinium oxycoccos*) of the *calibration* and *external validation sample sets I and V* were used. These models were compared with the mixtures models constructed for the same species using samples from the site in Sweden in the *distant validation sample set VII*. We further explored the possibility to apply this pure root approach on increased number of species. We used pure roots of 16 species (graminoids, forbs, ferns, shrubs and trees; *calibration* and *external validation sample sets I and IV*) and constructed pure root models for each of the species. Then we used artificial mixtures containing known mass proportions of roots of the given species (prepared from the pure roots used for the calibration) for the model validation, which allowed us to test whether there is any interference between the individual pure root components in the mixture spectra.

For woody plants, our data set allowed us to perform detailed testing of models that distinguish both species and diameter class. We started by "narrow" calibration models at the level of plant species as well as diameter class (diameter < 2 mm and 2–10 mm); narrow in the sense that only one study site and pooled samples for each species and diameter class were included in the calibrations (*external validation sample set IV*). Such narrow models were presented in most earlier root studies (e.g., Roumet et al., 2006; Picon-Cochard et al., 2009; Meinen and Rauber, 2015; Tong et al., 2016) and if such models worked outside the calibration sample set, they would offer a rather fast and simple way of root quantification using a limited number of samples.

Then, "broader" calibration models were prepared so that for each species roots of several different study sites collected at several sampling times were used. Thus, we increased the variation covered by the samples (*calibration sample set II*). Diameters < 2 mm and 2–10 mm were in this case pooled for each species, site and sampling time in known proportions.

Additionally, we constructed species level calibration models focusing on very fine roots (diameter ≤ 0.5 mm, *external validation sample set V*). Due to the very limited amount of available root material of the very fine diameter, root mixtures were not prepared and the models were constructed including only a limited number of pure root samples.

Dead roots (H3)

We first made attempts to construct calibration models for quantification of dead roots of four plant species in composite samples with other species. We used dead roots of *Carex rostrata, Eriophorum vaginatum, Pinus sylvestris* and *Vaccinium myrtillus* in mixtures with other plant species within the *calibration datasets I and II* (**Table 1** and **Figure 1**). This approach did not work well due to insufficient root material for creating robust calibration models and their validation (results not shown).

So we selected a different approach and explored the possibilities and limits of dead roots quantification using calibration models that quantified dead roots only within the species, not in composite samples with other species. The models were constructed for *Eriophorum vaginatum* and *Vaccinium myrtillus* using pure root samples from the site in Czechia in the *distant validation sample set VII*. The samples were collected at three different times of the growing season which allowed us to construct the calibration models on samples from one sampling time and validate them on samples from the same site but different sampling time (external validation). The models were then validated on living and dead roots of the given species from Finland (distant validation). For *Eriophorum vaginatum* the models were also validated on living roots from Canada, United Kingdom and Sweden (distant validation).

RESULTS

Defining Root Types Based on FTIR Spectra

Plant roots were separated into nine root types, or root "chemotypes", based on their FTIR spectra (**Figures 2–4**), roots of: (1) graminoids; (2) forbs; (3) ferns; (4) shrubs and birch: very fine roots (diameter ≤ 0.5 mm); (5) shrubs and birch:

FIGURE 2 | PCA analysis of FTIR spectra of plant roots. Ordination diagrams from principal component analysis (PCA) showing **(A)** variation in FTIR-derived chemistry of pure root samples from Finland (calibration and external validation datasets I–V; **Table 1** and **Figure 1**), with **(B)** centroids for the different plant species and diameter class, and arrows showing the main FTIR absorption bands that defined the Axis 1 and 2. Pure roots from Canada, Czechia, Sweden and United Kingdom (distant validation sample set VII) were **(C)** projected into the ordination space defined by the samples from Finland, with **(D)** centroids for the different plant species and diameter class. Species abbreviations are in **Table 1**. Symbols for the root types: graminoids, red circles; forbs, green crosses; ferns, brown diamonds; conifers diameter ≤ 0.5 mm, blue squares; conifers diameter < 2 mm, gray squares; conifers diameter 2–10 mm, open squares; shrubs and birch diameter ≤ 0.5 mm, black circles; shrubs and birch diameter < 2 mm, gray circles; shrubs and birch diameter 2–10 mm, open circles.

fine roots (diameter < 2 mm); (6) shrubs and birch: coarser roots (diameter 2–10 mm); (7) conifers: very fine roots (diameter ≤ 0.5 mm); (8) conifers: fine roots (diameter < 2 mm); (9) conifers: coarser roots (diameter 2–10 mm). Roots of graminoids with coarser roots of shrubs and birch differed from the other root types along the first PCA axis, which accounted for 43% of the variation in the root FTIR data (**Figure 2**). The first axis was largely defined by absorbance at 1033 cm^{-1}, assigned to polysaccharides (relatively higher absorbance in graminoids and coarser shrub and birch roots). Woody (shrub and tree) roots differed from the herbaceous (graminoid, forb and fern) roots along the second PCA axis, which accounted for 19% of the variation in the root FTIR data. The second axis was defined by absorbance at 1650 cm^{-1} and 1550 cm^{-1} assigned to polypeptides (amide I and II; relatively higher absorbance in the herbaceous roots), and 1606 cm^{-1} and 1450 cm^{-1} assigned to polyphenolics (relatively higher absorbance in shrub and tree roots). The FTIR patterns related to root type were consistent across different sample sets (**Figure 2**).

Roots of the herbaceous plants showed rather high chemical variation: forbs and ferns were not similar to graminoids but rather grouped with the roots of conifers and fine or very fine roots of shrubs and birch (**Figures 2**, **3**). Roots of forb *Rubus chamaemorus* were similar to roots of shrubs and birch and thus were added to shrubs and birch in the root type level calibration models. The other forbs and the ferns were more similar to conifers, but still formed distinct clusters (**Figure 3**).

Within woody species, our only broadleaf tree (birch *B. pubescens*) could not be distinguished from shrubs (**Figure 2**). Among shrubs, two species (*Andromeda polifolia* and *Vaccinium oxycoccos*) separated from others and grouped more closely with coniferous trees, forming however a distinct group in the cluster analysis (**Figure 3**). After testing different options we decided yet to add those to the other shrubs and birch in the root type level calibration models. Noteworthy, for all woody species, the within-species variation related to root diameter was in general higher than the difference from the other species (**Figure 5**). Conifers showed smaller

FIGURE 3 | Cluster analysis of the plant root FTIR spectra. Dendrogram of the second derivative of the root FTIR spectra, calculated with Ward's method and Squared Euclidean Distance, on pure root samples from Finland, the *calibration* and *external validation datasets I–V* (**Table 1** and **Figure 1**). For simplicity only the main clusters are shown, not the individual samples.
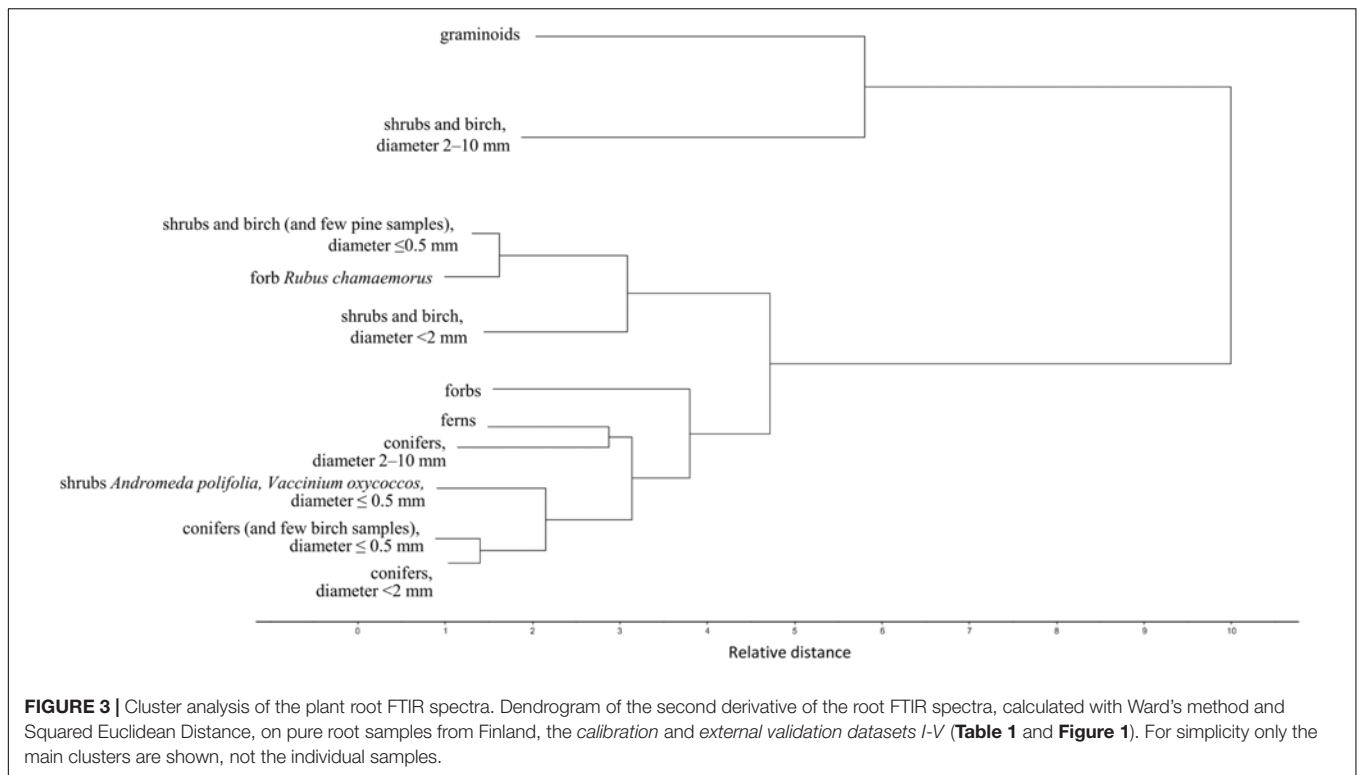
within-species variation related to the root diameter than shrubs and birch (**Figures 2**, **3**, **5**). In general, the very fine roots (diameter $\leq$ 0.5 mm) of all the woody species were rather similar. With increasing root diameter the difference between shrubs and birch on one hand and conifers on the other hand increased. Fine and very fine roots were characterized by higher absorbance at 1606 cm^{-1}, 1515 cm^{-1} and 1450 cm^{-1} cm^{-1} assigned to polyphenolics (lignin) as well as 2920 cm^{-1} and 2850 cm^{-1} assigned to aliphatic (wax, lipids) compounds. The coarser roots in turn were characterized by higher absorbance at 1033 cm^{-1} assigned to polysaccharides (**Figures 2**, **3**, **5**).

There were few trends in the variation in FTIR-derived root characteristics related to season, and these trends were marginal compared to the variation related to root type. The absorbance at wavenumbers assigned to polysaccharides tended to increase from early spring to summer and decrease in autumn, while absorbance assigned to polyphenolics had the opposite pattern (data not shown).

Estimating Mass Proportions at the Level of Root Type

Using root mixtures, calibration models for graminoids, forbs, ferns, and all woody roots (diameter < 10 mm) had RMSE of 5.7, 3.2, 2.7, and 6.1%, respectively, and r^2 0.96–0.97 (**Figure 6A**). Calibration models sorting woody roots to conifers and shrubs with birch also performed reasonably well, with RMSE 9.9–10.9% and r^2 0.92–0.93. Calibration models further sorting the woody roots based on their diameter (very fine $\leq$ 0.5 mm, fine < 2 mm,

coarser 2–10 mm) had RMSE of 4.4–9.2% for conifers and 9.5–14.5% for shrubs with birch, with r^2 0.83–0.96. Calibration models that included only pure root samples had RMSEs comparable to the mixtures models, often with even lower number of factors (PCs) used (**Figure 6B**). For all the models, the internal validation outcomes were in similar range as for the calibrations (**Figures 6A,B**).

Distant validation of the root type level models provided acceptable predictions even at far distance sites, with predictions being, similarly to the calibration models, generally somewhat better for herbaceous plant roots than for the woody plant roots (**Figure 7**, **Supplementary Figures S1–S3**). All woody roots present in the distant validation dataset were of diameter $\leq$ 0.5 mm and woody models for this diameter class generally provided more precise predictions than the woody models that contained all diameter classes < 10 mm. Noteworthy, root type level calibration models that included only pure root samples yielded comparable or better predictions than the models constructed using the mixtures. This finding was consistent across the different distant validation sites (**Figure 7**, **Supplementary Figures S1–S3**).

Using pure root models, the root type level predictions for sites with species that were included in the calibration models had RMSE of 0.0–14.1% at the Swedish site (**Figure 7B**), or 3.3–20.1% at the Czech site (**Supplementary Figure S1B**; if for woody roots only the models for very fine diameter were considered then RMSE ranged 3.3–7.9%). Predictions for roots from the sites in Canada and United Kingdom that also included shrub species not present in the calibration models had RMSE of 0.0–17.1% (**Supplementary Figures S2B, S3B**; if for woody roots only
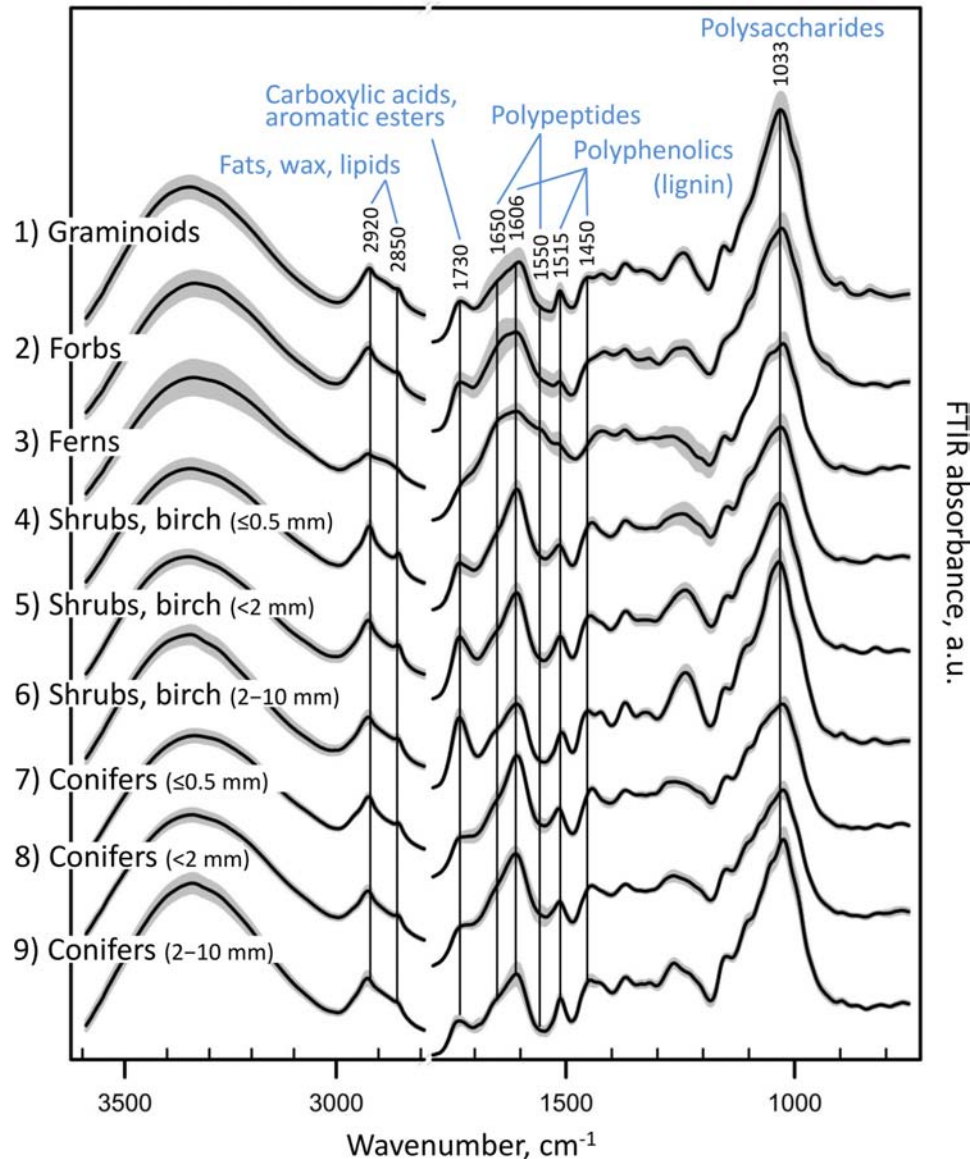
FIGURE 4 | FTIR spectra of the different root types. FTIR absorbance spectra of pure root samples from Finland (calibration and external validation datasets I–V; **Table 1** and **Figure 1**), grouped into nine root types: **(1)** graminoids; **(2)** forbs; **(3)** ferns; **(4)** shrubs and birch: very fine roots (diameter ≤ 0.5 mm); **(5)** shrubs and birch: fine roots (diameter < 2 mm); **(6)** shrubs and birch: coarser roots (diameter 2–10 mm); **(7)** conifers: very fine roots (diameter ≤ 0.5 mm); **(8)** conifers: fine roots (diameter < 2 mm); **(9)** conifers: coarser roots (diameter 2–10 mm). Black lines show means for each group, and gray area the standard deviation.

the models for very fine diameter were considered then RMSE ranged 1.4–15.8%).

Estimating Mass Proportions at the Level of Plant Species

Herbaceous Species: Graminoids, Forbs and Ferns

Calibration models for roots of 8 herbaceous species had RMSE ranging from 2.6% to 6.2% and r^2 from 0.97 to 0.99, with internal validation outcomes in similar range (**Supplementary Figure S4**). However, the calibration models did not yield reliable estimates of herbaceous species composition during

the external and distant validation (results not shown). For graminoids, the RMSEs ranged from 0–10% when models correctly estimated that the graminoid species roots were not forb or fern species, to 30–70% when the models did not distinguish the graminoid species *Carex lasiocarpa*, *Carex rostrata* or *Eriophorum vaginatum* from each other. For forbs, the only species present in the external validation dataset was *Menyanthes trifoliata* and its roots were successfully recognized by the calibration model for the species (RMSE 11.0%). There were two species present in the external validation dataset but not included in the calibration models, and those species were not estimated as "non-present" (prediction 0%) but rather partly

FIGURE 5 | PCA analysis of FTIR spectra of woody roots of different diameter class. Ordination diagrams from principal component analysis (PCA) showing **(A)** variation in FTIR-derived chemistry of pure root samples of woody roots, in relation to root diameter. The roots were collected at one study site in Finland (drained bog Kalevansuo) in one date. Diameter classes: ≤ 0.5 mm, < 1 mm, 1–2 mm, 2–5 mm, 5–10 mm. Shrub and broadleaf tree species are shown by circles while conifer tree species are shown by squares. Pure shrub roots from Canada, Czechia, Sweden and United Kingdom (distant validation sample set VII) and their rhizomes (1–2 mm) were **(B)** projected into the ordination space defined by the samples from Finland, only the centroids for the different plant species and diameter class are shown. Species abbreviations are in **Table 1**.

overlapped with similar species: *Trichophorum cespitosum* was estimated as being 38% *Carex lasiocarpa* or 44% *Eriophorum vaginatum,* while *Equisetum fluviatile* was estimated as 62% *Dryopteris carthusiana.*

Woody Species: Shrubs, Broadleaf Tree, Coniferous Trees

Narrow calibration models for shrubs and trees at the level of plant species and diameter class (diameter < 2 mm and 2–10 mm) fitted the calibration samples well with RMSE 2.4–6.5% and r^2 0.92–0.97, and internal validation outcomes in similar range (**Supplementary Figure S5**). The calibration models well distinguished and quantified even species of the same genus: *Betula nana* and *Betula pubescens*, *Vaccinium myrtillus* and *Vaccinium uliginosum*.

However, the narrow calibration models did not yield reliable estimates during the external validation on roots of same species and diameter, with RMSE 7.2–27.2% for species level estimates and 8.1–41.3% for diameter class level estimates, and r^2 0.02–0.85 (**Supplementary Figure S6**). When the models were applied on two species that were not included in the narrow calibration models, *Picea abies* and *Vaccinium vitis-idaea*, those species were not estimated as "non-present" but largely overlapped with the other woody species (data not shown).

Broader calibration models for woody roots at the level of plant species, that compared to the narrow models covered broader variation for each species, fitted the calibration samples with RMSE 4.5–8.6% and r^2 0.88–0.95, and internal validation outcomes in similar range (**Figure 8A**).

External validation of the broader models on roots of the same species and diameter provided better estimates than

did the narrow models, with RMSE 8.5–15.5% and r^2 0.64–0.89 (**Figure 8B**). For several species, most obviously for *Rhododendron tomentosum*, the estimates showed a trend of forming two clusters with different slopes in their regression lines. The clusters were related to the two diameter classes (diameter < 2 mm and 2–10 mm) with underestimated proportions of the fine diameter (< 2 mm) class (**Figure 8B**). Calibration models that merged roots of the same genus, *Betula nana* with *Betula pubescens*, *Vaccinium myrtillus* with *Vaccinium uliginosum* and *Vaccinium vitis-idaea,* or the two coniferous tree species *Pinus sylvestris* with *Picea abies,* provided marginal improvement of the external validation predictions compared to the species-level models (**Figure 8**).

External validation of the broader calibration models on very fine woody roots (diameter < 0.5 mm), provided reliable estimates only for *Calluna vulgaris* with RMSE 5.5%, for the other woody species RMSE was 12.5–47.5% (**Supplementary Figure S7**). Distant validation of the broader calibration models on very fine roots again provided acceptable estimates only for *Calluna vulgaris* with RMSE 14.3% (not shown).

Noteworthy, calibration models focusing on very fine woody roots that were constructed including only a limited number of pure root samples (no mixtures) fitted the root samples well with RMSE 0.9–6.2% for the species or grouped species, with internal validation outcomes in similar range (**Supplementary Figure S8A**). We had only limited number of samples of the given species and diameter class available for external and distant validation of these models. Still, the validation results indicate that these models provide better estimates for the very fine woody roots than the broader models above that included roots of different diameter classes < 10 mm. Distant
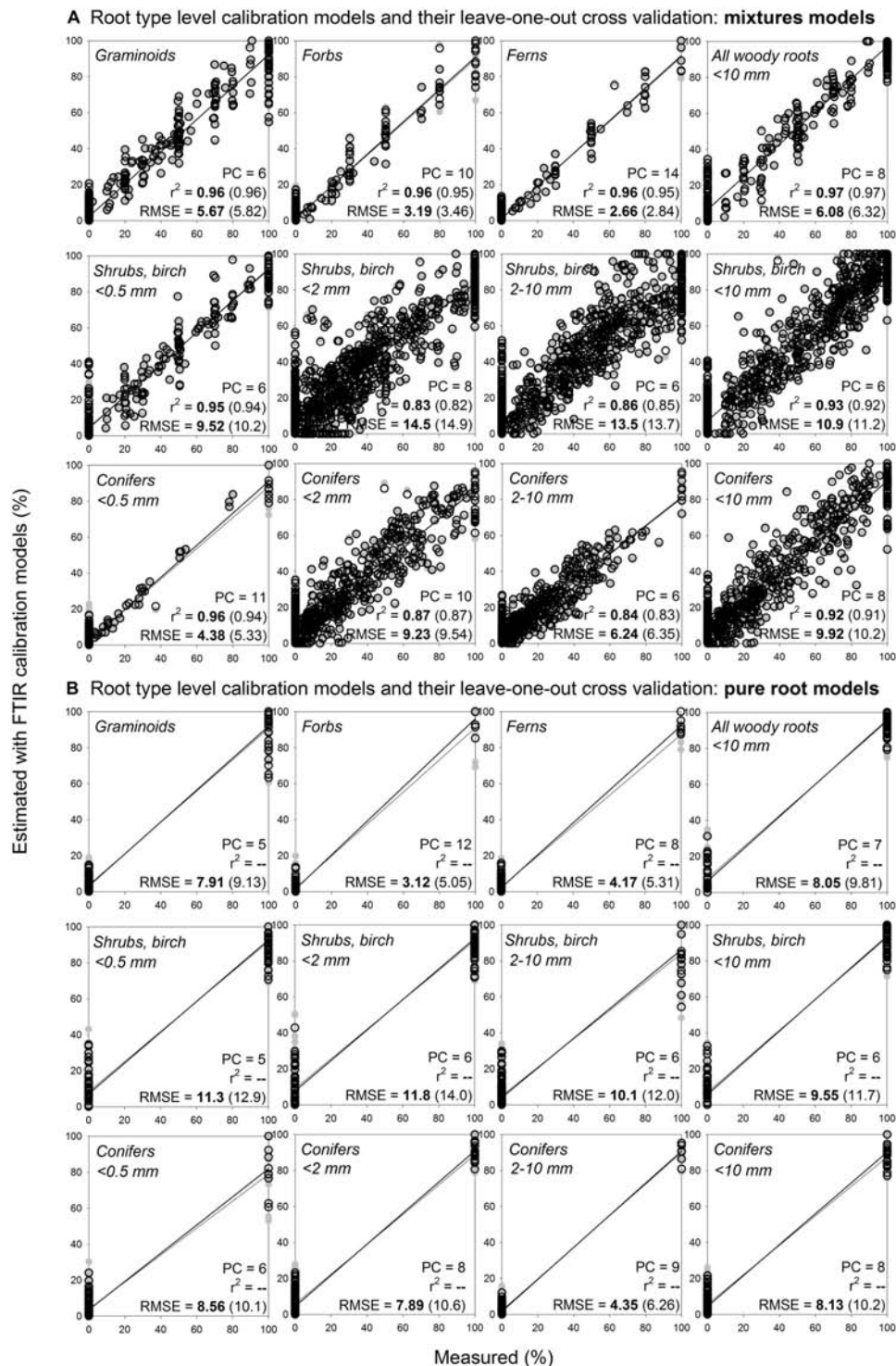
FIGURE 6 | Root type level calibration models and their internal validation: comparison of mixtures models and pure roots models. Relationships between the measured percentage of roots of the specific root type in composite root samples and the percentage estimated with FTIR calibration models. **(A)** For calibration of the mixtures models for shrubs and birch diameter ≤ 0.5 mm and conifers diameter ≤ 0.5 mm, the calibration and external validation sample sets I, III, V and VI (**Table 1** and **Figure 1**) were used, *n* = 333. For calibration of the mixtures models for all the other root types, samples from the calibration and external validation sample sets I–VI were used, *n* = 1218. **(B)** For calibration of the pure root models, only the pure roots (no mixtures) of the sample sets specified for **(A)** were used, *n* = 128 for woody roots diameter ≤ 0.5 while *n* = 216 for all the other root types. PC is the number of factors ("principal components") included in the calibration models and RMSE is the root mean square error. Calibration values are shown by black open symbols with RMSE and *r*[2] in bold letters, the internal full-cross validation values are shown by gray symbols with values of RMSE and *r*[2] in parentheses.
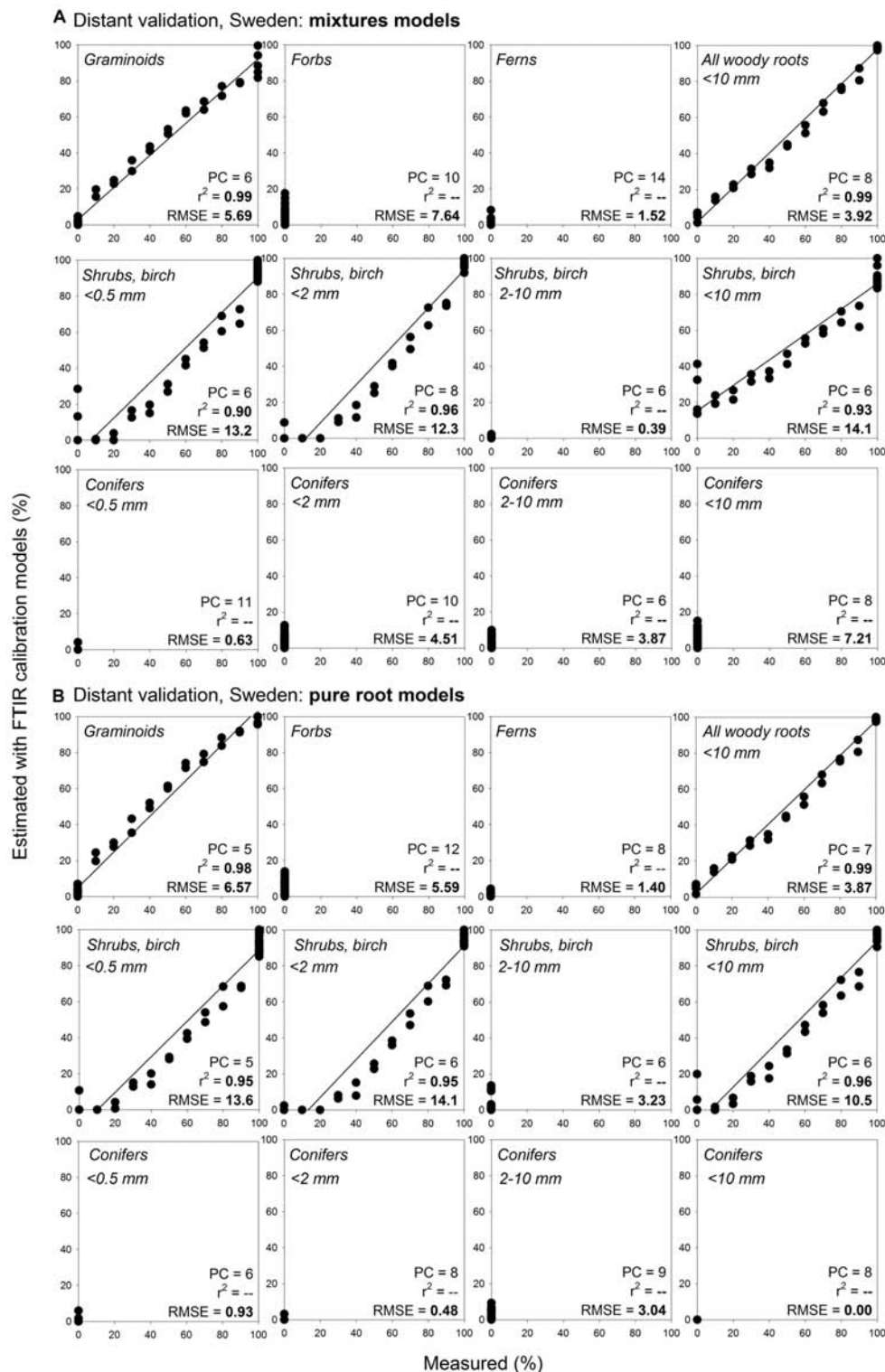
FIGURE 7 | Distant validation, Sweden, of the root type level calibration models: comparison of estimates using mixtures models and pure roots models. Relationships between the measured percentage of roots of the specific root type in composite root samples and the percentage estimated using FTIR calibration models: comparison of estimates using **(A)** mixtures models and **(B)** pure roots models. The calibration models are presented in **Figure 5**. Distant validation of the models show samples from wet fen site in Sweden (distant validation sample set VII, **Table 1**) that included plant species present in the calibration, $n = 44$. PC is the number of factors ("principal components") included in the calibration models and RMSE is the root mean square error. Distant validation of the pure root models on the remaining samples from the distant validation sample set VII is shown in **Supplementary Figures S1–S3, S13**.
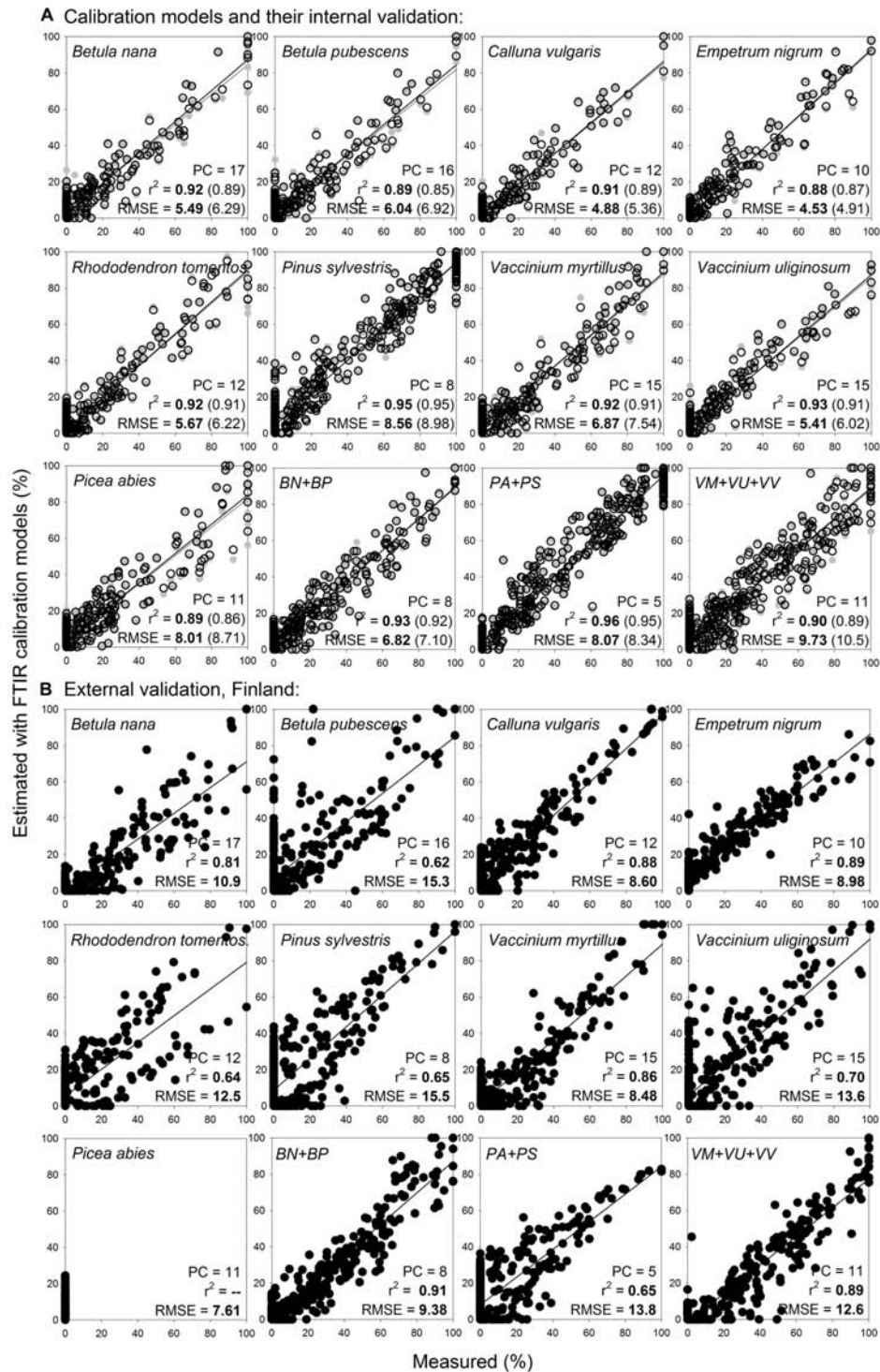
FIGURE 8 | Species level calibration models (broader) for woody roots, with external validation on newly collected samples. Relationships between the measured percentage of roots of the different woody plant species in composite root samples and the percentage estimated with FTIR calibration models. **(A)** For calibration, samples from calibration sample set II (**Table 1** and **Figure 1**) that included fine (< 2 mm) and coarser (2–10 mm) roots were used, n = 518. **(B)** External validation of the models show samples from external validation sample set IV, that included woody roots of the same species and diameter classes as present in the calibrations, n = 367. PC is the number of factors ("principal components") included in the calibration models and RMSE is the root mean square error. Calibration values are shown by black open symbols with RMSE and r^2 in bold letters, the internal full-cross validation values are shown by gray symbols with values of RMSE and r^2 in parentheses, and the external validation values are shown by black full symbols. Species abbreviations are in **Table 1**. External validation of the models on very fine (≤ 0.5 mm) woody roots of the same species is shown in **Supplementary Figure S7**.

validation of the models provided estimates with RMSE 4.3–23.2% (**Supplementary Figures S8B,C**).

Herbaceous and Woody Species Together in Calibration Models

Merging herbaceous and woody roots in general calibration models (**Supplementary Figure S9**) did not decrease the prediction ability of the models. Compared to the models constructed separately for herbaceous and woody species (**Figure 8** and **Supplementary Figure S4**), the calibration and internal validation RMSE of the general models were even lower as the two types of roots (woody vs. herbaceous) were well distinguished. External and distant validation of the general models provided similar estimates as did the herbaceous root models for herbaceous roots and the woody root models for woody roots (not shown). Similarly to the woody models, however, the general models calibrated on woody samples with different diameter roots (diameter < 10 mm) did not provide reliable estimates for the very fine woody roots (diameter ≤ 0.5 mm).

Noteworthy, pure root calibration models that were constructed for selected herbaceous and woody roots (3 species: *Eriophorum vaginatum*, *Andromeda polifolia*, *Vaccinium oxycoccos*; **Figure 9A**) provided very good estimates for the same species and diameter class during distant validation, with RMSE 5.9–8.4% (**Figure 9B**). Prediction abilities of these three-species "pure root" models were comparable with the predictions using "mixtures" models (**Figures 9C,D**). Furthermore, when we applied this approach on 16 species of herbaceous and woody plants, the pure root models (**Figure 10A**) provided very good estimates of proportions of the given species in composite samples (prepared from the pure roots used for the calibration), with RMSE 3.0–14.9% (**Figure 10B**, **Supplementary Figure S10**).

Dead Roots

Compared to the living roots of the given species, the field-dead roots showed a pattern of relatively higher absorbance at FTIR regions assigned to polyphenolics (1606 cm^{-1}, 1515 cm^{-1}, 1450 cm^{-1}) and polypeptides (1650 cm^{-1} and 1550 cm^{-1}). In contrast, lower absorbance at FTIR regions assigned to polysaccharides (1033 cm^{-1}), aliphatics: fats, wax, lipids (2920 cm^{-1} and 2850 cm^{-1}) and carboxylic acids or aromatic esters (1730 cm^{-1}) was found for field-dead roots.

There was only negligible change in FTIR-derived chemistry from living to artificially dead roots of *Vaccinium myrtillus* and *Pinus sylvestris* from the greenhouse experiment. Still, similarly to the field-dead roots, there was a trend of relatively higher absorbance at FTIR regions assigned to polyphenolics and lower absorbance at FTIR regions assigned to polysaccharides in the artificially dead roots (**Supplementary Figure S11**).

Within species, the calibration models constructed on pure root samples provided very good estimates of field-dead roots of *Vaccinium myrtillus,* with external validation RMSE 2.7% and 8.4% (**Supplementary Figure S12**). The artificially dead roots did not represent well the field-dead roots and were estimated as 100% living during the distant validation (RMSE 100%), while the living roots were estimated correctly, with RMSE 4.5%

(not shown). For *Eriophorum vaginatum*, the calibration models estimated field-dead roots correctly with external validation RMSE 6.7% and 13.2% (**Supplementary Figure S12**). Distant validation provided correct estimates of field-dead and/or living *Eriophorum vaginatum* roots for four of seven tested study sites, with RMSE < 10%, while for the three remaining sites the estimates were unacceptable with RMSE 30–40% (not shown).

The root type level models showed somewhat different predictions for living (**Supplementary Figure S1**) and field-dead (**Supplementary Figure S12**) roots of *Vaccinium myrtillus* during distant validation at the Czech site, the field-dead roots were estimated like conifers rather than shrubs.

DISCUSSION

Quantification at the Level of Plant Functional Type or Root Type Was Possible Even for Distant Samples; Simple Identification of Graminoids

We were able to construct generally applicable FTIR calibration models for quantification of roots of the main PFTs of northern peatlands: graminoids, forbs, ferns, shrubs (including a broadleaf tree), and coniferous trees. It was also possible to distinguish diameter classes for roots of woody plants. The models estimated mass proportions of these root types in composite samples with relatively low error, even for roots from far distance sites that were of different plant species than those included in the calibration models. These results provide robust support to the hypotheses that roots of peatland PFTs (H1) and diameter classes (H2) can be distinguished and quantified using FTIR spectra. Such root-type level models have to our knowledge not been built before, and could be more widely applicable than species-level models.

Graminoid roots clearly differed from all the other root types. This enabled their reliable quantification in composite samples using the calibration models. They could also be identified using PCA or cluster analysis, in our extensive dataset with 100% success. Forbs and ferns belong to herbaceous plants, but their root FTIR-derived chemistry was quite different from graminoids. Accordingly, Legner et al. (2018) showed clear separation of four graminoid species from a dicot species using cluster analysis of their root FTIR spectra. Forbs and ferns showed similarity with the roots of coniferous trees and fine or very fine roots of shrubs and birch. Still they formed separate clusters, which is a precondition for correct quantification by the calibration models. Yet, the distant validation for non-presence of forbs and ferns in root samples indicated that there may be some overlap with trees and shrubs in the estimates. In reality, the accuracy of the estimates can be improved with information on plant species or PFT presence at target sites, which is easily based on aboveground observations. Thus, if there is no evidence of, e.g., fern presence, fern models need not be applied at all.

In earlier studies all herbaceous roots, and sometimes all ground vegetation roots have usually been pooled, irrespective of whether NIR spectra (Lei and Bauhus, 2010; Domisch et al., 2015; Finér et al., 2017) or visual identification
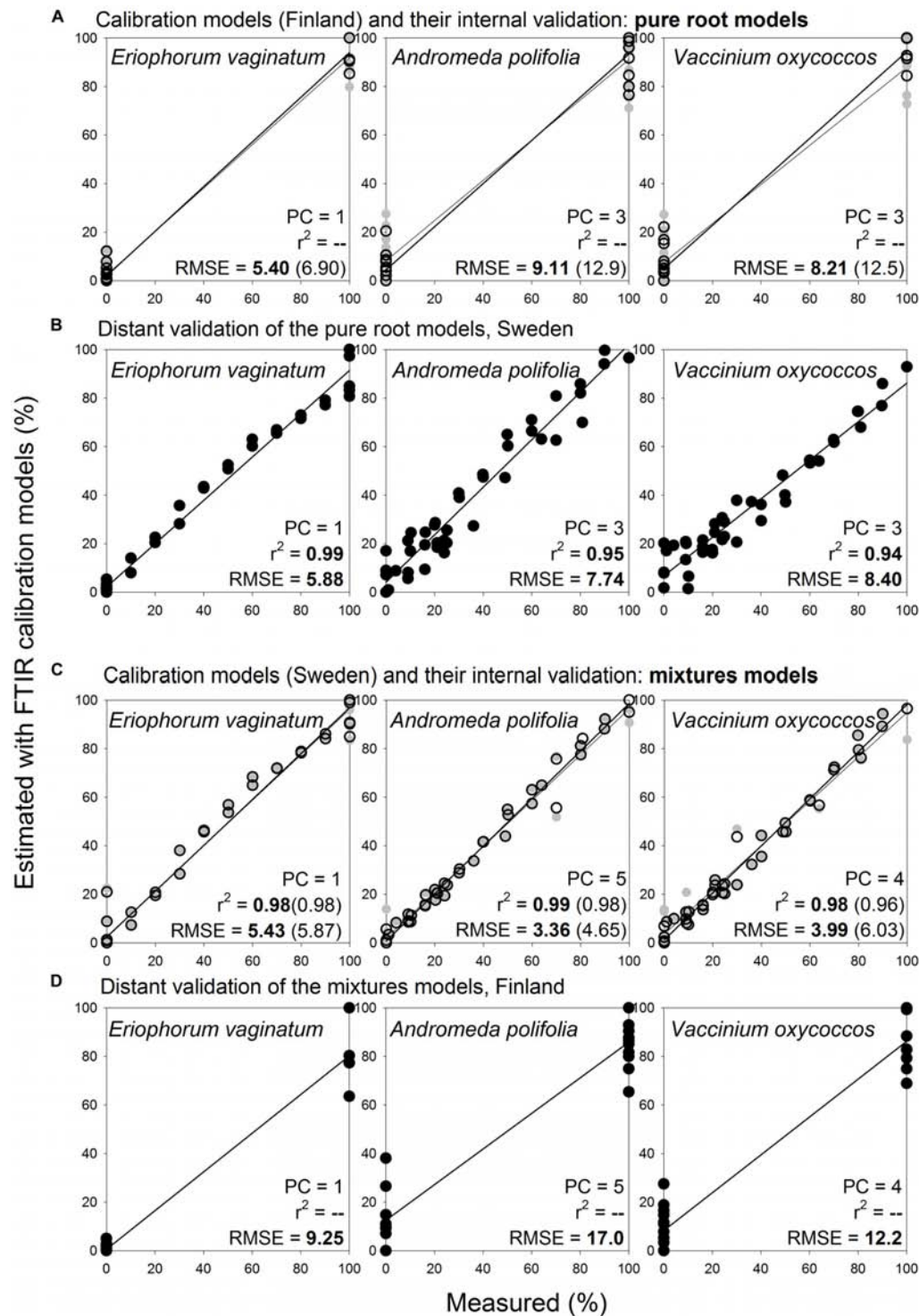
FIGURE 9 | Species level calibration models, comparison of results obtained with pure root and mixtures models for three species, distant validation between Finnish and Swedish sites. Relationships between the measured percentage of roots of the different plant species in composite root samples and the percentage estimated with FTIR calibration models. **(A)** For calibration, pure roots of the selected herbaceous and woody roots of the calibration and external validation sample sets I and V (**Table 1**) were used, $n = 25$. **(B)** Distant validation of the models show samples (both pure roots and their mixtures) of the same species as used in the calibrations, from the site in Sweden in the distant validation sample set VII, $n = 44$. **(C)** Calibration models were constructed using the same samples described for **(B)**. **(D)** Distant validation of the models was done using the same samples described for **(A)**. PC is the number of factors ("principal components") included in the calibration models and RMSE is the root mean square error. Calibration values are shown by black open symbols with RMSE and r^2 in bold letters, the internal full-cross validation values are shown by gray symbols with values of RMSE and r^2 in parentheses, and the distant validation values are shown by black full symbols.
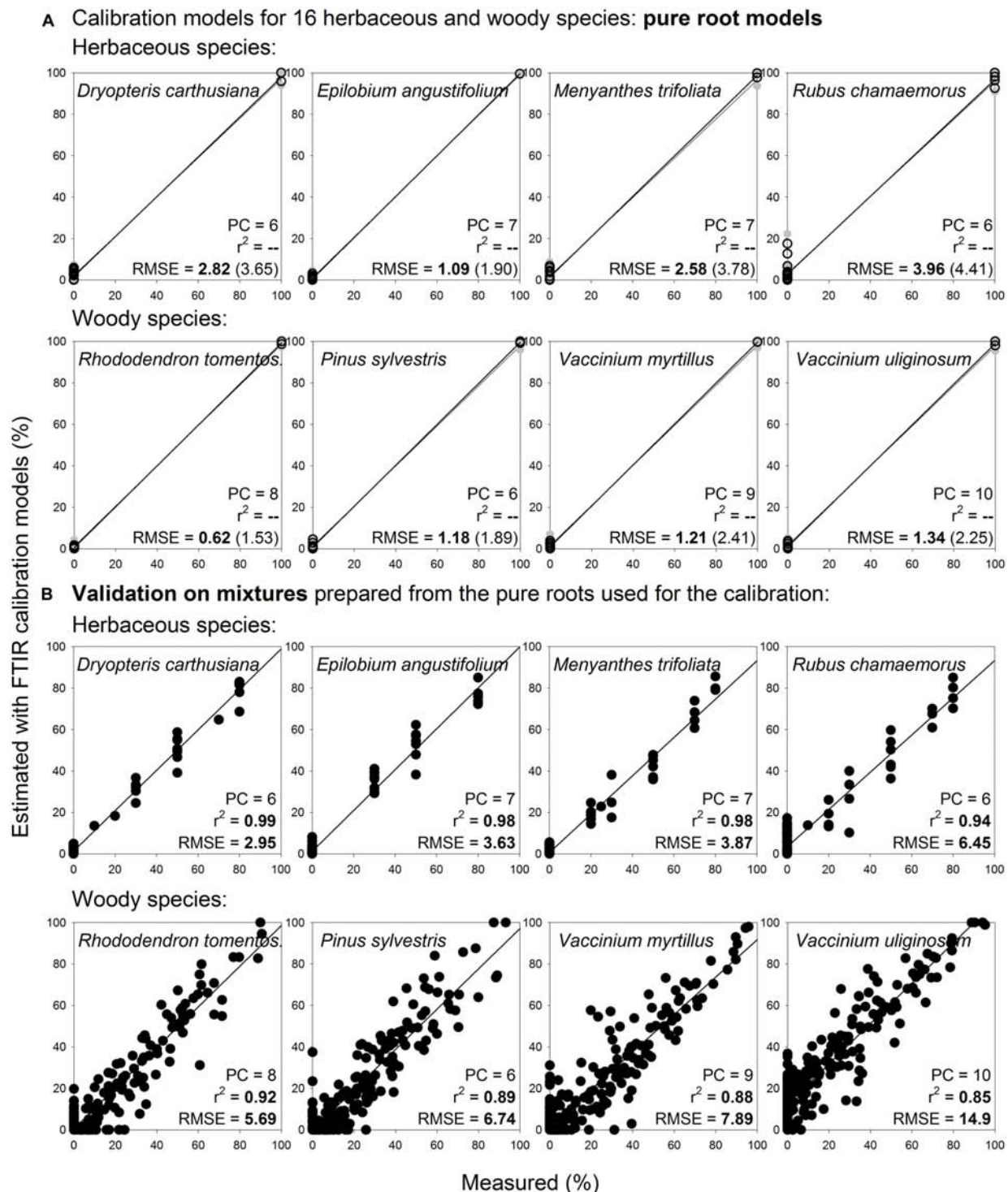
FIGURE 10 | Species level calibration models constructed on pure roots of 16 species and their validation on multispecies mixtures prepared from the pure roots used for the calibration. Relationships between the measured percentage of roots of the different plant species in composite root samples and the percentage estimated with FTIR calibration models. **(A)** For calibration, pure roots of 16 herbaceous and woody species of the calibration and external validation sample sets I and IV (**Table 1**) were used, only living roots were included, $n = 38$. **(B)** Validation of the models show artificial mixtures containing known mass proportions of roots of the given species, that were prepared from the pure roots used for the calibration, $n = 466$. PC is the number of factors ("principal components") included in the calibration models and RMSE is the root mean square error. Calibration values are shown by black open symbols with RMSE and r^2 in bold letters, the internal full-cross validation values are shown by gray symbols with values of RMSE and r^2 in parentheses, and the distant validation values are shown by black full symbols. Only graphs for 8 species are shown, graphs for the remaining 8 species are shown in **Supplementary Figure S10**.

(Bhuiyan et al., 2017) was applied. Our plant functional type or root type level calibration models offer a possibility for more detailed identification of understory vegetation roots. This is important since it will make it possible to produce data for ecosystem models to represent PFTs correctly also belowground. This will allow, for instance, accounting for the different turnover rates (Gill and Jackson, 2000) and decomposability (Straková et al., 2010, 2012) of the PFTs. In peatlands, the high water-table levels and soil anoxia largely shape the root depth distributions of plant species and PFTs (Murphy et al., 2009a,b; Murphy and Moore, 2010). Thus, these distributions may greatly differ from patterns typical in oxic mineral soils (e.g., Westoby and Wright, 2006). Consequently, estimation of the rooting patterns of PFTs cannot be based on insights from mineral soils.

Diameter Class Affected Estimates of Conifers, Shrubs and Birch, but Not Woody Roots in Total

The diameter-related results on woody roots partly disagree with our initial hypothesis stating that roots of different diameter class can be distinguished and quantified using FTIR spectra (H2). Fine and coarser roots showed different FTIR signatures, as we expected. However, the diameter class differences overrode the between-species chemical variation and negatively affected estimates at the level of species or PFT (coniferous tree vs. shrubs with birch). Still, it was well possible to quantify woody roots in total. Our results demonstrated clear diameter class differences in woody roots, and ability of the models to quantify the different classes in artificial mixtures. However, we must conclude that in practise, we are not able to quantify their proportions in field samples, where there are no clearly defined classes (only) present, but a continuum of roots of varying diameters. The practical conclusion is thus that when applying the models for identifying roots in field samples we should focus on very fine or fine roots only. Roots coarser than 2 mm should be manually removed, if present, and analyzed separately. The good thing is that coarse roots are rather easily separated, unlike fine roots. The unfortunate thing is that the compositional differences between the roots of conifers and shrubs with birch decrease with decreasing root diameter and consequently, also the ability of calibration models to correctly distinguish and quantify these root types decreases.

The diameter-related chemical variation that we present for roots is supported by earlier findings that show increase in polyphenolic (lignin) and nutrient concentration and decrease in polysaccharide (cellulose) concentration with decreased root diameter (Pregitzer et al., 2002; Thomas et al., 2014; Zhang et al., 2014). They are also in line with the woody plant chemistry aboveground: similar differences were observed for branch litter, with finer branches having higher concentration of lignin and nutrients and lower holocellulose concentration compared to coarser branches (Vávřová et al., 2009; Straková et al., 2010).

Compared to branches or coarser roots, the very fine root chemistry and consequently, FTIR signature is, however, additionally affected by mycorrhizal colonization (Pena et al., 2014), which may also complicate the root species identification.

While there is little information from peatlands, in general both birch and pine species have shown 100% mycorrhizal colonization rate and no secondary xylem or continuous cork layer for the finest (branching order 1) roots (Guo et al., 2008), which may make the root FTIR signatures rather similar between species. In the same study birch then showed decrease in mycorrhizal colonization and increase in secondary xylem and continuous cork layer from branching order 2 and pine from branching order 3, providing support for our finding that species-specific compositional patterns become more evident with increasing root thickness. Other studies also reported some overlap of NIR-derived chemistry of coniferous and broadleaf tree fine roots that was reflected in prediction abilities of the calibration models (Lei and Bauhus, 2010; Domisch et al., 2015; Finér et al., 2017).

Concerning herbaceous plants, we did not analyse diameter effect as the actual diameters were not determined for all samples. The roots were all of diameter < 2 mm, thus belonging in the fine root class. This class, however, included also very fine roots, if they were formed by the given species. In another study, species-dependent diameter class differences were observed for graminoid and forb forage species (White et al., 2011). Two of their four species had largely the same composition regardless of diameter, while the other two showed different composition for roots of diameter < 1mm and coarser than 1mm. Finer roots showed higher absorbance at wavenumbers assigned to polyphenolics relative to polysaccharides, in line with our results for woody roots, but higher absorbance at wavenumbers assigned to carbonyls, which is in contrast with our results for woody roots that showed higher absorbance at this region in the coarser diameter class only. The finest part of the roots with yet undifferentiated cells, the root tips, had to be removed in Meinen and Rauber (2015) and Legner et al. (2018) to get full species differentiation of various segments of forage roots by cluster analysis of their FTIR spectra. Thus, we may conclude that diameter variation possibly affected the model fit to graminoids.

Quantifications at Species Level Are Not Routinely Applicable in Field Conditions

The results on species level quantification do not fully support our hypothesis that roots of all peatland plant species can be distinguished and quantified using FTIR (H1). We were not able to clearly distinguish and quantify species belonging to the same PFT. The within-species (intraspecific) heterogeneity captured by FTIR was higher for the species sampled at different sites and/or sampling times than the between-species (interspecific) chemical variation, limiting routine application of the species-level models in field conditions.

This outcome seems to be in contrast with a study that examined FTIR spectra of five herbaceous agriculture plants and concluded that the roots of the same species are similar despite differences in climate, soil and fertilization, while important differences were noted between roots of different species (White et al., 2011). Several other FTIR or NIR studies that quantified roots of closely related species in mixtures also ended in more optimistic conclusions concerning the predictive power than we

(Rumbaugh et al., 1988; Roumet et al., 2006; Picon-Cochard et al., 2009; Naumann et al., 2010; Kusumo et al., 2011; White et al., 2011; Meinen and Rauber, 2015; Legner et al., 2018; Streit et al., 2019; Lei and Bauhus, 2010; Domisch et al., 2015; Tong et al., 2016; Finér et al., 2017; Picon-Cochard et al., 2009). However, to our knowledge, none of the earlier studies tested their models on newly collected independent samples. Also our models and their internal cross-validation results look very promising as such. However, when applied on newly collected samples of the same species coming from various sites, the models did not reliably distinguish and quantify the species. We thus have to conclude that despite of our effort to include high natural variation in our samples, the calibration models cannot be successfully applied on new samples outside the calibration dataset. We suggest that the previous studies using the same methodology were overoptimistic, and the models should not be used for routine application without careful external validation. They may, naturally, still be valid for the specific study settings that they were created for.

Dead Roots Differ From Living, but the Calibration Models Are Not Routinely Applicable in Field Conditions

We found FTIR-derived differences between living and dead roots and demonstrated the potential to quantify the dead and living roots within the given species. These results provide robust support to the hypotheses stating that proportions of dead and living roots can be quantified using FTIR, assuming that dead and living roots have different FTIR signatures (H3). However, models estimating dead root proportions only within a given species are insufficient for routine application in field conditions when multiple species are present. Theoretically, if combined with pure roots of other species, the models could be reconstructed to quantify proportions of dead roots in multiple species composite samples, and this will be the direction of our further work.

The need to include also dead roots in the calibration models in field studies has been concluded earlier (Lei and Bauhus, 2010). Still, so far only Picon-Cochard et al. (2009) attempted to quantify the proportions of dead and living roots in composite samples. They constructed models for separation of artificially produced, 1 and 2 months old dead roots, in mixtures of 5 graminoid species, with RMSE of one-leave out full cross validation 15%. Their mixtures were, however, prepared from a bulk sample for each species and living or dead root variant and thus are unlikely to give reliable estimates outside their experimental conditions.

Dead roots in field conditions may vary greatly in time passed since the root death, from recently dead to dead for years, in anoxic peat soils actually even millennia. This, in combination with the root initial chemistry and the environmental characteristics, determines their decomposition stage and thus their FTIR signatures (e.g., Duboc et al., 2012). Unfortunately, our experimental production did not yield suitable samples for the calibrations, the artificially dead roots did not represent well the field-dead roots. This means that one of the major challenges in creating dead root models is being able to

harvest representative materials. Yet, with the pure root approach this is made at least a bit easier.

Pure Root Models Are More Practical Than the Traditional Models Constructed on Root Mixtures

Our results confirm the extreme power of FTIR with chemometrics to distinguish and quantify different substrates mixed into a composite sample. We present distinction and quantification of even 16 substrates. This is, as such, nothing new. Noteworthy, we demonstrate that the substrates can be successfully quantified in composite samples using calibration models that are constructed on pure substrates only, without the need to prepare artificial mixtures of the different substrates for the model calibration. This finding has several practical implications for studies dealing with fine or very fine plant roots:

1) Significantly reduced amount of root material that needs to be collected for model calibration.
2) Consequently, possibility to cover higher within-species (intraspecific) heterogeneity, without the need to increase sample amount by pooling roots for each species.
3) Flexibility to add/remove particular sample type (e.g., species, or diameter class) from calibrations.
4) Reducing technical errors introduced by weighing and mixing the samples for mixtures

For example, 45 g of dry roots were used for creating narrow 5-species NIR mixtures models by Tong et al. (2016). A minimum of 500 mg dry root material was necessary to record the NIR spectrum of one sample (Tong et al., 2016; Lei and Bauhus, 2010). Such amount is quite hard to harvest in the very fine root diameter. Using FTIR-ATR method instead of NIR already significantly reduces the amount of roots needed to about 5–10 mg of dry root material per sample. Combined with our pure root approach, 5-species model covering sufficient within-species variation (10 samples per species) could be achieved with only 250–500 mg dry root material (5 species × 10 samples per species × 5–10 mg per sample).

Natural variation makes the difference between our attempts to quantify roots grown in field conditions, and roots produced in controlled greenhouse experiment, or even more clearly, between strictly defined chemicals. We need to capture the natural variation into the calibration sample library to make reliable predictions for new samples. Natural variation may be captured by well designed sampling, following objectives of the particular study, and sufficient number of samples per each species. However, most of the published root studies (including ours, to a large extent) needed to pool the collected roots to obtain enough material for creating the artificial mixtures. The only study that clearly stated that their mixtures were prepared from different individuals, not pooled samples, is the study by Lei and Bauhus (2010). The pure root approach allows us to cover much more within-species variation for more robust calibration models.

Previous studies recommended to increase the prediction quality of the calibration models by extending the calibration

sample size by creating more artificial mixtures (e.g., Lei and Bauhus, 2010; Meinen and Rauber, 2015), giving the impression that the more mixtures, the better.

We had two assumptions based on earlier work that we took for granted when beginning our research:

– For successful quantification of species in composite samples, we need to achieve all possible combinations of the species in artificial mixtures for calibration,
– The more mixtures the better.

Now we argue that both of these were in fact wrong. We do agree that increasing the calibration sample size improves the models. However, we argue that while mixtures improve the fit of the calibration models, they do not improve the essential, which is the prediction ability of the models. Mixtures are basically just re-runs of samples that were used for the mixtures preparation. Instead of creating the artificial mixtures with all possible combinations of the species, we now suggest a different way of building models for routine applications in field conditions. That is creating an extensive root spectra database covering the species (that may be used to represent PFTs) and root types with several independent samples, and then selecting suitable samples from the database for each specific research goal using the pure root approach. This approach leaves the flexibility for selecting the combination of species or sample types best suited for each application, instead of being stuck with mixtures that include also irrelevant material that may distort the analyses. Seasonal variation in root chemistry (Lei and Bauhus, 2010; Kaštovská et al., 2018) should also be considered; either by including it extensively in the calibration sample library, or sampling at the same season as the forthcoming samples to be estimated by the models are to be collected.

CONCLUSION

Our results confirm the extreme power of FTIR with chemometrics to distinguish and quantify different substrates mixed into composite samples. Noteworthy, we demonstrate that the substrates can be successfully quantified in composite samples using calibration models that are constructed on pure substrates only, without the need to prepare artificial mixtures with varying concentrations of the different substrates for calibrations of the models.

We were able to construct generally applicable FTIR calibration models for quantification of roots of the main PFTs of northern peatlands: graminoids, forbs, ferns, shrubs (including a broadleaf tree), and coniferous trees.

More detailed root quantifications, e.g., at species level or distinguishing dead roots from living, are in field conditions restricted by natural variation in root chemistry, and unclear boundaries between the different root classes. We were not yet able to construct such models for general application in field conditions, although we present a possibility for further development of such models using the pure root approach.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

RL brought the initial idea for this study. PS, TL, KE, KM, and RL planned and designed the research, at various stages of this study. PS, TL, JA, NI, PL, KE, and RL participated on root sample collection and processing and/or FTIR measurements. NI and PL analyzed subsets of the data in their M.Sc. theses. PS completed and organized the final dataset, performed all the data analyses and prepared all the figures and the tables for this manuscript and wrote the manuscript, with inputs from RL and helpful comments from the other co-authors.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00597/full#supplementary-material

REFERENCES

Bellon-Maurel, V., and McBratney, A. (2011). Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils–Critical review and research perspectives. *Soil Biol. Biochem.* 43, 1398–1410.

Bernard, J. M., Solander, D., and Kvet, J. (1988). Production and nutrient dynamics in Carex wetlands. *Aq. Bot.* 30, 125–147.

Bhuiyan, R., Minkkinen, K., Helmisaari, H., Ojanen, P., Penttilä, T., and Laiho, R. (2017). Estimating fine-root production by tree species and understorey functional groups in two contrasting peatland forests. *Plant Soil* 412, 299–316.

Clemmensen, K. E., Bahr, A., Ovaskainen, O., Dahlberg, A., Ekblad, A., Wallander, H., et al. (2013). Roots and associated fungi drive long-term carbon sequestration in boreal forest. *Science* 339, 1615–1618. doi: 10.1126/science. 1231923

Dawson, L., Mayes, R., Elston, D., and Smart, T. (2000). Root hydrocarbons as potential markers for determining species composition. *Plant, Cell Environ.* 23, 743–750.

Domisch, T., Finér, L., Dawud, S. M., Vesterdal, L., and Raulund-Rasmussen, K. (2015). Does species richness affect fine root biomass and production in young forest plantations? *Oecologia* 177, 581–594. doi: 10.1007/s00442-014-3107-3

Duboc, O., Zehetner, F., Djukic, I., Tatzber, M., Berger, T., and Gerzabek, M. (2012). Decomposition of European beech and Black pine foliar litter along an Alpine elevation gradient: mass loss and molecular characteristics. *Geoderma* 189, 522–531. doi: 10.1016/j.geoderma.2015.03.024

Esbensen, K. H., Guyot, D., Westad, F., and Houmoller, L. P. (2002). *Multivariate Data Analysis - in Practice: an Introduction to Multivariate Data Analysis and Experimental Design.* Oslo: CAMO Software.

Finér, L., Domisch, T., Dawud, S. M., Raulund-Rasmussen, K., Vesterdal, L., Bouriaud, O., et al. (2017). Conifer proportion explains fine root biomass more than tree species diversity and site factors in major European forest types. *For. Ecol. Manage.* 406, 330–350.

Finér, L., and Laine, J. (1998). Root dynamics at drained peatland sites of different fertility in southern Finland. *Plant Soil* 201, 27–36.

Finér, L., and Laine, J. (2000). The ingrowth bag method in measuring root production on peatland sites. *Scand. J. For. Res.* 15, 75–80.

Frolking, S., Roulet, N. T., Tuittila, E., Bubier, J. L., Quillet, A., Talbot, J., et al. (2010). A new model of Holocene peatland net primary production, decomposition, water balance, and peat accumulation. *Earth System Dynamics* 1, 1–21. doi: 10.1111/gcb.12672

Frolking, S., Talbot, J., Jones, M. C., Treat, C. C., Kauffman, J. B., Tuittila, E., et al. (2011). Peatlands in the Earth's 21st century climate system. *Env. Rev.* 19, 371–396.

Gill, R. A., and Jackson, R. B. (2000). Global patterns of root turnover for terrestrial ecosystems. *New Phytol.* 147, 13–31.

Gower, S., Krankina, O., Olson, R., Apps, M., Linder, S., and Wang, C. (2001). Net primary production and carbon allocation patterns of boreal forest ecosystems. *Ecol. Appl.* 11, 1395–1411.

Guo, D., Xia, M., Wei, X., Chang, W., Liu, Y., and Wang, Z. (2008). Anatomical traits associated with absorption and mycorrhizal colonization are linked to root branch order in twenty-three Chinese temperate tree species. *New Phytol.* 180, 673–683. doi: 10.1111/j.1469-8137.2008.02573.x

Iversen, C. M., Childs, J., Norby, R. J., Ontl, T. A., Kolka, R. K., Brice, D. J., et al. (2018). Fine-root growth in a forested bog is seasonally dynamic, but shallowly distributed in nutrient-poor peat. *Plant Soil* 424, 123–143.

Jackson, R., Canadell, J., Ehleringer, J. R., Mooney, H., Sala, O., and Schulze, E. (1996). A global analysis of root distributions for terrestrial biomes. *Oecologia* 108, 389–411. doi: 10.1007/BF00333714

Jandl, R., Rodeghiero, M., Martinez, C., Cotrufo, M. F., Bampa, F., van Wesemael, B., et al. (2014). Current status, uncertainty and future needs in soil organic carbon monitoring. *Sci. Total Environ.* 468, 376–383. doi: 10.1016/j.scitotenv. 2013.08.026

Joosten, H., and Clarke, D. (2002). *Wise Use of Mires and Peatlands: Background and Principles Including a Framework for Decision-Making.* Saarijärvi: International Mire Conservation Group and International Peat Society.

Kaštovská, E., Straková, P., Edwards, K., Urbanová, Z., Bárta, J., Mastný, J., et al. (2018). Cotton-grass and blueberry have opposite effect on peat characteristics and nutrient transformation in peatland. *Ecosystems* 21, 443–458.

Kusumo, B. H., Hedley, M. J., Hedley, C. B., and Tuohy, M. P. (2011). Measuring carbon dynamics in field soils using soil spectral reflectance: prediction of maize root density, soil organic carbon and nitrogen content. *Plant Soil* 338, 233–245.

Laiho, R., Bhuiyan, R., Straková, P., Mäkiranta, P., Badorek, T., and Penttilä, T. (2014). Modified ingrowth core method plus infrared calibration models for estimating fine root production in peatlands. *Plant Soil* 385, 311–327.

Laiho, R., and Finér, L. (1996). Changes in root biomass after water-level drawdown on pine mires in southern Finland. *Scand. J. For. Res.* 11, 251–260.

Laiho, R., Vasander, H., Penttilä, T., and Laine, J. (2003). Dynamics of plant-mediated organic matter and nutrient cycling following water-level

drawdown in boreal peatlands. *Global Biogeochem. Cycles* 17:1053. doi: 10.1029/ 2002GB002015

Laine, J., Vasander, H., and Laiho, R. (1995). Long-term effects of water level drawdown on the vegetation of drained pine mires in southern Finland. *J. Appl. Ecol.* 32, 785–802.

Legner, N., Meinen, C., and Rauber, R. (2018). Root differentiation of agricultural plant cultivars and proveniences using FTIR spectroscopy. *Front. Plant Sci.* 9:748. doi: 10.3389/fpls.2018.00748

Lei, P., and Bauhus, J. (2010). Use of near-infrared reflectance spectroscopy to predict species composition in tree fine-root mixtures. *Plant Soil* 333, 93–103.

Lorenz, K., and Lal, R. (2010). "Carbon dynamics and pools in major forest biomes of the world," in *Carbon Sequestration in Forest Ecosystems* (Dordrecht: Springer), 159–205.

Meinen, C., and Rauber, R. (2015). Root discrimination of closely related crop and weed species using FT MIR-ATR spectroscopy. *Front. Plant Sci.* 6:765. doi: 10.3389/fpls.2015.00765

Minkkinen, K., Ojanen, P., Penttilä, T., Aurela, M., Laurila, T., Tuovinen, J., et al. (2018). Persistent carbon sink at a boreal drained bog forest. *Biogeosciences* 15, 3603–3624. doi: 10.5194/bg-15-3603-2018

Mommer, L., Dumbrell, A. J., Wagemaker, C. N. A., and Ouborg, N. J. (2011). Belowground DNA-based techniques: untangling the network of plant root interactions. *Plant Soil* 348, 115–121. doi: 10.1007/s11104-011-0962-0

Murphy, M., Laiho, R., and Moore, T. R. (2009a). Effects of water table drawdown on root production and aboveground biomass in a boreal bog. *Ecosystems* 12, 1268–1282.

Murphy, M., McKinley, A., and Moore, T. (2009b). Variations in above-and below-ground vascular plant biomass and water table on a temperate ombrotrophic peatland. *Botany* 87, 845–853.

Murphy, M. T., and Moore, T. R. (2010). Linking root production to aboveground plant characteristics and water table in a temperate bog. *Plant Soil* 336, 219–231.

Naumann, A., Heine, G., and Rauber, R. (2010). Efficient discrimination of oat and pea roots by cluster analysis of Fourier transform infrared (FTIR) spectra. *Field Crops Res.* 119, 78–84.

Nichols, J.E., and Peteet, D. M. (2019). Rapid expansion of northern peatlands and doubled estimate of carbon storage. *Nat. Geosci.* 12, 917–921. doi: 10.1038/ s41561-019-0454-z

Paavilainen, E., and Päivänen, J. (1995). *Peatland Forestry: Ecology and Principles.* Berlin: Springer Science & Business Media.

Page, S. E., Rieley, J. O., and Banks, C. J. (2011). Global and regional importance of the tropical peatland carbon pool. *Global Change Biol.* 17, 798–818.

Peltoniemi, K., Straková, P., Fritze, H., Iráizoz, P. A., Pennanen, T., and Laiho, R. (2012). How water-level drawdown modifies litter-decomposing fungal and actinobacterial communities in boreal peatlands. *Soil Biol. Biochem.* 51, 20–34.

Pena, R., Lang, C., Naumann, A., and Polle, A. (2014). Ectomycorrhizal identification in environmental samples of tree roots by Fourier-transform infrared (FTIR) spectroscopy. *Front Plant Sci.* 5:229. doi: 10.3389/fpls.2014. 00229

Persson, H. Å (1983). The distribution and productivity of fine roots in boreal forests. *Plant Soil* 71, 87–101.

Picon-Cochard, C., Pilon, R., Revaillot, S., Jestin, M., and Dawson, L. (2009). Use of near-infrared reflectance spectroscopy to predict the percentage of dead versus living grass roots. *Plant Soil* 317, 309–320.

Pregitzer, K. S., DeForest, J. L., Burton, A. J., Allen, M. F., Ruess, R. W., and Hendrick, R. L. (2002). Fine root architecture of nine North American trees. *Ecol. Monogr.* 72, 293–309.

Proctor, C., and He, Y. (2019). Quantifying wetland plant vertical root distribution for estimating the Interface with the anoxic zone. *Plant Soil* 440, 381–398.

Rewald, B., and Meinen, C. (2013). Plant roots and spectroscopic methods– analyzing species, biomass and vitality. *Front Plant Sci.* 4:393. doi: 10.3389/fpls. 2013.00393

Robroek, B. J., Jassey, V. E., Kox, M. A., Berendsen, R. L., Mills, R. T., Cécillon, L., et al. (2015). Peatland vascular plant functional types affect methane dynamics by altering microbial community structure. *J. Ecol.* 103, 925–934.

Roumet, C., Picon-Cochard, C., Dawson, L. A., Joffre, R., Mayes, R., Blanchard, A., et al. (2006). Quantifying species composition in root mixtures using two

methods: near-infrared reflectance spectroscopy and plant wax markers. *New Phytol.* 170, 631–638. doi: 10.1111/j.1469-8137.2006.01698.x

Rumbaugh, M., Clark, D., and Pendery, B. (1988). Determination of root mass ratios in alfalfa-grass mixtures using near infrared reflectance spectroscopy. *J. Range Manag.* 41, 488–490.

Rydin, H., and Jeglum, J. K. (2013). *The Biology of Peatlands.* Oxford: Oxford university press.

Saarinen, T. (1996). Biomass and production of two vascular plants in a boreal mesotrophic fen. *Can. J. Bot.* 74, 934–938.

Sjörs, H. (1991). Phyto-and necromass above and below ground in a fen. *Ecography* 14, 208–218.

Straková, P., Anttila, J., Spetz, P., Kitunen, V., Tapanila, T., and Laiho, R. (2010). Litter quality and its response to water level drawdown in boreal peatlands at plant species and community level. *Plant Soil* 335, 501–520.

Straková, P., Niemi, M., Freeman, C., Peltoniemi, K., Toberman, H., Heiskanen, I., et al. (2011). Litter type affects the activity of aerobic decomposers in a boreal peatland more than site nutrient and water table regimes. *Biogeosciences* 8, 2741–2755. doi: 10.5194/bg-8-2741-2011

Straková, P., Penttilä, T., Laine, J., and Laiho, R. (2012). Disentangling direct and indirect effects of water table drawdown on above-and belowground plant litter decomposition: consequences for accumulation of organic matter in boreal peatlands. *Global Change Biol.* 18, 322–335.

Streit, J., Meinen, C., Nelson, W. C. D., Siebrecht-Schöll, D. J., and Rauber, R. (2019). Above-and belowground biomass in a mixed cropping system with eight novel winter faba bean genotypes and winter wheat using FTIR spectroscopy for root species discrimination. *Plant Soil* 436, 141–158.

Thomas, F. M., Molitor, F., and Werner, W. (2014). Lignin and cellulose concentrations in roots of Douglas fir and European beech of different diameter classes and soil depths. *Trees* 28, 309–315.

Tong, J., Xiang, W., Lei, P., Liu, J., Tian, D., Deng, X., et al. (2016). Prediction of tree species composition in fine root mixed samples using near-infrared reflectance spectroscopy. *Plant Biosyst.* 150, 412–419.

Vávřová, P., Penttilä, T., and Laiho, R. (2009). Decomposition of Scots pine fine woody debris in boreal conditions: Implications for estimating carbon pools and fluxes. *For. Ecol. Manag.* 257, 401–412.

Weishampel, P., Kolka, R., and King, J. Y. (2009). Carbon pools and productivity in a 1-km2 heterogeneous forest and peatland mosaic in Minnesota. *USA. For. Ecol. Manag.* 257, 747–754.

Westoby, M., and Wright, I. J. (2006). Land-plant ecology on the basis of functional traits. *Trends Ecol. Evol.* 21, 261–268. doi: 10.1016/j.tree.2006.02.004

White, K. E., Reeves, J. B. III, and Coale, F. J. (2011). Mid-infrared diffuse reflectance spectroscopy for the rapid analysis of plant root composition. *Geoderma* 167, 197–203.

Zhang, C., Chen, L., and Jiang, J. (2014). Why fine tree roots are stronger than thicker roots: the role of cellulose and lignin in relation to slope stability. *Geomorphology* 206, 196–202.

Hydrophobic and Hydrophilic Extractives in Norway Spruce and Kurile Larch and Their Role in Brown-Rot Degradation

Sophie Füchtner[1]*, Theis Brock-Nannestad[2], Annika Smeds[3], Maria Fredriksson[4], Annica Pilgård[5,6] and Lisbeth Garbrecht Thygesen[1]

[1] Department of Geoscience and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark, [2] Department of Chemistry, University of Copenhagen, Copenhagen, Denmark, [3] Laboratory of Wood and Paper Chemistry, Johan Gadolin Process Chemistry Centre, Åbo Akademi University, Turku, Finland, [4] Faculty of Engineering, Division of Building Materials, Lund University, Lund, Sweden, [5] Wood Research Munich, Technical University of Munich, Munich, Germany, [6] Research Institutes of Sweden (RISE), Gothenburg, Sweden

Extractives found in the heartwood of a moderately durable conifer (*Larix gmelinii var. japonica*) were compared with those found in a non-durable one (*Picea abies*). We identified and quantified heartwood extractives by extraction with solvents of different polarities and gas chromatography with mass spectral detection (GC-MS). Among the extracted compounds, there was a much higher amount of hydrophilic phenolics in larch (flavonoids) than in spruce (lignans). Both species had similar resin acid and fatty acid contents. The hydrophobic resin components are considered fungitoxic and the more hydrophilic components are known for their antioxidant activity. To ascertain the importance of the different classes of extractives, samples were partially extracted prior to subjection to the brown-rot fungus *Rhodonia placenta* for 2–8 weeks. Results indicated that the most important (but rather inefficient) defense in spruce came from the fungitoxic resin, while large amounts of flavonoids played a key role in larch defense. Possible moisture exclusion effects of larch extractives were quantified via the equilibrium moisture content of partially extracted samples, but were found to be too small to play any significant role in the defense against incipient brow-rot attack.

Keywords: extractives, brown-rot, spruce, larch, durability, moisture content, heartwood, GC-MS

1. INTRODUCTION

Wood used as building material in an outdoor environment (i.e., European Standard EN 335-1:2006 use class 3–4) is frequently exposed to high relative humidity or wetting, highly increasing the risk of degradation by wood-degrading micro-organism. The resistance of wood to degradation in such a setting is mainly determined by its inherent durability (natural or artificial) and its moisture sorption properties, apart from environmental factors and design (Brischke et al., 2006; Meyer-Veltrup et al., 2017; Brischke and Alfredsen, 2020). Naturally durable heartwood is formed in the center of the stem of some tree species by deposition of various metabolites in the tissue, called extractives (Rowe, 1989). For the heartwood of many species and all of the sapwood this is not the case, and thus artificial wood protection is needed (Kutnik et al., 2017). In Europe, many of the old wood-preservatives were banned due to their high toxicity, and consequently,

research and development focused their efforts on more environmentally benign ways to protect wood (Schultz and Nicholas, 2002; Singh and Singh, 2012). Of the new preservative compounds, many still have restricted use (ECHA, 2020), and thus further developments are necessary. One approach is to achieve a better understanding of the mechanisms underlying natural durability in trees, and the current study aims to contribute to this.

In northern Europe, wood products from conifers are widely used for construction purposes. Most of these are susceptible to degradation by brown-rot fungi, which are known to cause a rapid loss of strength already at an early stage of degradation (Bader et al., 2012; Arantes and Goodell, 2014; Wagner et al., 2015). Brown-rot fungi are widely spread, cellulose-degrading fungi, currently believed to start their attacks employing non-enzymatic oxidative degradation, followed by an enzymatic stage (Bader et al., 2012; Arantes and Goodell, 2014; Zhang et al., 2016). Non-enzymatic oxidative degradation of the cell wall relies on secretion and diffusion of low molecular weight substances and metal ions into the cell wall, where they react to create radicals that disrupt the cell wall polymers. The success of this process depends heavily on the presence of water in and around the cell wall, but also on the amount and mechanisms of extractives present (Schultz and Nicholas, 2002; Jebrane et al., 2014). In this work, we chose to compare Kurile larch (*Larix gmelinii var. japonica*), a moderately durable conifer (Scheffer and Morrell, 1998; Bergstedt and Lyck, 2007; Metsä-Kortelainen and Viitanen, 2009), to non-durable Norway spruce (*Picea abies*) (Scheffer and Morrell, 1998; Metsä-Kortelainen and Viitanen, 2009) in terms of their extractive composition and their susceptibility to the brown-rot fungus *Rhodonia placenta* after various extraction procedures. Their similar xylem anatomy should provide similar geometrical conditions to the fungus, and accentuate differences arising due to their extractive composition.

Extractives belong to many different chemical groups, and may be divided into more hydrophilic and more hydrophobic types (Giwa, 1973; Willför et al., 2003a, 2006a). Both types are found in the tree species studied here.

One role of the more hydrophobic extractives is possibly to repel water, as the amount of moisture in the wood is critical for the fungus' successful establishment (Meyer and Brischke, 2015; Brischke et al., 2017). For instance, it has been suggested that a certain cell wall moisture content is needed to form pathways within the cell walls to allow diffusion of the fungal low molecular weight substances (Zelinka et al., 2015; Hunt et al., 2018). Hydrophobic molecules, such as those found in the oleoresin of conifers, are good candidates for such functionality by decreasing the wettability of the cell wall (Eberhardt et al., 1994; Harju et al., 2002; Nzokou and Kamdem, 2004; Belt et al., 2017; Sjökvist et al., 2018). Oleoresin is present throughout the xylem, but its composition varies within the stem and changes upon injury or infection (Ekman, 1980; Hillis, 1987; Bohlmann et al., 2000; Holmbom et al., 2008; Mason et al., 2015, p. 73). Its primary components are fats and fatty acids (FAs) and various terpenoids. The non-volatile fraction is mainly composed of diterpenoids (DTs), among which resin acids (RAs) are the most abundant (50–75%) (Higuchi, 1997; Bohlmann et al., 2000; Holmbom et al.,

2008). Apart from a role in moisture regulation, the potential of RAs as biocides against multiple insects and micro-organisms was shown in numerous *in vivo* and *in vitro* assays (Micales et al., 1994; Nerg et al., 2004; Keeling and Bohlmann, 2006; Mason et al., 2015), including white rot fungi (Eberhardt et al., 1994), and brown rot fungi (Micales et al., 1994; Nerg et al., 2004). Sterols (STs), which are triterpenoids, have potential as growth retardants in bacteria, and a synergistic role with RAs has been proposed (Burčová et al., 2018). Representative chemical structures of compounds, belonging to these families and relevant for this study, can be viewed in **Figure S1.1–7**.

Spruce and larch heartwood also contain extractives that are more hydrophilic, mainly lignans (LI) in spruce and flavonoids (FL) in larch (see **Figure S1.9–12**, Willför et al., 2003a; Gierlinger et al., 2004; Nisula, 2018). Their functions include hindering the creation of (fungal) radicals by chelation of metals needed for the latter, neutralizing occurring radicals (antioxidants) and/or directly harming the fungus (biocide) (Rice-Evans et al., 1996; Willför et al., 2003c; Binbuga et al., 2008; Donoso-Fierro et al., 2009; Chen et al., 2014). Phenolic extractives may additionally play a role in moisture exclusion by bulking of the cell wall (Wangaard and Granados, 1967; Choong and Achmadi, 1991; Nzokou and Kamdem, 2004; Vahtikari et al., 2017). Both lignans (Rowe, 1989; Smith et al., 1989) and flavonoids (Dellus et al., 1997; Ostroukhova et al., 2012) partially form oligo- and polymers in heartwood, and so do the more hydrophobic resin components (Schaller, 2008; Smeds et al., 2016, 2018, p. 164).

Larch is somewhat a special case among the conifers, because it contains up to 30% w/w of a non-structural, hemicellulose-type polysaccharide - arabinogalactan (ArGal). Its amounts increase drastically at the sapwood-heartwood boundary, and fills the lumen of tracheid cells in the heartwood, especially those closer to rays (Côté et al., 1966; Giwa, 1973; Grabner et al., 2005a,b). Experiments show that ArGal has influence on the mechanical properties of larch heartwood (Luostarinen and Heräjärvi, 2013), but its role in fungal degradation, if any, remains controversial (Côté et al., 1966; Gierlinger et al., 2004; Hill et al., 2015).

Wood degrading fungi generally need a minimum water potential in the range of −4 to 0.1 MPa in order to grow (Griffin, 1977; Boddy, 1983; Griffith and Boddy, 1991; Schmidt, 2006), which corresponds to 97–99% relative humidity (RH). Above this range, liquid water accumulates in pits and cell lumina via capillary condensation (Engelund et al., 2013; Fredriksson and Thybring, 2019), while in the hygroscopic range (0 to about 98% RH) water is bound to hydroxyl groups in the cell wall. Although it has been shown that extractives alter the equilibrium moisture content (MC) in the hygroscopic range (Wangaard and Granados, 1967; Choong and Achmadi, 1991; Nzokou and Kamdem, 2004; Vahtikari et al., 2017), to the best of our knowledge, no publications exist on how extractives affect the MC in the over-hygroscopic range. The pressure plate technique allows investigation of the wood's MC in this range by precise regulation of the pressure applied to a water containing, tight cell (Fredriksson and Thybring, 2019). After a long equilibration time, the moisture content is determined gravimetrically. One aim of this study was, to explore whether

extractives also influence the over-hygroscopic moisture range in Kurile larch.

The best results for artificial impregnation is obtained by the combination of different functionalities (Schultz and Nicholas, 2002). Even more so, is it important to understand the individual roles of natural extractives, in order to be able to mimic the mechanisms at play when designing novel wood protection systems. Thus, partial removal of extractives of a certain polarity may give insight on their influence on wood durability against fungal degradation, and by assessing their effect on the MC of the wood possible interrelations between moisture and extractives content can be identified. With the aim of understanding the roles of hydrophobic and hydrophilic extractives in degradation and moisture regulation, we developed a multi-step extraction procedure that partially removes more extractives of a certain polarity from wooden sticks. The extractive composition was obtained by gas chromatography coupled to mass spectrometry and flame ionization detection to gain insights on the quantities of molecules present. Additionally, since moisture is a prerequisite for fungal attack, a specific aim of our study was to explore whether semi-selective removal of extractives with different polarity would affect the equilibrium moisture content in Kurile larch, especially in the important but under-explored over-hygroscopic range.

2. MATERIALS AND METHODS

A summary of procedures and methods used in this work can be viewed in **Figure 1**.

2.1. Wood Sampling and Preparation

Two clones of spruce (*P. abies*, 48 years old) and larch (*Larix gmelinii var. japonica*, 65 years old) were harvested near Hørsholm, Denmark, in the autumn of 2017. The individual trees will be referred to as spruce 1, spruce 2, larch 1 and larch 2. **Figure 1A** illustrates the sampling process. Disks (about 80 mm thick) were taken at 1.3 m stem height and stored at −20°C within 10 h from sampling. The disks' diameters were 323 and 280 mm for spruce 1 and 2, respectively. For larch 1 and 2 they were 381 and 406 mm, respectively. Using a band saw, 3–4 mm thick slices were radially cut from all around the frozen stem disks. The mature heartwood (red square in **Figure 1A**) was separated from sapwood and juvenile heartwood. For spruce, 10 growth rings from the pith were considered juvenile (Lindström, 2002). For larch, on average 15 year rings were similarly discarded

FIGURE 1 | (A) Preparation of wooden sticks, exemplified with Larch. The frozen wood disk was cut radially into slices. The mature heartwood (marked in red) was selected and one part was cut into sticks with dimensions 50 × 3–4 × 3–4 mm [longitudinal × radial × tangential]. The other part was milled. **(B)** Sticks and milled samples were extracted with different solvents according to the scheme. The extracts from the milled samples were quantified and identified using GC-MS/-FID. Some larch sticks were made into blocks and used for determination of the sorption isotherm. All extracted and native sticks of both species were subjected to brown-rot degradation. NAT, native; PHO, hydrophobic; PHI, hydrophilic; TOT, total; DCM, dichloromethane.

(Gierlinger and Wimmer, 2004; Luostarinen and Heräjärvi, 2013). The preselected mature heartwood was stored at $-80\,^{\circ}$C until further processing.

The samples were split into two groups: slices with 3–4 mm width were used to prepare sticks of 50 mm × 3 or 4 mm × 4 or 3 mm (L × W × B), as shown in **Figure 1A**. The sticks were freeze dried for 30–48 h and categorized into different growth ring patterns, i.e., broad EW + broad LW, thin repeating EW+LW, EW-LW-EW, and LW-EW-LW. These were then distributed evenly into four groups of 13–17 sticks each, so as to avoid possible bias from the radial position in the heartwood, where each stick was comprised of 1–3 annual rings. The rest of the frozen material was made into 10 mm^2 chips with garden shears. After freeze drying for 24 h, the chips were milled in a Retch ZM100 mill (1 mm sieve), freeze dried again and sieved through a 60 mesh filter, corresponding to a hole size of 250 μm. Before extraction, all samples were placed in a desiccator under vacuum with freshly dried molecular sieves for 2–18 h at room temperature.

2.2. Extraction and Further Processing

Extraction of milled samples and sticks was done using the Accelerated Solvent Extractor (Dionex$^{\text{TM}}$ ASE$^{\text{TM}}$ 350, Thermo Electron A/S, Scientific Instrument Division, 2650 Hvidovre, Denmark). The apparatus works under N$_2$ atmosphere, thereby lowering chances of oxidation artifacts possibly arising during extraction. Additionally, 1.38 MPa are applied to the extraction cell, allowing the use of volatile solvents at temperatures above their boiling points.

An extraction procedure was developed, aiming at producing samples where the hydrophobic (PHO) or hydrophilic (PHI) part of extractives had been removed, as well as a total extraction, removing both portions (TOT, **Figure 1B**). For the TOT extraction a sequence of four solvents was used in the following order: Heptane (anhydrous, 99%, Sigma-Aldrich), dichloromethane (DCM, SupraSolv$^{\circledR}$, Sigma-Aldrich), 96% ethanol (EtOH, SupraSolv$^{\circledR}$, Sigma-Aldrich), and demineralized water. For the PHO batches, only the hydrophobic solvents heptane and DCM were used. The PHI batches were treated only with hydrophilic solvents: 96% EtOH and demineralized water. For each procedure 13–17 sticks were used.

To counteract losses in extraction efficiency due to the geometry of the sticks, the maximum possible number of extraction cycles was used, increasing the probability of analytes being washed out of the maze of wood cells. The extraction conditions for each solvent and batch were 9 × 5 min cycles, 90°C and 150% rinse volume, except for ethanol, for which 100°C were used. Void volume in the ASE extraction cells was filled with clean quartz sand (50–70 mesh, Sigma-Aldrich).

In order to get an estimate of the extraction efficiency of the sticks, an amount corresponding to 13–17 sticks of milled material was extracted with the same procedure. For spruce 3–3.7 g were used per sample, while 5–5.6 g were needed for larch. Void volume in the ASE extraction cells was again filled with quartz sand, with a cellulose paper separating sample and sand. The ratio between the gravimetric yields of sticks and milled material (in mg/g) was used as a measure of extraction efficiency.

To test for residues in the total extract, the milled samples were additionally extracted with 95:5 acetone:water (2 cycles à 5 min, 100°C). Furthermore, in order to double check the results, 4 g of milled material from each of the trees was extracted with a control sequence adapted from Willför et al. (2003a), employing one hydrophobic and one hydrophilic solvent. Heptane was used instead of hexane for the sake of lower toxicity and the acetone step was repeated 3 times instead of 2 times.

The extracts were stored inside the pressurized extraction bottles at 4°C until further processing. The sample volume was reduced to 50 ml and 2–4 × 10 ml aliquots were used to determine the gravimetric yield. Two of the water samples were lost due to a mistake in the laboratory.

For the larch water samples, the ArGal was precipitated out of the solution in triplicates, using the procedure described in Luostarinen and Heräjärvi (2013). The precipitate of each sample was dried at 60°C over night in a ventilated oven and the weight determined.

2.3. Identification and Quantification of the Extracts
2.3.1. Gas Chromatography
Gas Chromatography (GC) was used for the separation of individual analytes of the respective extracts. Identification was done using the response from the Mass Spectrometrometer (MS). Quantification was done based on the signal of the Flame Ionization Detector (FID) obtained from one run per solvent fraction of each clone ($n = 2$). The extracts of all milled samples were run.

Gas chromatography of the water fraction of both species did not show any peaks, likely because the majority of the material extracted by water are polysaccharides or other polymers. Thus, water-ethanol (WE) supernatant of the ArGal-free samples was pooled from 3 determinations and subsequently dried over Na$_2$SO$_4$ (ACS reagent, $\geq$ 99.0%, anhydrous, Sigma-Aldrich). After filtration and ethanol-wash of the filtrate, the now alcoholic solution was reduced to 5 ml. Only the hydrophilic extracted larch 1 WE-samples were analyzed by GC-MS, but here the extracts of the milled material and of the sticks were compared.

An aliquot of 0.5 mg/ml of each extract was mixed with 200 μl internal standard (0.2 mg/ml heneicosanoic acid—HIA and betulinol—BET, both from Sigma-Aldrich, in methyl-tertiary-butyl-ether) and derivatized before subsequent separation according to the procedures found in Nisula (2018), Willför et al. (2003b), and Zule et al. (2015). Derivatization reagents were acquired from commercial sources and used as received.

GC-MS and -FID experiments were performed on an Agilent 6890N/5973N-system (MS Consult, 2740 Skovlunde, Denmark), equipped with an S/SL inlet for sample introduction, using bleed and temperature optimized (BTO) high temperature septa and an Agilent Ultra Inert, split, low pressure drop liner with glass wool. The inlet was connected to the analytical column (HP-1, 25 m, 0.20 mm ID, 0.11 μm) by way of 5 m Agilent Ultimate Plus deactivated fused silica tubing (0.25 mm ID), used as a sacrificial pre-column. Eluents were split between the MSD for identification and the FID for quantification using an EPC

pressure controlled CFT-splitter, sending 10% to the MSD and 90% to the FID. The original protocol for the GC-MS can be found in Örså and Holmbom (1994). Before and after all runs, injections of neat derivatizing reagent were used to passivate the chromatographic system. The GC-Program was as follows: Injection volume 1 μl, split 10:1. Starting temperature 120°C, heating rate 6°C /min to 325°C, hold time 4 min. Flow rate 0.9 ml He/min, solvent delay 3 min, FID data were collected at 10 Hz.

To estimate the linearity of the method over several orders of magnitude, a calibration row was run with one varying (HIA) and one constant standard (BET). The concentrations of HIA where 0, 0.005, 0.025, 0.050, 0.502, 1.003, 3.010, and 5.017 mg/ml, and BET was constant at 0.55 mg/ml. The series was run once with all concentrations, and twice leaving out the 1 and 5 mg/ml samples. A linear regression based on the ratio of the two peaks was made for each of the runs (in Microsoft® Excel 2016), on a 95% confidence level. All three regressions gave an $R^2 > 99.97\%$.

The limits of detection (LOD) and of quantification (LOQ) were determined for each calibration row by means of the first 4 points (0–0.05 mg/g), using Equations (1) and (2), respectively.

$$LOD = 3S/b \qquad (1)$$
$$LOQ = 10S/b \qquad (2)$$

where S is the standard deviation of the y-intercepts and b the slope of the corresponding regression. The average of the three determinations was used as the final value. The LOD was found to be 0.004 ± 0.001 mg/ml, corresponding to 0.005–0.008 mg/g dry wood (sample size 3–5 g). The LOQ was found to be 0.013 ± 0.004 mg/ml, which corresponds to 0.015–0.025 mg/g dry wood (sample size 3–5 g).

2.4. Sorption Isotherm of Kurile Larch

Samples of Kurile larch were used to obtain the sorption isotherm, using two techniques: conditioning above saturated salt solutions (4 points, 64–95% RH, hygroscopic range) and the pressure plate technique (3 points, 99.64–99.99%, over-hygroscopic range). An overview of the different RH-levels and techniques used is given in **Table 1**.

The TOT and NAT samples of larch 1 were used to obtain the respective absorption and desorption isotherms, in both the hygroscopic and over-hygroscopic moisture ranges, covering a total of 7 points. Thus, 7 sticks with different EW-LW patterns from the NAT and TOT groups were used. However, due to limited sample availability, sorption isotherms for the PHO and PHI samples were only determined in desorption and in the in the over-hygroscopic range. Three sticks of each treatment were used. All the sticks were labeled and cut into 5–7 equally sized pieces ($\approx 0.7 \times 0.3 \times 0.4$ mm, L × W × B). Each piece of the same stick was assigned to a different RH-level (7 for NAT and Tot, 3 for PHO and PHI). Because we suspected that the sticks were not equally well-extracted in the central part as on the borders, the pieces were distributed over the levels so that one level did not for instance contain solely center pieces.

For the absorption isotherm, the samples were dried, put into individual open Eppendorf tubes in a vacuum oven at 60°C for 24 h. The oven was then allowed to cool under reduced pressure and the cooling of the samples was finalized in a vacuum desiccator. Freshly dried molecular sieves were added to each of the tubes for storage until weighing. For determination of desorption isotherms, the samples were initially water saturated. First, they were placed in round bottom flasks RH-level-wise and set under vacuum for 15 min. Degassed Milli-Q water was added to each flask using a syringe, followed by another minute of degassing with the vacuum pump. Then the samples were allowed to stand under reduced pressure for 1h, after which atmospheric pressure was re-established and they were stored in water until weighing. Due to this water saturation step, we expected that the hydrophobic extracted pieces were extracted to a certain degree—at the very least the arabinogalactan must have been affected.

Weighing before conditioning: For the water-saturated samples, the surface water of the samples was removed by rolling each piece over a wet cellulose based cloth (Wettex, Vileda, Freudenberg home & cleaning, solutions, AB, Malmö). Then, the piece was quickly placed on the balance and the mass recorded with a resolution of 0.01 mg. The dry samples were weighed inside of a tared weighing glass filled with dry molecular sieves.

TABLE 1 | Measurement techniques and respective conditions for the determination of the sorption isotherm of Kurile larch.

Method	Salt used	Relative humidity [%]	Pressure [bar]	Water potential [J kg⁻¹]	Calculation of RH % by
Saturated salt solution	NH_4NO_3	64.1	42.01	-60364	Equation 30 in Mozurkewich (1993)
	NaCl	75.5	26.6	-38172	
	KCl	85.0	15.3	-21972	Linear interpolation of RH values at 20 and 25°C in Table 2 of Greenspan (1977)
	KNO_3	94.5	5.3	-7643	
Pressure plate	n.a.	99.67	4.44*	-44.0	Rearrangement of Equation 1 in Fredriksson and Johansson (2016)
	n.a.	99.89	1.46*	-15.0	
	n.a.	99.97	0.45*	-5	

*All the RH values shown for the salt solutions were calculated as detailed in the table, and the pressure as well as the water potential were determined from these. For the pressure plate method, the pressure is the pressure that was applied in the experiment (marked with *), which was then used to calculate the corresponding RH and water potential as explained in the main text.*

For the hygroscopic range, saturated salt solutions according to **Table 1** were placed in small climate boxes, equipped with RH sensors. Once the RH stabilized, the samples were introduced. The relative humidities generated by the different salt solutions were determined as detailed in **Table 1**.

The pressure plate technique, which gives information on the relation between the water potential and the moisture content of the material (Defo et al., 1999), was used to determine sorption isotherms in the over-hygroscopic range. In this study, a custom-built pressure plate system was used (Fredriksson and Thybring, 2019) where specimens were conditioned in the range 0.4–4.4 bar, corresponding roughly to 99–100% RH (see **Table 1**). The experimental procedure as described by Fredriksson and Thybring (2019) was used. All the samples were kept in the climate boxes/pressure plate cells for a period of 2 months at 20.5 ± 0.3°C.

The conversion from relative humidities to pressure (salt solutions) and vice versa (pressure plate) was done by rearrangement of Equation (1) in Fredriksson and Johansson (2016). The water potential was calculated from the relative humidities by Equation (6) in Cloutier and Fortin (1994).

Weighing after treatment: A glove box containing an analytical balance (resolution 0.01 mg) was used. For each of the humidity levels, the RH in the glove box was adjusted accordingly using a humidity generator (2500 Humidity Generator, Thunder Scientific Corporation, Albuquerque, New Mexico, USA). For the highest levels wet cloths were additionally placed in the glove box. Finally, the desorption samples were dried and weighed as described above. The final sample size after all procedures was 4–7 replicates per group.

To enable comparison between unextracted and differently extracted samples, an adjusted moisture content, u (g/g), was determined relative to the unextracted weight, as shown in Equation (3) below:

$$u = (m_{\mathrm{w}}/m_{\mathrm{dry}}) * (1 + [\%extractives]) \tag{3}$$

where m_{w} (g) is the mass of water in the specimen and m_{dry} (g) is the dry mass of the specimen, which for the extracted material was the dry mass after extraction and [% extractives] corresponds to the amount of gravimetrically determined extractives in percent; i.e., for the NAT samples, this term is zero.

2.5. Fungal Degradation

Cultures of European *R. Placenta*, European strain FPRL280, were cultivated on malt agar and stored at 4°C for 2 weeks. Mycelial flocks were taken from these cultures for the experiments below.

To get a view on a more advanced state of degradation and to test for the virulence of the fungus an agar-block test was made. Six sterilized NAT sticks of spruce 2 and larch 2 were horizontally placed two by two on plastic grids in agar-filled petri dishes and incubated at 23°C and 70% RH. After 8 weeks, the samples were harvested, the mycelium removed, and the wet as well as dry weight determined.

With a very simplified version of a pole in ground-contact, we chose a modified soil-block test inspired by Zhang et al. (2016) to assess the impact of extractives on the initial phase of degradation. The fungus was allowed to grow bottom-up in the longitudinal direction of the wood for 2 weeks as shown in **Figure 2**. Nine sticks from each of the different extraction treatments, and native controls were marked at 2/3 of the height. The dry weight was determined, and all sticks were equilibrated at 23°C and 70% RH for 5 weeks prior to sterilization by autoclaving and inoculation. Larch 2 was additionally autoclaved before equilibration, due to suspicion of mold. The sticks were placed vertically on pre-inoculated pine feeder strips on an autoclaved soil mixture. All glasses were incubated until the hyphal front (HF) reached the marked threshold (33.3 mm). The average height of the HF in spruce was 35.5 ± 4.5 mm and was reached within 14–18 days. For larch the average HF height was 39.8 ± 5.7 mm and was reached in 15 days. Because of the geometry of the sample, there is no guarantee the hyphal front inside of the sticks reached the same height as on the outside, and thus we consider the weight loss the appropriate measure to describe the degradation. Statistical evidence for this can be found in the **Supplementary Information**.

After harvest, the samples were weighed immediately, stored at −25°C until all the samples had been harvested, and then dried for weight loss determination.
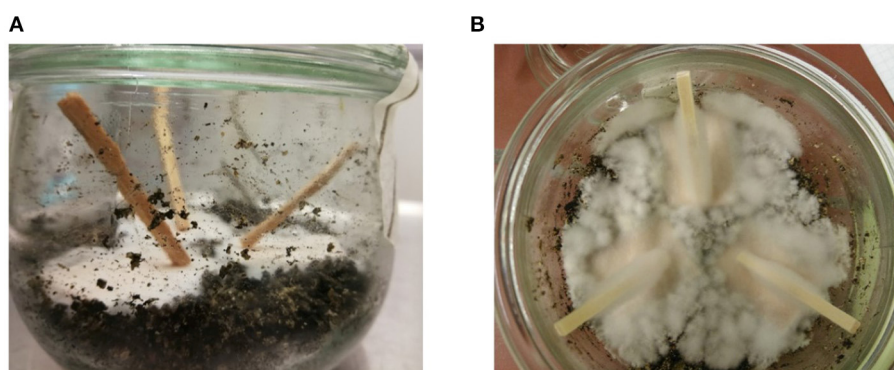

FIGURE 2 | Sample setup for directional growth of *R. placenta*, inspired by Zhang et al. (2016). **(A)** Kurile larch before degradation. **(B)** Norway spruce with hyphal front after degradation.

2.6. Data Analysis

2.6.1. Gravimetric Yields

The gravimetric yields were analyzed in OriginPro® 2017 (OriginLab Corporation, Northamtom, MA 01060, USA, www.OriginLab.com). The standard deviations of sums and quotients of variables were calculated according to formulas taking into account statistical error propagation, i.e., formula (1) and (2) in Andraos (1996).

2.6.2. Chromatograms

The mass spectral data were used to identify the peaks in the chromatograms by using a mixture of the NIST database and the database created at the Laboratory of Wood and Paper Chemistry, Åbo Akademi University. The corresponding retention times (RT) were taken from the total ion current chromatograms. For quantification of all the extracts, the peak areas of the FID-chromatogram were obtained using MSD Enhanced ChemStation© (Agilent Technologies Inc., CA 95051, USA, www.agilent.com) and the amounts calculated based on the area of the internal standard peaks. Fatty acids/alcohols, resin acids, other diterpenoids, flavonoids, carbohydrates, and unknown compounds were quantified using the HIA standard, while lignans and sterols where quantified using betulinol. A correction factor of 1.2 was used for lignans, as recommended elsewhere (Willför et al., 2003a; Zule et al., 2015; Nisula, 2018; Zule et al., 2017, a.o.).

Due to the high split ratio and higher detection rate of the FID, these spectra show more resolved peaks as well as many small peaks not seen with the MS detector. The latter are quantified within the unknown (UNK) class, together with peaks that could not be identified with MS. Fatty alcohols were pooled with the fatty acids, but resin acids in different oxidation states were classified as "other diterpenoids." Due to variations in peak intensity, some peaks below the LOD were actually detected, but disregarded from analysis. Analytes detected below the LOQ, where nevertheless included in the total sums of each chemical class.

2.6.3. Estimation of Sticks Composition

Under the assumption that similar proportions of extractives were removed in sticks as in milled material, we estimated the amount of each chemical group (i) left in the sticks, as compared to the milled TOT samples ($f_{i, sticks}$, see Equation 4). This was achieved by weighing the chromatographic yields (y) of each milled solvent fraction (j) and chemical group (i) with the respective extraction efficiency (EF). Then each chemical group was summed for each procedure and divided by the sum of yields of the milled TOT samples. For the PHO samples, the chromatographic yields obtained from the respective milled TOT ethanol fractions were added. The average of 2 trees is reported.

$$f_{i,sticks} = \sum_{j=1}^{N} y_{i,j} \times (1 - EF_j)/ \sum y_{i,j} \qquad (4)$$

Where N is the solvents relevant for the respective treatment. We have also added the remaining ArGal present in larch, which was directly calculated from the gravimetric yields, as $(1- m_{sticks} / m_{milled})$.

2.6.4. Sorption Isotherm

The data were analyzed in OriginPro® 2017. Absorption and desorption data were tested individually. First a test for normality (Shapiro-Wilk) was performed for each extraction treatment and humidity level tested. For the desorption samples, all groups were normally distributed, except the groups NAT at 94% and TOT at 74% RH, hence these were excluded from follow up analysis. Among the absorption data, only the NAT at 94% group was not normally distributed, and thus excluded from further statistical analysis.

A 2-way ANOVA was performed on the rest of the desorption and absorption data, separately. The extraction treatment and the RH were used as factors, as well as the interaction terms. A power analysis was added, to test for Type II errors.

2.6.5. Fungal Degradation

The statistical analysis of the fungal degradation was done in OriginPro® 2017. The variation in growth height and sampling days was tested using ANOVA as detailed in the **Supplementary Information**. It showed that WL was a suitable factor to describe the decay resistance.

Normality tests showed that weight loss in both species had to be log transformed for further analysis. The equality of variance was tested before the ANOVA, showing equal variance for spruce ($p < 0.05$, $\alpha = 0.05$), and unequal variance for larch ($p < 0.02$, α 0.05). The normalized data were subjected to 2-way-ANOVA for each species separately, testing for tree, treatment and interaction terms. Tukey's test for comparison of means and power analysis were used *post-hoc*.

3. RESULTS

3.1. Milled Samples: Extract Yields and Composition

Figure 3 shows the average gravimetric yields of the different extraction strategies of milled spruce and larch samples. The heptane and DCM fractions of the total extracted samples serve as the references for the milled PHO samples, which are therefore not shown separately. From total extraction (**Figure 3A**) it is evident that the amounts of hydrophobic extracts were very similar for both species, with 7–8 mg/g dry wood for heptane and 2–4 mg/g dry wood for DCM. In spruce, the yields obtained for both hydrophilic fractions were at similar levels as the heptane extracts ($\approx$ 10 mg/g dry wood). In larch on the other hand, ethanol yielded about 5x more material on average, and the water fraction even more with about 9x higher yields, making up about 50–60% of the total sum of all solvent fractions (see also **Table S1**). Substantial yield differences between the two larch trees were found, reflected by the larger standard deviations (error bars in **Figure 3**).

When ethanol was used as the first solvent (**Figure 3B**), the gravimetric yields where higher than when ethanol was the third solvent. The spruce yields were slightly lower than for the sum of heptane, DCM and EtOH fractions of the TOT extraction. The

FIGURE 3 | Gravimetric yields of all the solvent fractions and extraction strategies for milled Norway spruce and Kurile larch. **(A)** Average yields of the milled, total extracted samples (TOT) obtained from two clones of *Picea abies* (green) and *Larix gmelinii var. japonica* (blue), with respective standard deviations. The TOT heptane and DCM yields should be considered as reference for the hydrophobic extracted samples (PHO). **(B)** Average yields per species obtained from the extraction with hydrophilic solvents, Ethanol (EtOH) and Water.

average ethanol yield of the two larch trees was again much higher than for spruce, and comparable to the summed yields of the first three TOT solvents. For both species, the water fractions yielded about the same amounts as for the TOT extraction.

Upon the addition of cold ethanol to the water fractions of spruce, only very small amounts of material precipitated and could not be reliably quantified. In the case of larch, addition of cold ethanol resulted in large amounts of precipitate, known to be arabinogalactan (Côté et al., 1966; Luostarinen and Heräjärvi, 2013). Nearly all the yield of these water fractions was composed of the polysaccharide. For larch 1, the average yield of dry precipitate for the milled TOT and PHI extractions was 58.6 ± 2.7 mg/g. Reflecting the higher water extract yields, the ArGal precipitate of larch 2 amounted to 95.6 ± 3.8 mg/g on average.

To test whether the extraction was complete after the use of four different solvents, we added a fifth extraction step for the milled TOT samples. Using an acetone:water mixture (95:5), we found that in both species, 2–3 mg/g of additional material could be extracted.

In the following, the composition of the heptane and ethanol extracts of the total extraction (as assessed by GC-MS) will be described first, because the other fractions were mixtures of the latter. The composition of the additional replicates, where a control extraction procedure (heptane and acetone) was used, can be viewed in **Table S2**. As detailed further in the **Supplementary Information**, the summed yields of the individual extractive groups agreed well with the quantities obtained with the four step extraction procedure.

3.1.1. Hydrophobic Extracts - Heptane
As shown in **Figures 4A,B**, the hydrophobic heptane extracts of both species contained fatty acids and alcohols (FAs), resin

acids (RAs), other diterpenoids (DTs), and sterols (STs), but their proportions, as well as the number of detected analytes differed. The numbers above each column in **Figure 4** show the total chomatographic yield of the respective fraction in mg/g dry wood. The corresponding chromatograms can be viewed in the **Figures S2**, **S4**. The relative proportions of FAs and DTs were about 2 times higher in larch (20 and 22%, resp.) than in spruce (13 and 8%, resp.). On the other hand, spruce had 10 times higher proportions of STs (around 15%) and 3 times higher contribution from unidentified compounds (UNK, 30%). Only resin acids were found in similar proportions in both species, and where the most abundant group of hydrophobic analytes, amounting to around 40% of detected analytes. Please view **Table 2** for a summary of the compounds detected using four solvents.

Structures of the most abundant and/or relevant analytes are shown in **Figure S1**.

3.1.2. Hydrophilic Extracts - Ethanol
Compared to the heptane extracts, the composition of the ethanol extracts was less complex in terms of the number of different chemical classes detected (**Figures 4A,B**). The chromatograms can be found in **Figures S2**, **S4**.

The spruce ethanol extracts of the TOT extraction were largely dominated by lignans (around 90%; **Figure 4A**), while the dominant phenolics in larch were flavonoids (also around 90%; **Figure 4B**). It may also be noteworthy that in both species the main phenolic compounds (bold in **Table 2**) were present at much greater concentrations than the other compounds, which was not the case for the hydrophobic analytes. In spruce, the difference was about 9-fold, while in larch it was much greater with over 50-fold. Note also, that the absolute amounts we found for the larch flavonoids are comparable to *L. decidua*,

FIGURE 4 | Relative chemical composition of all the detected solvent fractions and extraction strategies for milled Norway spruce and Kurile larch. For reference the absolute chromatographic yields are shown as numbers above the respective bars in mg/g dry wood (*n*=1). **(A)** Total extracted samples (TOT) of spruce (S). Each tree is shown separately, as indicated by the index. **(B)** Total extracted samples of larch (L). **(C)** Ethanol (EtOH) fractions of the hydrophilic extracted samples (PHI). The water-ethanol fractions were only run for larch 1 (sticks (s) and milled (m) extracts). Indexes before solvent designations indicate the sequence by which the solvents were used. DCM, Dichloromethane; FA, fatty acid; RA, resin acid; DT, diterpenoids; ST, sterols; FL, flavonoids; SL, sesquilignans; LI, lignans; CA, carbohydrates and other structural components; UNK, unknown.

rather then *L. gmelinii var. japonica* as reported by Nisula (2018). Additionally, larch 2 had only one largely dominating flavonoid (90% of detected flavonoids)—taxifolin (TAX), while larch 1 had dihydrokaempferol (DHK) and taxifolin at an almost 50:50 ratio, DHK being slightly higher. The total ethanol yield was also higher for larch 2 (**Figure S4**). In both species, low amounts of monomeric carbohydrates and lignin components were also detected ($\approx 3\%$), and $\approx 9\%$ of the ethanol extracts were unknown compounds.

3.1.3. Other Solvents and Solvent Sequences
In spruce, the low-yielding DMC fraction was composed of lignans to 82–88%, while 77–80% of the corresponding larch extracts consisted of flavonoids (**Figures 4A,B**). The rest was composed of mainly unknowns (8–12%) and residual FAs, RAs, DTs, and STs, as well as 2–4% carbohydrates and lignin monomers. The chromatograms are shown in **Figures S2**, **S4**.

Figure **4C** shows the composition of the extracts obtained from the hydrophilic extraction procedure in spruce and larch was not specific for the hydrophilic compounds only. The PHI ethanol extracts of spruce (**Figure 4C** and **Figure S3**) resulted in a mixture of the analytes found in the heptane and ethanol fractions discussed above. Thus, this extract gives an overview of the proportions of GC-detectable extractives, that is 15–20% hydrophobic analytes (FAs, RAs, DTs, and STs) and around 70–75% of lignans. The rest was small amounts of carbohydrates/lignin components and unknown compounds. Similarly, the PHI ethanol extracts of milled larch were also a mixture of the heptane and ethanol fractions (**Figure 4C** and **Figure S5**). The hydrophobic analytes amounted to 4–7% of this extract, and flavonoids to 81–86% and the rest was again carbohydrates/lignin components and unknown analytes.

The water fraction of spruce did not show any peaks in the chromatograms, thus, no information on their composition is

available (see the chromatogram in **Figure S2**). In the case of larch, removal of ArGal together with a dehydration step of the remaining supernatant, enabled the now alcoholic fraction to be successfully run on the GC. For the milled specimen of larch 1, (**Figure 4C**), flavonoids accounted for 55% of the yield, and carbohydrates for about 30%.

3.1.4. Differences in Gravimetric and Chromatographic Yields
We compared the gravimetric and chromatographic yields of all milled samples in **Figure 5** and found that the GC-detectable fraction varies between solvents and species. Only about 30% of the less polar spruce extracts were reliably detected with GC-FID (including unknown peaks), while up to 38% of the ethanol fractions (TOT and PHI) were detected as compared to the gravimetric yields. In case of larch, 40–50% of the heptane and DCM fractions were detected, and around 65% of the ethanol fractions (TOT and PHI).

3.2. Composition of the Sticks
The gravimetric yields obtained for the sticks are listed in **Table S1**. The resulting extraction efficiency of the sticks compared to the milled material is shown in **Table 3**. Note that the composition of the milled TOT heptane and DCM fractions were used for the calculation of the extraction efficiency of the PHO sticks, because we considered them representative for both groups. We found that the extraction efficiency was similar in spruce and larch. With some exceptions, the sticks yielded 50–90% for the heptane, DCM and water fractions relative to the milled material. With ethanol maximally 30–60% could be extracted from the sticks.

Precipitation of ArGal from the larch water fractions of TOT and PHI sticks yielded about 40% of what the corresponding milled specimen yielded (not shown in **Table 3**).

TABLE 2 | Analytes detected in each solvent fraction of Norway spruce and Kurile larch, grouped by chemical class and the class sum ($n = 2$), as compared to the literature.

Chemical group	*P. abies*		*L. gmelinii var. japonica*	
	Detected	Literature References	Detected	Literature References
Fatty acids /alcohols (FA), free	Total: 0.20–0.36 mg/g **Palimitic acid** (16:0), margaric acid (17:0), **pinoleic acid** (5,9,12–18:3), **linoleic acid** (9,12–18:2), **oleic acid** (9–18:1), behenyl alcohol (22:0), linoceryl alcohol (24:0)	0.6–3.4 mg/g (Willför et al., 2003a) 0.25–1.4 mg/g depending on latitude, detected more types (Nisula, 2018)	Total: 0.63–0.67 mg/g palmitic acid, margaric acid, **Pinoleic acid**, **linoleic acid**, **oleic acid**, sciadonic acid (5,11,14–20:3)	1.2–3.2 mg/g, found more types (Nisula, 2018) 0.86-1.1 mg/g in *L. decidua* (Zule et al., 2015)
Resin acids (RA)	Total: 0.55–1.00 mg/g levopimaric acid, pimaric acid, sandarapimaric acid, isopimaric acid, **palustric acid**, **dehydroabietic acid**, abietic acid	0.48–1.5 mg/g (Nisula, 2018) 0.8–2.4 mg/g (Willför et al., 2003a)	Total: 1.17–1.38 mg/g levopimaric acid, communic acid, **isopimaric acid**, **palustric acid**, dehydroabietic acid, abietic acid, neoabietic acid	0.47–1.7 mg/g (Nisula, 2018) 4.6–7.3 mg/g (RA+DT) in *L. decidua* (Zule et al., 2015)
Other diterpenoids (DT), free	Total: 0.09–0.26 mg/g Thunbergol, neoabienol, only S1: palustral[*], palustrol[*]	Present (Nisula, 2018) 0.1–1.3 mg/g (Willför et al., 2003a) [*] RA precursors (Higuchi, 1997; Keeling and Bohlmann, 2006), were not reported in (Willför et al., 2003a; Nisula, 2018)	Total: 0.60–0.75 mg/g Thunbergol, manool (1 and 2), palustral[*], (iso)pimarol[*], neoabietol[*], **cedrol**	Thunbergol - present, manool - 0.07–0.38 mg/g (Nisula, 2018) cedrol has not been reported for *Larix gmelini var. japonica* nor any other larch species [*] RA precursors (Higuchi, 1997; Keeling and Bohlmann, 2006) (Higuchi p 258)
Sterols (ST), free	Total: 0.26–0.33 mg/g campesterol, **sitosterol**, sitostanol	0.24–0.36 mg/g (Nisula, 2018) 0.2–0.3 mg/g (Willför et al., 2003a)	Total: 0.05–0.06 mg/g campesterol, **sitosterol**	0.12–0.14 mg/g (Nisula, 2018) 0.14–0.18 mg/g in *L. decidua* (Zule et al., 2015) Cycloartenol was additionally found in these references (Zule et al., 2015; Nisula, 2018)
Lignans (LI), monomeric	Total: 4.24–5.03 mg/g Dihydroxymatairesinol, 7r-todolactol, iso-lariciresinol, todolactol A, alpha-conidendric acid, lignan a, lignan b, matairesinol, hydroxymatairesinol 1, **hydroxymatairesinol 2**, mixed lignan signal, lariciresinol, iso-hydroxymatairesinol	0.87–15 mg/g (Nisula, 2018) 0.0–2.5 mg/g (Willför et al., 2003a) secoisolariciresinol, pinoresinol and nortrachelogenin were not detected in our case (Willför et al., 2003a; Nisula, 2018)	Detected, but not confirmed: isoliovil and dihydroxymatairesinol	0.2–1 mg/g (Nisula, 2018) trace amounts (Zule et al., 2017)
Flavonoids (FL)	Detected, but not confirmed: naringenin, taxifolin	Trace amounts of dihydrokaempferol, taxifolin (Willför et al., 2004a)	Total: 18.45-35.73 mg/g naringenin, dihydrokaempferol, taxifolin 1, taxifolin 2	0.49-5.4 mg/g (Nisula, 2018) 15-36 mg/g in *L. decidua* (Nisula, 2018) 10-15 mg/g (FL + LI) in *L. decidua* (Zule et al., 2017)

*See **Table S2** for the quantities obtained by the 2-solvent sequence. Fatty acids are given by their common names, the numbers in brackets show (position of double bond-number of carbons:number of double bonds).*

The chromatograms of the remaining dehydrated water extract of larch PHI sticks revealed that the composition was not the same as for the milled counterpart, as shown in **Figure 4C**. Flavonoids and carbohydrates were identified in both samples, but in the sticks' extract the amount of flavonoids was much higher than for milled specimen—in absolute and relative amounts—making up 80% of the yield ($\approx$ 5 mg/g).

Despite uncertainties, we tentatively estimated the residual amounts of different extractive groups in the sticks after the different extraction procedures (**Figure 6**), using Equation (4). The milled TOT samples were used as a reference, because we

consider it the most complete extraction. Residual ArGal was estimated based on the gravimetric yields of the precipitates.

The levels of retained hydrophobic extractives was slightly higher for the PHO extractions than for the TOT extraction in both species, even though a contribution from ethanol in the TOT procedure can be excluded, as no hydrophobic analytes were detected in this fraction. Differences may arise from experimental variations, such as the moisture content of the material upon starting the extraction. On the other hand, total extraction removed lignans (LI, SL) more efficiently in spruce (approx. 50% left), than flavonoids in larch (approx. 70% left).
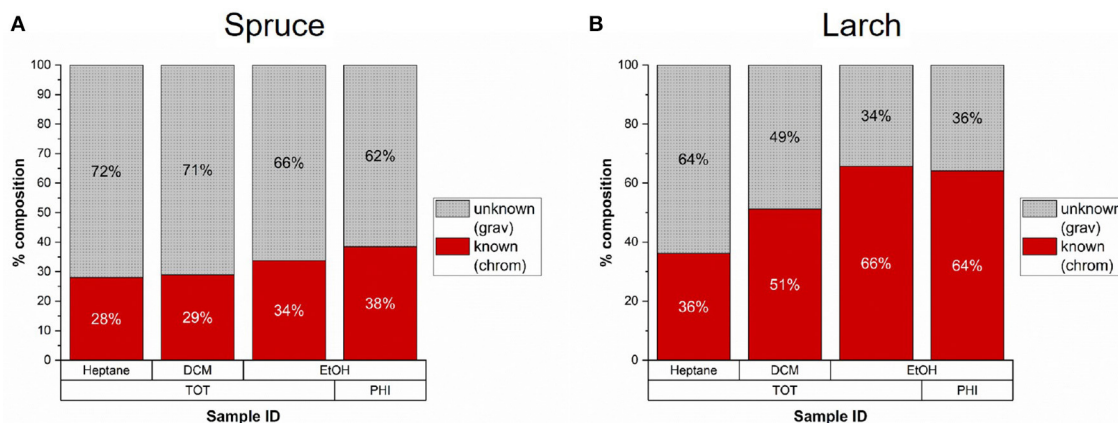
FIGURE 5 | Average percentage of extractives from milled samples with known composition as determined by gas chromatography compared to the total gravimetric yield ($n = 2$). **(A)** In Norway spruce 28–38% of the gravimetric yields were detected by chromatography. **(B)** In Larch 36–66% of the extracts were detected by chromatography.

TABLE 3 | Average extraction efficiency of the sticks that were used in the sorption isotherm and fungal degradation experiments.

Tree ID	Stick/milled	Heptane	DCM	EtOH	H_2O
S1	TOT-s/TOT-m	0.4	0.3	0.4	n.a.
	PHO-s/TOT-m	0.5	0.5	–	–
	PHI-s/PHI-m	–	–	0.4	0.9
S2	TOT-s/TOT-m	0.8	0.5	0.6	1.1
	PHO-s/TOT-m	0.6	0.6	–	–
	PHI-s/PHI-m	–	–	0.4	1.2
L1	TOT-s/TOT-m	0.6	0.6	0.3	0.5
	PHO-s/TOT-m	0.5	1.2	–	–
	PHI-s/PHI-m	–	–	0.3	0.6
L2	TOT-s/TOT-m	0.8	1.0	0.2	n.a.
	PHO-s/TOT-m	0.6	0.5	–	–
	PHI-s/PHI-m	–	–	0.2	0.5

The extraction efficiency was calculated as the ratio between the average gravimetric yield of sticks and associated milled sample. The yields of heptane and DCM fraction of the TOT extraction were considered representative for the respective fractions of the PHO procedure, and the calculations therefore based on the TOT yields. DCM, dichloromethane; EtOH, ethanol; S, spruce; L, larch; 1/2, tree 1 or 2; s, stick; m, milled; TOT, total extracted; PHO, hydrophobic extracted; PHI, hydrophilic extracted.

One reason could be that DCM extracted about 20% of the total monomeric lignans in spruce, while flavonoids in larch seem to be more polar, because DCM removed only about 1–4 % of total flavonoids.

The ethanol step of the PHI extraction removed a significant portion of the rather hydrophobic extractives (FA, RA, DT, ST), although lower amounts as compared the other extraction procedures in both species. The extraction efficiency of ethanol as the first solvent (PHI) was lower in larch than in spruce. A contributing factor here could be the presence of ArGal in larch, which is highly insoluble in alcohol, and while occupying the

lumina of many tracheid cells, may reduce the flux of solvent during extraction. More than half of the polysaccharide remained in the TOT and the PHI groups (**Figure 6B**), although it is likely that the water extraction step relocated parts of it. As water was not used in the PHO group, we consider the relocation issue to be smaller for these samples. Nevertheless, solvent flux during extraction might be different in spruce than in larch.

The metabolic profiles described above can be summarized as follows: While the TOT samples contain the lowest amount of extractives in all categories, the PHO samples contain the highest amounts of hydrophilic extractives, including most unidentified compounds and large amounts of ArGal in larch. The PHI extraction resulted in samples having a higher proportion of hydrophobic material compared to the TOT group, but also higher amounts of hydrophilic compounds. It should thus be considered the least efficient extraction, especially in larch. A likely consequence of the incomplete extraction of the sticks is a relocation of extractives to areas where they might not be present in the native wood, which might influence their mode of action and/or efficacy.

3.3. Sorption Isotherm of *Larix gmelinii var. japonica*

The sorption isotherm for Kurile larch is shown in **Figure 7A** for native and total extracted wood in the range of 64–100% RH. The MC of the TOT samples was generally slightly higher than for NAT in both absorption and desorption. The adjusted MC of the TOT samples deviated more from the NAT samples in the range of 94.5–99.89% RH in desorption mode and up to 99.97% in absorption mode. Above and below that, the curves coincide (**Figure 7A**). Nevertheless, 2-way ANOVA showed that this difference was only significant in absorption mode ($p = 0.028$, $\alpha = 0.05$, power $= 0.59$) in the RH range of 64–99.97%. Also, as expected, the differences in MC at different RH-levels was significant ($p = 0$, $\alpha = 0.05$).

FIGURE 6 | Estimates of the sticks' monomeric extractive composition after extraction: The chromatographic yields of total extracted, milled samples were weighed by corresponding extraction efficiency and extractive groups summed over the solvent fractions. The water fractions were disregarded. The average of both trees is shown, respectively. **(A)** Spruce sticks. **(B)** Larch sticks. FA, fatty acid; RA, resin acid; DT, diterpenoids; ST, sterols; FL, flavonoids; SL, sesquilignans; LI, lignans; CA, carbohydrates and other structural components; ArGal, arabinogalactan; UNK, unknown.
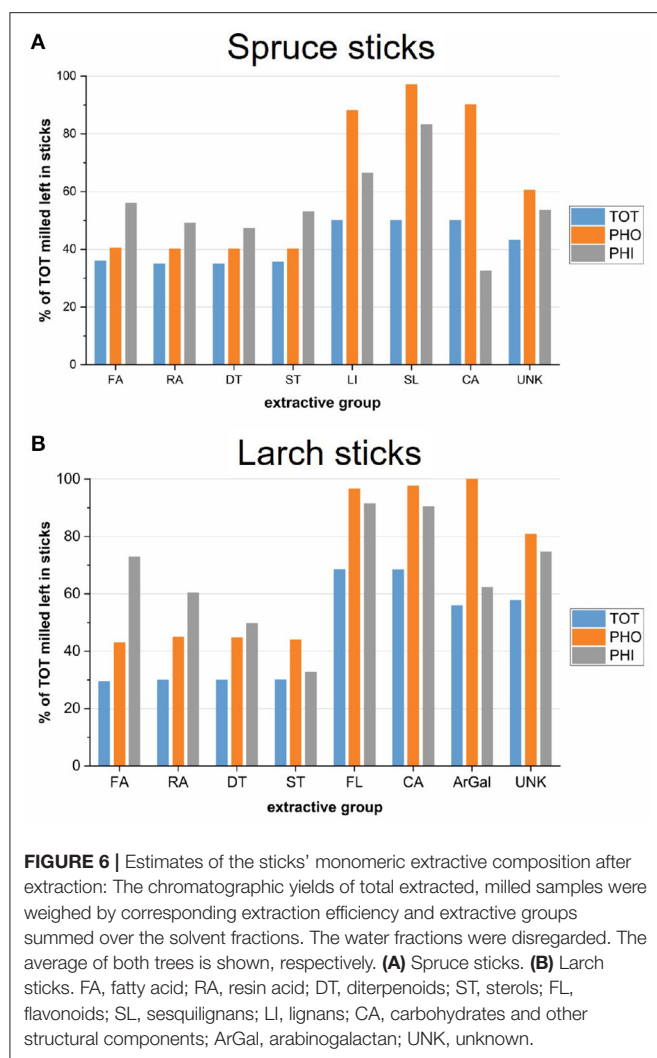
Figure 7B shows the desorption isotherm in the over-hygroscopic range for native, totally extracted, hydrophobic and hydrophilic extracted larch. Although NAT had the lowest MC compared to the extracted samples, none of these differences were significant ($p > 0.05$, $\alpha = 0.05$). Again, the difference in MC was different between the RH-levels as expected ($p = 0$, $\alpha = 0.05$).

3.4. Fungal Degradation

The 8 week degradation test of native Norway spruce 2 and native Kurile larch 2 with *R. placenta* resulted in an average weight loss of $30.0 \pm 12\%$ for spruce ($n = 5$), while it was about half for larch, with $14.8 \pm 6\%$ ($n = 6$). This shows that, as expected, Kurile larch was more resistant to degradation by this brown-rot fungus.

A similar pattern was found for the 2 weeks degradation test, after which a small weight loss could already be observed. For every treatment group, spruce lost close to twice the amount of mass compared to larch, except for the hydrophilic extraction, where the weight loss was almost identical for both species. The

average weight loss percentage of each species and treatment is shown in **Table 4**.

The average weight loss of spruce was between 3.5 and 5% over all groups. The 2-way ANOVA performed on the spruce data showed that the trees were not significantly different from each other ($p > 0.05$, $\alpha = 0.05$, power = 0.15), but that the treatments were different ($p \ll 0.001$, $\alpha = 0.05$, power = 0.99), but also that the trees were differently affected by the treatments (interaction terms, $p = 0.046$, $\alpha = 0.05$, power = 0.94). Tukey's test showed that the TOT and the PHO groups were more affected than the NAT group, and that the TOT group was also different from the PHI group. For the interaction terms, the same was found for spruce 1, but in the case of spruce 2 only the TOT group was different from the NAT group.

For Kurile larch, the range of weight loss was between 1.8 and 3% over all groups. The ANOVA showed that the trees were significantly different from each other ($p \ll 0.001$, $\alpha = 0.05$, power = 0.999), which can be seen from the individual weight loss of the trees, were larch 2 lost about half the weight compared to larch 1 in all treatment groups (**Table 4**). In fact, the weight loss recorded for larch 1 was much more comparable to the weight loss of the spruce specimen, despite having much higher phenolics content. The treatments affected the weight loss of larch differently ($p \ll 0.001$, $\alpha = 0.05$, power = 0.999) and also the trees were affected differently by different treatments (interaction terms, $p = 0.03$, $\alpha = 0.05$, power = 0.96). Tukey's test revealed that the WL after all extraction procedures were different from the native group, but not from each other. For the interactions, we found that in both trees the WL of the PHI group was different from the NAT group. However, in larch 1, only the weight loss of the PHO group was higher than the NAT group, while in larch 2 the weight loss of the PHI group was higher than for NAT.

We also looked at the moisture content, calculated relative to non-degraded mass of the samples. After 8 weeks of degradation the reduced moisture content was the same for native spruce ($47 \pm 13\%$) and native larch ($43 \pm 9\%$). A similar relation is found for samples after 2 weeks of exposure to *R. placenta*, although the MC was lower. The reduced MC of 35% was again the same for all four trees, irrespective of the extraction procedure or weight loss.

4. DISCUSSION

4.1. Extraction and Analysis

Before any further conclusions are drawn from the collected data, a few methodological issues should be acknowledged.

The performance of a living tree is yet different than what we can observe here, as we removed the majority of volatile monoterpenes during drying, which are known to play an important role in defense (Schaller, 2008, p. 164). Furthermore, despite having increased the number of extraction cycles to a maximum, the yields of the sticks were much lower than for the milled counterparts. It is questionable whether an increase in cycle duration would have significantly increased the yields further. We also found differences in the composition of the water extracts of larch 1 sticks and milled specimen. Similarly,
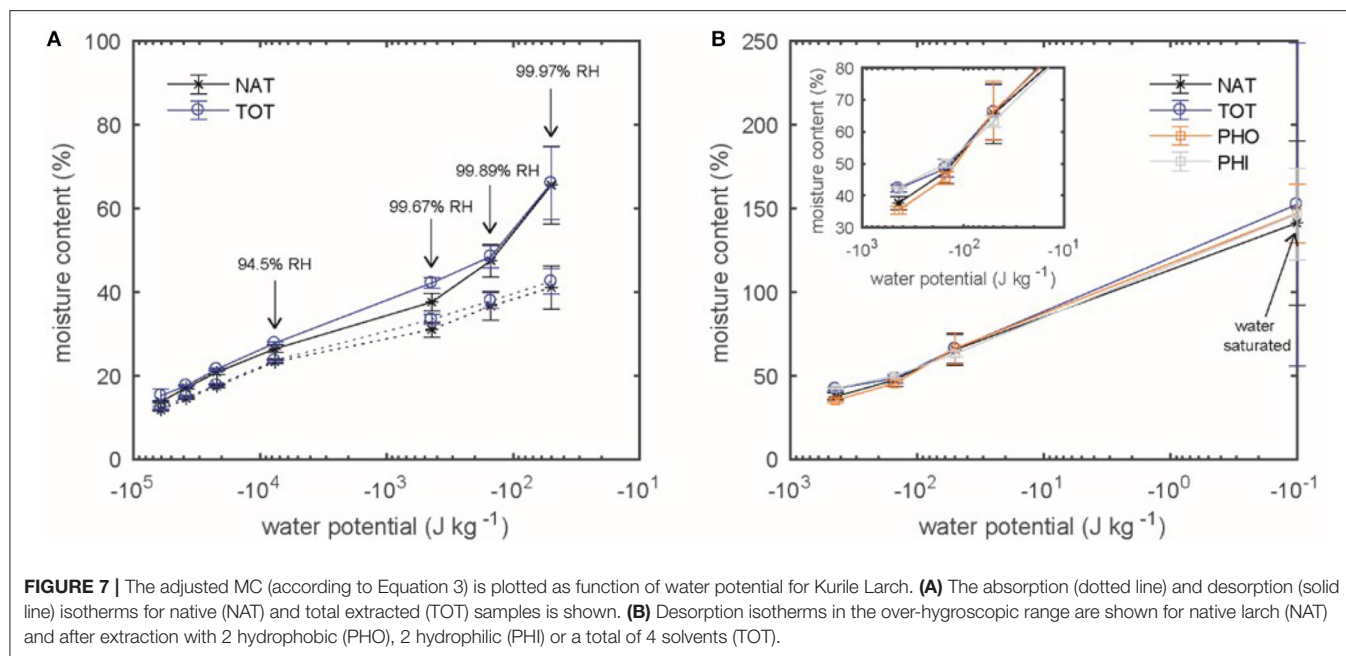
FIGURE 7 | The adjusted MC (according to Equation 3) is plotted as function of water potential for Kurile Larch. **(A)** The absorption (dotted line) and desorption (solid line) isotherms for native (NAT) and total extracted (TOT) samples is shown. **(B)** Desorption isotherms in the over-hygroscopic range are shown for native larch (NAT) and after extraction with 2 hydrophobic (PHO), 2 hydrophilic (PHI) or a total of 4 solvents (TOT).

TABLE 4 | Average weight loss percentage and associated standard deviations of sticks of two clones of Norway spruce and Kurile Larch after different extraction treatments.

Treatment	Spruce 1[A]	Spruce 2[A]	Average spruce	Larch 1[D]	Larch 2[E]	Average larch
Native	3.7 ± 1.2 [a, b]	3.3 ± 0.8 [a]	3.5 ± 1.0 [a]	2.5 ± 1.5 [d]	1.2 ± 0.3 [d]	1.9 ± 0.9 [d]
Total	5.3 ± 1.5 [a, c]	5.1 ± 1.3 [b]	5.2 ± 1.4 [b]	3.2 ± 1.0 [d, f]	2.8 ± 0.8 [e]	3.0 ± 0.9 [e]
Hydrophobic	6.0 ± 1.9 [c]	4.2 ± 0.5 [a, b]	5.1 ± 1.2 [b, c]	3.8 ± 1.0 [e, f]	1.8 ± 0.4 [d, e]	2.8 ± 0.7 [e]
Hydrophilic	3.6 ± 1.1 [a, b]	4.6 ± 1.7 [a,b]	4.1 ± 1.4 [a, c]	4.3 ± 1.2 [e, f]	2.6 ± 0.4 [e]	3.4 ± 0.8 [e]

For each treatment n = 18, except native of trees with index 1, where n=17. Statistically significant differences on a 95% confidence level are denoted with different letters pairs, all at p ≪ 0.001. Uppercase letters are used for the average trees (column-wise), lowercase for the treatments (row-wise); Letters "a-c" were used for spruce, "d-f" for larch.

Ekeberg et al. (2006) compared the yields and content of acetone-extracted powders and blocks of *Pinus sylvestris L.* and found that i.e., fatty acids were extracted more efficiently from blocks than resin acids or stilbenes. As we did not run the sticks' extracts on the GC, we can base our discussion only on the estimated composition of the sticks (**Figure 6**).

4.1.1. Gas Chromatography and Polymers

We obtained a detailed view on the monomeric composition of our extracts, but noticed that, depending on the solvent, up to 70% of the gravimetrically determined mass, was not detected with the GC-MS/FID method used in this work. There can be various (cumulative) reasons for this, which will be shortly outlined below:

Low-molecular-weight compounds, especially many of the acidic extractives, are known to form esters with different alcohols, such as glycerol, sterols and other diterpenyl and aliphatic alcohols, as well as with various carbohydrates (Fengel and Wegener, 1989; Rowe, 1989; Otto and Wilde, 2001; Willför et al., 2003a). As a consequence, some compounds do not have sufficient volatility to reach the detector with the column used in this work (Zule et al., 2015).

The most significant contribution is certainly the presence of polymers, the molecular weight of which is too high for regular gas chromatography. Polymerization reactions may also occur during extraction and work-up (Holmbom, 1999, i.e., p.128), and thus introduce artifacts. Our extracts were analyzed fresh, but we saw some precipitates in the acetone extracts of both species as the samples aged. Detailed information, as well as respective countermeasures can be found in the literature (i.e., Willför et al., 2006b).

Nevertheless, most of the mass not detected with the GC-MS method used int his work is expected to be from natural polymers. The rather hydrophilic lignans and flavonoids (Willför et al., 2004b, 2006b; Fedorova et al., 2015), and more recently, also the rather hydrophobic resin acids and fatty acids (Smeds et al., 2016, 2018) have been shown to be present as oligomers and polymers. In *Picea obovata Ledeb.* from Siberia, oligomeric lignans were found to constitute up to 40% of ethyl acetate extracts (Fedorova et al., 2015). Several studies on spruce knotwood found that up to 70% of the ethanolic extracts were polymeric lignans (Willför et al., 2003a, 2004b), but also polymerized fatty acids, resin acids and other diterpenoids (Smeds et al., 2016). This agrees surprisingly well with our

findings, since only 30% of the spruce extracts could be quantified by GC-FID. Therefore we suggest, that there may be similar proportions of polymeric material in the stemwood as in the knotwood of spruce, although with much higher abundance in knots (Willför et al., 2003a). Similarly in larch, Ostroukhova et al. (2012) mention a publication where the resin of Siberian larch should contain 30% of a polymer, consisting of taxifolin subunits. This again corresponds rather well with the amount of substance we were able to quantify from the ethanol/acetone extracts using gas chromatography.

Investigations of pine knotwood determined large amounts (> 50%) of polymerized fatty acids, resin acids and other diterpenoids (Smeds et al., 2018). Although, to our knowledge, such studies have not been conducted for larch, and only in knotwood of pine (Smeds et al., 2018) and spruce (Smeds et al., 2016), it may very well be that it is a property of resin in general to form polymers of its constituents. Thus, it could explain the difference in gravimetric and chromatographic yields of the heptane fractions of spruce, as well as larch. There are many techniques that allow identification and quantification of the polymeric fractions (Holmbom, 1999; Willför et al., 2003a, 2006b; Fang et al., 2013; Smeds et al., 2016; Zule et al., 2017). However, none of these options were used in the present study.

4.1.2. Extractives in Spruce and Larch

Our Norway spruce specimens showed fewer analytes in some of the extractive groups than reported in literature, and slight differences between the two clones were seen. Our data from the 2-solvent sequence confirm the quantities (**Table S2**), and these also corresponded well with literature references (**Table 2**). We found relatively low amounts of fatty acids and fewer types than reported in the references (Nisula, 2018); also levopimaric acid (a RA) was not found. This might be due to natural variation with regard to genetics or the environment, or because these analytes were below the LOD of our setup, or even due to artifacts from extraction and/or work-up. Alternatively, some infection might have happened already in the live tree, which has been shown to decrease the amount of free FAs, but also deplete the amount of levopimaric acid (Ekman, 1980). The fatty alcohols lignoceryl alcohol and behenyl alcohol were also detected and are known to be the most common alcohols found in *P. abies* (Fengel and Wegener, 1989, p. 194). Due to the detection method used, we have no information about the amounts and types of esterified FAs, but they are known to be less abundant in spruce heartwood than in sapwood (Willför et al., 2003a). Among the diterpenoids, neoabienol had a higher concentration in spruce 1, explaining the different proportions of DTs among the two trees (**Figure 4**). The relative amounts of sterols in our spruce specimen were also different, but the absolute amounts were the same, the difference likely arising from the difference in unknown compounds. Also the lignans detected in our specimen were fewer than described in the literature (Willför et al., 2003a; Nisula, 2018), but again, some of these might have been below our LOD. The precipitate obtained from the water fraction of spruce suggests the presence of hemicelluloses, possibly of arabinogalactan type (Willför and Holmbom, 2004).

Similarly, in Kurile larch a few analytes were not detected—or detected and not found in the literature. Apart from this, the extractive profiles corresponded well with previous literature for the same (Nisula, 2018) and other species of larch (Zule et al., 2015, 2017; Nisula, 2018). As opposed to spruce, the heptane fraction of Kurile larch did not contain fatty alcohols, but triglycerides are also expected to be present (Zule et al., 2015). We found relatively large amounts of cedrol in our specimens, but could not find evidence for this in the literature for any species of larch. We do not know if this is an artifact or not. As mentioned, there was a rather important difference in the flavonoid profiles of the two larch clones. Firstly, larch 2 contained almost twice the amount of flavonoids of larch 1. Secondly, and as reported by several sources, Taxifolin 1 (TAX) was the dominating flavonoid of larch 2 (Venäläinen et al., 2006; Zule et al., 2017; Nisula, 2018), whereas in larch 1, DHK and TAX were present at a 50:50 ratio. According to literature, DHK is one of the direct precursor molecules of TAX (Winkel-Shirley, 2001), indicating that larch 1 had a problem with the conversion to TAX. The large difference in yields of the water fractions resulted in a 30 mg/g difference of precipitated arabinogalactan, being higher in larch 2. The amounts detected correspond to 6–10% w/w. These quantities are in agreement with findings from other authors and species of larch, where the amounts range from 5 to 20% (Côté et al., 1966; Luostarinen and Heräjärvi, 2013). The ArGal yields for the sticks were approximately 40–50% of the milled counterpart. After removal of the hemicellulose, the water fractions were successfully run on the GC and revealed that the water fraction was composed of carbohydrates and flavonoids, but the relative composition between sticks and milled specimens was not the same. Not surprisingly, a larger contribution of carbohydrates was found for the milled samples, but in absolute numbers this difference was not so drastic. The relatively higher release of flavonoids in the sticks was probably due to the limited extraction efficiency as compared to the milled material. Because the extraction efficiency of the ethanol fraction in spruce was higher on average (50%) than for larch (30%), we also suggest that the presence of ArGal in the lumen of tracheids might hinder the flow of solvent and extractives during extraction with solvents other than water.

Summarizing, we should keep in mind that (1) we do not know the composition of the fraction not detected by GC, (2) that we cannot say for sure whether the extracts of the sticks contain the same proportions of molecules as do their milled counterparts and (3) that extraction was incomplete for the sticks, possibly causing migration of analytes within the wooden tissue.

4.2. Extractives in Relation to Moisture and Fungal Degradation

4.2.1. Moisture Content at Isothermal Equilibrium and After Fungal Degradation

Investigating the moisture content of the Kurile larch samples, we found that the TOT samples had a higher MC than the NAT samples in absorption, especially in the range between 94.5 and 99.89% RH. This difference, however, was not statistically significant in desorption, although we observed that at saturation

(100% RH), the NAT samples had the lowest MC (about 140%) and the TOT samples the highest with 150% MC.

Previous studies, performed in the hygroscopic range, indicate that there is a difference in both absorption and desorption modes and that it is dependent on the amount of extractives (Wangaard and Granados, 1967; Choong and Achmadi, 1991; Nzokou and Kamdem, 2004; Vahtikari et al., 2017). In the hygroscopic moisture range, this observation can be attributed to a bulking effect of (polymeric) extractives, occupying space in cell walls. However, changes seen in the over-hygroscopic range where water is present also outside cell walls should rather be related to presence of extractives in pits and macro voids.

It seems that also the proportions of hydrophilic and hydrophobic extractives influence the MC in the hygroscopic range, which can be seen from the data of Nzokou and Kamdem (2004) where the resin-rich pine had a lower equilibrium MC as compared to cherry and oak at comparable extractives content. No such trend was seen from our desorption data in the over-hygroscopic range, obtained for all types of extractions of larch. Another factor may be the radial position the sample was taken from, as the MC of earlywood and latewood may differ from year to year (Hill et al., 2015). The differences in MC reported in the above mentioned references are all in the same range with the standard deviations of our findings. Therefore, it might not be so straightforward to interpret the cause of these observations and studies should be conducted to assess whether these effects can actually be distinguished from each other.

Different fungi have different moisture requirements for successful infestation of wood (Meyer and Brischke, 2015; Brischke et al., 2017). Yet, our data suggest that the extractives do not play a key role in the moisture regulation of our larch specimen, as the MC was not different enough between the extracted and native samples to explain the differences we found in degradation. This is supported by the fact that we found the same moisture content for all samples after degradation by the brown-rot, no matter what kind of extractives had been removed, and also independent of the species. It would be necessary to repeat this experiment with a non-degraded control sample, incubated under the same conditions, to see how much of that water comes from the humidity conditions inside the petri-dish and the soil-block jar. This would allow the estimation of how much water comes from fungal respiration and/or active water transport by the fungus (Thybring, 2017).

Similar to our study, it was also concluded for Scots pine (Jebrane et al., 2014) and several species of larch (Venäläinen et al., 2006; Jebrane et al., 2014) that the natural durability of wood is not necessarily determined only by its moisture content, but also by chemical properties of extractives, as will be further discussed below.

4.2.2. Defense Strategy of Spruce

The weight loss of the Norway spruce samples was in the same range with Finnish spruce (Metsä-Kortelainen and Viitanen, 2009), but higher than found for Austrian spruce, which was more in range with larch specimen of the present study (Fackler et al., 2010). Despite its overall low durability, we found that spruce hydrophobic extractives play a relevant role in defense of the cell wall material against fungal attack, supported by the

action of lignans. Very indicative was that the average weight loss we found for the PHO group of spruce was statistically the same as the TOT group, which both had a higher weight loss than the native samples. This suggests that the removal of the hydrophobic extractives alone had a more significant impact on the brown-rot attack, than did the additional removal of hydrophilic lignans, which are removed and/or relocated to higher extends by the PHI and TOT extractions.

We found that the major contributors to the heptane extracts were fatty acids, resin acids, some diterpenoids and sterols, out of which resin acids were dominant. These compound classes are all part of the typical oleoresin found in rays and resin channel of conifers (Higuchi, 1997), which is also used for the defense in active tissue (i.e., sapwood). Micales et al. (1994) tested pure and mixed RAs *in vitro* and detected fungitoxic effects at concentrations of 0.02% on pine wood. They noted that, generally, the abietan type resin acids (i.e., dehydro-/abietic acid, palustric acid, **Figure S1.3,4**) were more fungitoxic than the pimarane-types (i.e., iso-pimaric acid, sandaracopimaric acid, **Figure S1.5**). In our specimens the total resin acid concentration was higher than that (around 0.04%), and the most abundant analyte, dehydroabietic acid, was present at about 0.02%. The more hydrophilic lignans in the ethanol fraction exhibit relatively strong antioxidant behavior when used in higher concentrations (i.e., 2%) (Willför et al., 2003c). The total concentration of monomeric lignans of our specimens was only about 0.4%, with HMR as the main analyte. This molecule was found to be a good antioxidant *in vitro* (Rice-Evans et al., 1997; Saarinen et al., 2000; Willför et al., 2003c). In our study, resin acids were found at concentrations levels found to be fungitoxic, and lignans found at much lower concentrations than their proven effect levels. This could explain why the extraction of the hydrophobic analytes affected the spruce wood proportionally more than did the removal of lignans. Note also that spruce 1 was more severely affected, which might be due to the fact that it had a higher amount of hydrophobic substances, but a lower amount of hydrophilic ones than spruce 2. The fewer lignans in that tree could not back-up the loss of the RAs, while in spruce 2, which had a higher monolignan content, they could protect the cell wall material to a higher degree.

We saw that the TOT extraction affected the sticks samples significantly more than the PHI extraction (at least in spruce 1), but the slightly higher yield of the total extraction might not suffice to explain this difference alone. A possible additional effect could be the migration of hydrophobic molecules upon extraction, as mentioned in Nzokou and Kamdem (2004). Following this thought, and assuming more hydrophobic extractives are infiltrated in the cell wall, the extraction with heptane and DCM might have caused migration of hydrophobic substances to the cell wall surface, which were subsequently removed by ethanol extraction. In the case of the PHI samples, with ethanol as the first extraction step, the same migration may have happened, but the subsequent water extraction could not remove them efficiently, thus leaving a fungitoxic and hydrophobic layer on the cell wall surface. It was suggested that the absorption rate could be additionally reduced by such an effect and thus, potentially, early phase of fungal growth slowed down (Vahtikari et al., 2017; Sjökvist et al., 2018). Likely

we were not able to see this effect because our samples were pre-equilibrated for several weeks prior to inocculation.

4.2.3. Defense Strategy of Larch

For the 8 weeks degradation test, we found the same weight loss of Kurile larch grown in Denmark, as reported for European and Siberian larch of several provenances, degraded by *R. placenta* (Jebrane et al., 2014). The hydrophilic flavonoids, especially taxifolin, seemed to be the crucial ingredient to the defense of the wood via their antioxidative potential, but being supported by hydrophobic extractives. We saw that the two larch clones performed very differently in all tested groups, except in the TOT group, where the WL was almost the same. In this group the difference to the control was only significant for larch 2, indicating that the extractives it contained were more relevant to its durability than for larch 1. In contrast, larch 1 was more affected by the hydrophobic treatment, as compared to larch 2 and the controls, thus the hydrophobic extractives played a bigger role in this tree.

The phenolics content of larch 2 was about twice as high as for larch 1, while the hydrophobic extractives were present in similar amounts. This supports findings from previous studies that used the total phenolics or water-soluble extractives as a way to predict or explain decay resistance in larch trees (Scheffer and Cowling, 1966; Gierlinger et al., 2003; Jebrane et al., 2014; Nisula, 2018). Notably, the difference between the two trees lay in their respective taxifolin content, being much larger in larch 2. Venäläinen et al. (2006) also found a good correlation between the taxifolin content of several clones of Siberian larch to their weight loss. The major phenolic compounds in larch 1 were DHK and TAX at almost equal concentration, but this did not give nearly as good a protection as did TAX alone at a higher concentration in larch 2 (see **Table 4**). Interestingly, the only structural difference between DHK and TAX is an additional hydroxyl group at the 3' position of the B-ring (**Figure S1.8**) of TAX, as seen in **Figure S1.9,10**. It was shown that this greatly enhances the antioxidant potential in aqueous systems (Rice-Evans et al., 1996; Binbuga et al., 2008). One of the reasons is a better stabilization of the flavonoid radical when both of these positions are hydroxylated. The same configuration also increases their metal chelating abilities (Rice-Evans et al., 1996), which is important when considering that brown-rot fungi need iron to generate radicals (Binbuga et al., 2008; Ringman et al., 2014). The difference in durability of these two larch trees may thus be explained not only by the quantity, but also the increased antioxidative power, as well as metal chelating abilities of TAX over DHK.

The fact that larch 1 was more affected by the hydrophobic extraction points at the possibility that, similarly to spruce, the hydrophobic extracts are an important back-up to the more abundant hydrophilic compounds, especially when these are not so efficient. That is, larch 2 was less affected by the PHO extraction, because the high taxifolin content protected the wood well, which the mixture of TAX and DHK could not do for larch 1. Finally, the case of larch 1 gives a good example of why the taxifolin content may be a more reliable indicator for larch durability, rather than total phenolics or flavonoid

content, because it was found to have a similar WL as our spruce specimen, even though the phenolics content was much higher.

4.2.4. Degradation - Spruce vs. Larch

The total phenolics content is often used as a parameter in durability studies (Scheffer and Cowling, 1966; Gierlinger et al., 2003; Jebrane et al., 2014; Nisula, 2018). In the case of these two species the 4× higher phenolics content of larch, only lead to about half the weight loss within the first 2 and 8 weeks as compared to spruce. This suggests that lignans in spruce at higher concentrations could contribute to a similar or greater durability than larch flavonoids do, presumably through their antioxidative potentials. The work of Willför et al. (2003c) supports this suggestion, as they found that the performances of spruce and larch extracts, as well as isolated compounds thereof, were similar, but not equally good at scavenging different types of oxygen radicals, or inhibiting lipid peroxidation. We suggest that the antioxidative function of poly(phenolics) in heartwood could be especially relevant in the early-phase of degradation, where brown-rot fungi disrupt cell wall polymers by employing small radicals. Although there is no quantitative information on how efficient the cellulolytic enzymes of brow-rots could be without prior oxidative disruption of the cell wall, we find it unlikely that the fungus invests so much energy into a non-enzymatic oxidation phase, if it was not necessary. Apart from the moisture content of the wood, the amount and efficacy of antioxidants may play a significant role in the on-setting brown-rot attack, because they could potentially prolong the early-phase, as is the case with larch, showing a lower weight loss as compared to spruce in equal time. Ideally, it would even starve the fungus, because it cannot access the cell wall with its enzymes.

We found that the weight loss after 2 weeks was nearly the same for the PHI groups of both species, where spruce was not significantly affected compared to their native control, but both larch trees were less durable compared to their controls. This could be related to the presence of oligolignans in spruce, that may be harder to extract and have been shown to be good radical scavengers (Willför et al., 2003c). On the other hand, the larch specimen is also expected to contain polymeric material with similar activity. A more likely reason is that the removal of hydrophilic components did not affect our spruce samples as much, because the quantities of lignans were low, and are generally low in spruce heartwood. Therefor, spruce may simply have to rely on the fungicidal oleoresin more heavily, than larch does. In this scenario, the composition of the oleoresin is crucial, and known to differ among more and less susceptible trees (Holmbom et al., 2008; Mason et al., 2015).

We tentatively conclude, that hydrophobic extractives play a more important role in spruce than in larch, and vice versa. Note though, that we cannot exclude the possibility that the observed effects were caused by a relocalization of certain substances by the extraction, to places where they may not be as effective as they would have been in their native surroundings.

5. CONCLUSION

In order to investigate the influence of hydrophobic and hydrophilic extractives on water sorption and brown-rot

degradation, we prepared milled samples and sticks of Norway spruce and Kurile larch extracted with only hydrophobic or only hydrophilic solvents, as well as extracted with all four solvents (total extraction). The extraction efficiency was lowest for the ethanol fraction, indicating that ethanol-soluble analytes are harder to remove from the wood, which might be connected to their localization. For both species we found that most of the extracted material must be either polymeric or esterified in some way, because we only detected a fraction of the extracts with gas chromatography. This implies that the conclusions we draw could not consider much of the effect of the polymers in the extract, although it is known that they contribute to bulking of the cell wall. However, this effect was overall small as seen from the sorption isotherm we obtained for Kurile larch after different extractions, because of the incomplete extraction. We conclude that the changes observed are likely not responsible for the differences in degradation among the differently extracted samples.

The degradation test with European *R. placenta* resulted in measurable weight loss already after 2 weeks of incubation. We suggest that the more fungitoxic, hydrophobic substances found in spruce may play a bigger role in its protection, while on the other hand, larch durability is additionally boosted by its antioxidative flavonoid content, especially its taxifolin content. In this scenario, the hydrophilic lignans in spruce and hydrophobic extracts in larch can complement or back-up the function of the respective other part, especially in case of metabolic disturbances as likely seen with larch 1. Regarding non-toxic wood protection systems, this study indicates that large amounts of antioxidant material can provide some protection against the brown-rot *R. placenta*, but systems with multiple mechanisms of action are still preferrable.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

The research was initiated and designed by SF and LT. TB-N and SF did the GC-MS/GC-FID instrument setup and measurements. AS contributed significantly to the analysis of the GC-MS data and the manuscript revisions. MF and SF did the design and execution of the sorption isotherm experiments, interpretation of data together with LT. AP helped design and discuss the fungal degradation experiments and with manuscript revisions. SF did the extractions, data analysis and interpretation of all experiments. SF, TB-N, MF, and LT co-wrote the paper. All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00855/full#supplementary-material

REFERENCES

Andraos, J. (1996). On the propagation of statistical errors for a function of several variables. *J. Chem. Educ.* 73, 150–154. doi: 10.1021/ed073p150

Arantes, V., and Goodell, B. (2014). Current understanding of brown-rot fungal biodegradation mechanisms: a review. *ACS Symp. Ser.* 1158, 3–21. doi: 10.1021/bk-2014-1158.ch001

Bader, T. K., Hofstetter, K., Alfredsen, G., and Bollmus, S. (2012). Changes in microstructure and stiffness of Scots pine (*Pinus sylvestris* L) sapwood degraded by *Gloeophyllum trabeum* and *Trametes versicolor*–Part II: anisotropic stiffness properties. *Holzforschung* 66, 199–206. doi: 10.1515/HF.2011.153

Belt, T., Keplinger, T., Hänninen, T., and Rautkari, L. (2017). Cellular level distributions of Scots pine heartwood and knot heartwood extractives revealed by Raman spectroscopy imaging. *Indus. Crops Prod.* 108, 327–335. doi: 10.1016/j.indcrop.2017.06.056

Bergstedt, A., and Lyck, C. H. (eds.). (2007). *Larch Wood - a Literature Review. Forest & Landscape Working Papers no. 23-2007.* Copenhagen: Forest & Landscape.

Binbuga, N., Ruhs, C., Hasty, J. K., Henry, W. P., and Schultz, T. P. (2008). Developing environmentally benign and effective organic wood preservatives by understanding the biocidal and non-biocidal properties of extractives in naturally durable heartwood. *Holzforschung* 62, 264–269. doi: 10.1515/HF.2008.038

Boddy, L. (1983). Effect of temperature and water potential on growth rate of wood-rotting basidiomycetes. *Trans. Brit. Mycol. Soc.* 80, 141–149. doi: 10.1016/S0007-1536(83)80175-2

Bohlmann, J., Gershenzon, J., and Augbourg, S. (2000). "Biochemical, molecular genetic and evolutionary aspects of defense-related terpenoid metabolism in conifers," in *Evolution of Metabolic Pathways*, eds R. Ibrahim, L. Varin, V. De Luca, and J. Romeo (Elsevier), 109–150. doi: 10.1016/S0079-9920(00)80006-4

Brischke, C., and Alfredsen, G. (2020). Wood-water relationships and their role for wood susceptibility to fungal decay. *Appl. Microbiol. Biotechnol.* 104, 3781–3795. doi: 10.1007/s00253-020-10479-1

Brischke, C., Bayerbach, R., and Otto Rapp, A. (2006). Decay-influencing factors: A basis for service life prediction of wood and wood-based products. *Wood Mat. Sci. Eng.* 1, 91–107. doi: 10.1080/17480270601019658

Brischke, C., Soetbeer, A., and Meyer-Veltrup, L. (2017). The minimum moisture threshold for wood decay by basidiomycetes revisited. A review and modified pile experiments with Norway spruce and European beech decayed by

Coniophora puteana and *Trametes versicolor*. Holzforschung 71, 893–903. doi: 10.1515/hf-2017-0051

Burčová, Z., Kreps, F., Greifová, M., Jablonský, M., Ház, A., Schmidt, Š., et al. (2018). Antibacterial and antifungal activity of phytosterols and methyl dehydroabietate of Norway spruce bark extracts. *J. Biotechnol.* 282, 18–24. doi: 10.1016/j.jbiotec.2018.06.340

Côté, W. A., Day, A. C., Simson, B. W., and Timell, T. E. (1966). Studies on larch arabinogalactan. I. the distribution of arabinogalactan in larch wood. Holzforschung 20, 178–192. doi: 10.1515/hfsg.1966.20.6.178

Chen, P. S., Chen, Y. H., Yeh, T. F., and Chang, S. T. (2014). Mechanism of decay resistance of heartwood extracts from *Acacia confusa* against the brown-rot fungus *Laetiporus sulphureus*. *Wood Sci. Technol.* 48, 451–465. doi: 10.1007/s00226-014-0615-6

Choong, E., and Achmadi, S. (1991). Effects of extractives on moisture sorption and shrinkage in tropical woods. *Wood Fiber Sci.* 23, 185–196.

Cloutier, A., and Fortin, Y. (1994). Wood drying modelling based on the water potential concept: Hysteresis effects. *Drying Technol.* 12, 1793–1814. doi: 10.1080/07373939408962207

Defo, M., Fortin, Y., and Cloutier, A. (1999). Moisture content-water potential relationship of sugar maple and white spruce wood from green to dry conditions. *Wood Fiber Sci.* 31, 62–70.

Dellus, V., Mila, I., Scalbert, A., Menard, C., Michon, V., and Herve du Penhoat, C. L. (1997). Douglas-fir polyphenols and heartwood formation. *Phytochemistry* 45, 1573–1578. doi: 10.1016/S0031-9422(97)00245-8

Donoso-Fierro, C., Becerra, J., Bustos-Concha, E., and Silva, M. (2009). Chelating and antioxidant activity of lignans from Chilean woods (Cupressaceae). Holzforschung 63, 559–563. doi: 10.1515/HF.2009.123

Eberhardt, T. L., Young, R. A., Han, J. S., and Micales, J. A. (1994). Decay resistance in conifer seed cones: role of resin acids as inhibitors of decomposition by white-rot fungi. Holzforschung 48, 278–284. doi: 10.1515/hfsg.1994.48.4.278

ECHA (2020). *Wood Protection Chemicals (PT08)*.

Ekeberg, D., Flæte, P.-O., Eikenes, M., Fongen, M., and Naess-Andresen, C. F. (2006). Qualitative and quantitative determination of extractives in heartwood of Scots pine (*Pinus sylvestris* L.) by gas chromatography. *J. Chromatogr. A* 1109, 267–272. doi: 10.1016/j.chroma.2006.01.027

Ekman, R. (1980). *Wood extractives of norway spruce, a study of nonvolatile constituents and their effects on Fomes annosus* (Ph.D. thesis). Åbo Akademi University, Turku, Finland.

Engelund, E. T., Thygesen, L. G., Svensson, S., and Hill, C. A. (2013). A critical discussion of the physics of wood-water interactions. *Wood Sci. Technol.* 47, 141–161. doi: 10.1007/s00226-012-0514-7

Fackler, K., Stevanic, J. S., Ters, T., Hinterstoisser, B., Schwanninger, M., and Salmén, L. (2010). Localisation and characterisation of incipient brown-rot decay within spruce wood cell walls using FT-IR imaging microscopy. *Enzyme Microb. Technol.* 47, 257–267. doi: 10.1016/j.enzmictec.2010.07.009

Fang, W., Hemming, J., Reunanen, M., Eklund, P., Pineiro, E. C., Poljansek, I., et al. (2013). Evaluation of selective extraction methods for recovery of polyphenols from pine. Holzforschung 67, 843–851. doi: 10.1515/hf-2013-0002

Fedorova, T. E., Fedorov, S. V., and Babkin, V. A. (2015). Oligolignans in the wood of *Picea obovata* Ledeb. *Russ. J. Bioorgan. Chem.* 42, 712–715. doi: 10.1134/S1068162016070062

Fengel, D., and Wegener, G. (1989). *Wood, Chemistry, Ultrastucture, Reactions.* Berlin; New York, NY: de Gruyter,.

Fredriksson, M., and Johansson, P. (2016). A method for determination of absorption isotherms at high relative humidity levels: measurements on Lime-Silica Brick and Norway Spruce (*Picea abies* (L.) Karst.). *Dry. Technol.* 34, 132–141. doi: 10.1080/07373937.2015.1041035

Fredriksson, M., and Thybring, E. E. (2019). On sorption hysteresis in wood: separating hysteresis in cell wall water and capillary water in the full moisture range. *PLoS ONE* 14:e0225111. doi: 10.1371/journal.pone.0225111

Gierlinger, N., Jacques, D., Grabner, M., Wimmer, R., Schwanninger, M., Rozenberg, P., et al. (2003). Colour of larch heartwood and relationships to extractives and brown-rot decay resistance. *Trees Struct. Funct.* 18, 102–108. doi: 10.1007/s00468-003-0290-y

Gierlinger, N., Jacques, D., Schwanninger, M., Wimmer, R., and Pâques, L. E. (2004). Heartwood extractives and lignin content of different larch species (*Larix* sp.) and relationships to brown-rot decay-resistance. *Trees Struct. Funct.* 18, 230–236. doi: 10.1007/s00468-003-0300-0

Gierlinger, N., and Wimmer, R. (2004). Radial distribution of heartwood extractives and lignin in mature European larch. *Wood Fiber Sci.* 36, 387–394.

Giwa, S. A. O. (1973). *Extractives of western larch (Larix occidentalis nutt.)* (Ph.D. thesis). University of British Columbia, Vancouver, BC Canada.

Grabner, M., Müller, U., Gierlinger, N., and Wimmer, R. (2005a). Effects of heartwood extractives on mechanical properties of larch. *IAWA J.* 26, 211–220. doi: 10.1163/22941932-90000113

Grabner, M., Wimmer, R., Gierlinger, N., Evans, R., and Downes, G. (2005b). Heartwood extractives in larch and effects on x-ray densitometry. *Can. J. Forest Res.* 35, 2781–2786. doi: 10.1139/x05-196

Greenspan, L. (1977). Humidity fixed points of binary saturated aqueous solutions. *J. Res. Natl. Bureau Standards Sec A Phys. Chem.* 81A:89. doi: 10.6028/jres.081A.011

Griffin, D. M. (1977). Water potential and wood-decay fungi. *Annu. Rev. Phytopathol.* 15, 319–329. doi: 10.1146/annurev.py.15.090177.001535

Griffith, G. S., and Boddy, L. (1991). Fungal decomposition of attached angiosperm twigs: III. Effect of water potential and temperature on fungal growth, survival and decay of wood. *N. Phytol.* 117, 259–269. doi: 10.1111/j.1469-8137.1991.tb04907.x

Harju, A. M., Kainulainen, P., Venäläinen, M., Tiitta, M., and Viitanen, H. (2002). Differences in resin acid concentration between brown-rot resistant and susceptible scots pine heartwood. Holzforschung 56, 479–486. doi: 10.1515/HF.2002.074

Higuchi, T. (ed.). (1997). *Biochemistry and Molecular Biology of Wood.* Berlin; Heidelberg: Springer. doi: 10.1007/978-3-642-60469-0

Hill, C. A. S., Ramsay, J., and Gardiner, B. (2015). Variability in water vapour sorption isotherm in Japanese Larch (*Larix kaempferi* Lamb.)–earlywood and latewood influences. *Int. Wood Prod. J.* 6, 53–59. doi: 10.1179/2042645314Y.0000000090

Hillis, W. E. (1987). *Heartwood and Tree Exudates*, Vol. 4. Berlin; Heidelberg: Springer. doi: 10.1007/978-3-642-72534-0_2

Holmbom, B. (1999). "Extractives," in *Analytical Methods in Wood Chemistry, Pulping, and Papermaking*, eds E. Sjöström and R. Alén (Berlin; Heidelberg: Springer), 125–148. doi: 10.1007/978-3-662-03898-7_5

Holmbom, T., Reunanen, M., and Fardim, P. (2008). Composition of callus resin of Norway spruce, Scots pine, European larch and Douglas fir. Holzforschung 62, 417–422. doi: 10.1515/HF.2008.070

Hunt, C. G., Zelinka, S. L., Frihart, C. R., Lorenz, L., Yelle, D., Gleber, S. C., et al. (2018). Acetylation increases relative humidity threshold for ion transport in wood cell walls–A means to understanding decay resistance. *Int. Biodeter. Biodegr.* 133, 230–237. doi: 10.1016/j.ibiod.2018.06.014

Jebrane, M., Pockrandt, M., and Terziev, N. (2014). Natural durability of selected larch and Scots pine heartwoods in laboratory and field tests. *Int. Biodeter. Biodegr.* 91, 88–96. doi: 10.1016/j.ibiod.2014.03.018

Keeling, C. I., and Bohlmann, J. (2006). Diterpene resin acids in conifers. *Phytochemistry* 67, 2415–2423. doi: 10.1016/j.phytochem.2006.08.019

Kutnik, M., Klamer, M., Melcher, E., Kutnik, M., Klamer, M., and Melcher, E. (2017). "Ensuring quality of treated wood–regulations, certifications and associative background in the field of wood protection in Europe," in *Proceedings IRG Annual Meeting* (Stockholm: The International Research Group on Wood Protection).

Lindström, H. (2002). Intra-tree models of juvenile wood in Norway spruce as an input to simulation software. *Silva Fennica* 36, 521–534. doi: 10.14214/sf.542

Luostarinen, K., and Heräjärvi, H. (2013). Relation of arabinogalactans to density, growth rate and shear strength in wood of cultivated Siberian larch. *Eur. J. Wood Prod.* 71, 29–36. doi: 10.1007/s00107-012-0651-6

Mason, C. J., Klepzig, K. D., Kopper, B. J., Kersten, P. J., Illman, B. L., and Raffa, K. F. (2015). Contrasting patterns of diterpene acid induction by red pine and white spruce to simulated bark beetle attack, and interspecific differences in sensitivity among fungal associates. *J. Chem. Ecol.* 41, 524–532. doi: 10.1007/s10886-015-0588-4

Metsä-Kortelainen, S., and Viitanen, H. (2009). Decay resistance of sapwood and heartwood of untreated and thermally modified Scots pine and Norway spruce

compared with some other wood species. *Wood Mater. Sci. Eng.* 4, 105–114. doi: 10.1080/17480270903326140

Meyer, L., and Brischke, C. (2015). Fungal decay at different moisture levels of selected European-grown wood species. *Int. Biodeter. Biodegr.* 103, 23–29. doi: 10.1016/j.ibiod.2015.04.009

Meyer-Veltrup, L., Brischke, C., Alfredsen, G., Humar, M., Flæte, P. O., Isaksson, T., et al. (2017). The combined effect of wetting ability and durability on outdoor performance of wood: development and verification of a new prediction approach. *Wood Sci. Technol.* 51, 615–637. doi: 10.1007/s00226-017-0893-x

Micales, J. A., Han, J. S., Davis, J. L., and Young, R. A. (1994). "Chemical composition and fungitoxic activities of pine cone extractives," in *Biodeterioration Research 4: Mycotoxins, Wood Decay, Plant Stress, Biocorrosion, and General Biodeterioration*, eds G. C. Llewellyn, W. V. Dashek, and C. E. O'Rear (Boston, MA: Springer), 317–332. doi: 10.1007/978-1-4757-9450-2_25

Mozurkewich, M. (1993). The dissociation constant of ammonium nitrate and its dependence on temperature, relative humidity and particle size. *Atmos. Environ. A Gen. Top.* 27, 261–270. doi: 10.1016/0960-1686(93)90356-4

Nerg, A. M., Heijjari, J., Noldt, U., Viitanen, H., Vuorinen, M., Kainulainen, P., et al. (2004). Significance of wood terpenoids in the resistance of scots pine provenances against the old house borer, Hylotrupes bajulus, and brown-rot fungus, *Coniophora puteana. J. Chem. Ecol.* 30, 125–141. doi: 10.1023/B:JOEC.0000013186.75496.68

Nisula, L. (2018). *Wood extractives in conifers a study of stemwood and knots of industrially important species* (Ph.D. thesis). Åbo Akademi University, Turku, Finland.

Nzokou, P., and Kamdem, D. P. (2004). Influence of wood extractives on moisture sorption and wettability of Red Oak (*Quercus rubra*), Black Cherry (*Prunus serotina*), and Red Pine (*Pinus resinosa*). *Wood Fiber Sci.* 36, 483–492.

Örså, F., and Holmbom, B. R. (1994). A convenient method for the determination of wood extractives in papermaking process waters and effluents. *J. Pulp Pap. Sci.* 20, 361–366.

Ostroukhova, L. A., Raldugin, V. A., Babkin, V. A., Onuchina, N. A., and Levchuk, A. A. (2012). Investigation of the chemical composition of larch wood resin. *Russ. J. Bioorgan. Chem.* 38, 775–779. doi: 10.1134/S10681620 12070151

Otto, A., and Wilde, V. (2001). Sesqui-, di-, and triterpenoids as chemosystematic markers in extant conifers-a review. *Bot. Rev.* 67, 141–238. doi: 10.1007/BF02858076

Rice-Evans, C., Miller, N., and Paganga, G. (1997). Antioxidant properties of phenolic compounds. *Trends Plant Sci.* 2, 152–159. doi: 10.1016/S1360-1385(97)01018-2

Rice-Evans, C. A., Miller, N. J., and Paganga, G. (1996). Structure-antioxidant activity relationships of flavonoids and phenolic acids. *Free Radic. Biol. Med.* 20, 933–956. doi: 10.1016/0891-5849(95)02227-9

Ringman, R., Pilgard, A., Brischke, C., and Richter, K. (2014). Mode of action of brown rot decay resistance in modified wood: a review. *Holzforschung* 68, 239–246. doi: 10.1515/hf-2013-0057

Rowe, W. J. (ed.). (1989). *Natural Products of Woody Plants I - Chemicals Extraneous to the Lignocellulosic Cell Wall.* Berlin; Heidelberg: Springer.

Saarinen, N. M., Warri, A., Makela, S. I., Eckerman, C., Reunanen, M., Ahotupa, M., et al. (2000). Hydroxymatairesinol, a novel enterolactone precursor with antitumor properties from coniferous tree (*Picea abies*). *Nutr. Cancer* 36, 207–216. doi: 10.1207/S15327914NC3602_10

Schaller, A. (ed.). (2008). *Induced Plant Resistance to Herbivory.* Stuttgart: Springer Science+Business Media B.V. doi: 10.1007/978-1-4020-8182-8

Scheffer, T. C., and Cowling, E. B. (1966). Natural resistance of wood to microbial deterioration. *Annu. Rev. Phytopathol.* 4, 147–168. doi: 10.1146/annurev.py.04.090166.001051

Scheffer, T. C., and Morrell, J. J. (1998). *Natural Durability of Wood: A Worldwide Checklist of Species.* Technical Report 22, Forest Research Laboratory, Oregon State University.

Schmidt, O. (2006). *Wood and Tree Fungi.* Berlin; Heidelberg: Springer.

Schultz, T. P., and Nicholas, D. D. (2002). Development of environmentally-benign wood preservatives based on the combination of organic biocides with antioxidants and metal chelators. *Phytochemistry* 61, 555–560. doi: 10.1016/S0031-9422(02)00267-4

Singh, T., and Singh, A. P. (2012). A review on natural products as wood protectant. *Wood Sci. Technol.* 46, 851–870. doi: 10.1007/s00226-011-0448-5

Sjökvist, T., Wålinder, M. E., and Blom (2018). Liquid sorption characterisation of Norway spruce heartwood and sapwood using a multicycle Wilhelmy plate method. *Int. Wood Prod. J.* 9, 58–65. doi: 10.1080/20426445.2018. 1467602

Smeds, A. I., Eklund, P. C., and Willför, S. M. (2016). Chemical characterization of high-molar-mass fractions in a Norway spruce knotwood ethanol extract. *Phytochemistry* 130, 207–217. doi: 10.1016/j.phytochem.2016.05.006

Smeds, A. I., Eklund, P. C., and Willför, S. M. (2018). Characterization of high-molar-mass fractions in a Scots pine (*Pinus sylvestris* L.) knotwood ethanol extract. *Holzforschung* 72, 201–213. doi: 10.1515/hf-2017-0079

Smith, A. L., Campbell, C. L., Walker, D. B., and Hanover, J. W. (1989). Extracts from black locust as wood preservatives: extraction of decay resistance from black locust heartwood. *Holzforschung* 43, 293–296. doi: 10.1515/hfsg.1989.43.5.293

Thybring, E. (2017). Water relations in untreated and modified wood under brown-rot and white-rot decay. *Int. Biodeter. Biodegr.* 118, 134–142. doi: 10.1016/j.ibiod.2017.01.034

Vahtikari, K., Rautkari, L., Noponen, T., Lillqvist, K., and Hughes, M. (2017). The influence of extractives on the sorption characteristics of Scots pine (*Pinus sylvestris* L.). *J. Mater. Sci.* 52, 10840–10852. doi: 10.1007/s10853-017-1278-0

Venäläinen, M., Harju, A. M., Terziev, N., Laakso, T., and Saranpä, P. (2006). Decay resistance, extractive content, and water sorption capacity of Siberian larch (*Larix sibirica* Lebed.) heartwood timber. *Holzforschung* 60, 99–103. doi: 10.1515/HF.2006.017

Wagner, L., Bader, T. K., Ters, T., Fackler, K., and de Borst, K. (2015). A combined view on composition, molecular structure, and micromechanics of fungal degraded softwood. *Holzforschung* 69, 471–482. doi: 10.1515/hf-2014-0023

Wangaard, F. F., and Granados, L. A. (1967). The effect of extractives on water-vapor sorption by wood. *Wood Sci. Technol.* 1, 253–277. doi: 10.1007/BF00349758

Willför, S., Hemming, J., and Leppanen, A.-S. (2006a). Analysis of extractives in different pulps–method development, evaluation, and recommendations. *Analytical Tools With Applications for Wood and Pulping Chemistry.* 1–17.

Willför, S., Hemming, J., Reunanen, M., Eckerman, C., and Holmbom, B. (2003a). Lignans and lipophilic extractives in Norway spruce knots and stemwood. *Holzforschung* 57, 27–36. doi: 10.1515/HF.2003.005

Willför, S., Hemming, J., Reunanen, M., and Holmbom, B. (2003b). Phenolic and lipophilic extractives in scots pine knots and stemwood. *Holzforschung* 57, 359–372. doi: 10.1515/HF.2003.054

Willför, S., and Holmbom, B. (2004). Isolation and characterisation of water soluble polysaccharides from Norway spruce and Scots pine. *Wood Sci. Technol.* 38, 173–179. doi: 10.1007/s00226-003-0200-x

Willför, S., Nisula, L., Hemming, J., Reunanen, M., and Holmbom, B. (2004a). Bioactive phenolic substances in industrially important tree species. Part 1: knots and stemwood of different spruce species. *Holzforschung* 58, 334–344. doi: 10.1515/HF.2004.052

Willför, S., Reunanen, M., Eklund, P., Sjöholm, R., Kronberg, L., Fardim, P., et al. (2004b). Oligolignans in Norway spruce and Scots pine knots and Norway spruce stemwood. *Holzforschung* 58, 345–354. doi: 10.1515/HF. 2004.053

Willför, S. M., Ahotupa, M. O., Hemming, J. E., Reunanen, M. H. T., Eklund, P. C., Sjöholm, R. E., et al. (2003c). Antioxidant activity of knotwood extractives and phenolic compounds of selected tree species. *J. Agric. Food Chem.* 51, 7600–7606. doi: 10.1021/jf030445h

Willför, S. M., Smeds, A. I., and Holmbom, B. R. (2006b). Chromatographic analysis of lignans. *J. Chromatogr. A* 1112, 64–77. doi: 10.1016/j.chroma.2005.11.054

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126, 485–493. doi: 10.1104/pp.126.2.485

Zelinka, S. L., Gleber, S. C., Vogt, S., López, G. M., and Jakes, J. E. (2015). Threshold for ion movements in wood cell walls below fiber saturation observed by X-ray fluorescence microscopy (XFM). *Holzforschung* 69, 441–448. doi: 10.1515/hf-2014-0138

Zhang, J., Presley, G. N., Hammel, K. E., Ryu, J.-S., Menke, J. R., Figueroa, M., et al. (2016). Localizing gene regulation reveals a staggered wood decay mechanism for the brown rot fungus *Postia placenta*. *Proc. Natl. Acad. Sci. U.S.A.* 113:201608454. doi: 10.1073/pnas.1608 454113

Zule, J., Čufar, K., and Tišler, V. (2015). Lipophilic extractives in heartwood of European larch (*Larix decidua* Mill.). *Drvna Industrija* 66, 305–313. doi: 10.5552/drind.2015.1442

Zule, J., Čufar, K., and Tišler, V. (2017). Hydrophilic extractives in heartwood of European larch (*Larix decidua* Mill.). *Drvna Industrija* 67, 363–370. doi: 10.5552/drind.2016.1618

Variation of Secondary Metabolite Profile of *Zataria multiflora* Boiss. Populations Linked to Geographic, Climatic, and Edaphic Factors

Ali Karimi[1,2]*, Andrea Krähmer[1], Nadine Herwig[1], Hartwig Schulz[1], Javad Hadian[3] and Torsten Meiners[1]*

[1] Institute for Ecological Chemistry, Plant Analysis and Stored Product Protection, Julius Kühn Institute, Berlin, Germany, [2] Institute of Pharmacy, Freie Universität Berlin, Berlin, Germany, [3] Department of Agriculture, Medicinal Plants and Drug Research Institute, Shahid Beheshti University, Tehran, Iran

Geographic location and connected environmental and edaphic factors like temperature, rainfall, soil type, and composition influence the presence and the total content of specific plant compounds as well as the presence of a certain chemotype. This study evaluated whether geographic, edaphic, and climatic information can be utilized to predict the presence of specific compounds from medicinal or aromatic plants. Furthermore, we tested rapid analytical methods based on near infrared spectroscopy (NIR) coupled with gas chromatography/flame ionization (GC/FID) and gas chromatography/mass spectrometry (GC/MS) analytical methods for characterization and classification metabolite profiling of *Zataria multiflora* Boiss. populations. *Z. multiflora* is an aromatic, perennial plant with interesting pharmacological and biological properties. It is widely dispersed in Iran as well as in Pakistan and Afghanistan. Here, we studied the effect of environmental factors on essential oil (EO) content and the composition and distribution of chemotypes. Our results indicate that this species grows predominantly in areas rich in calcium, iron, potassium, and aluminum, with mean rainfall of 40.46 to 302.72 mm·year^{-1} and mean annual temperature of 14.90°C to 28.80°C. EO content ranged from 2.75% to 5.89%. Carvacrol (10.56–73.31%), thymol (3.51–48.12%), linalool (0.90–55.38%), and *p*-cymene (1.66–13.96%) were the major constituents, which classified 14 populations into three chemotypes. Corresponding to the phytochemical cluster analysis, the hierarchical cluster analysis (HCA) based on NIR data also recognized the carvacrol, thymol, and linalool chemotypes. Hence, NIR has the potential to be applied as a useful tool to determine rapidly the chemotypes of *Z. multiflora* and similar herbs. EO and EO constituent content correlated with different geographic location, climate, and edaphic factors. The structural equation models (SEMs) approach revealed direct effects of soil factors (texture, phosphor, pH) and mostly indirect effects of latitude and altitude directly

affecting, e.g., soil factors. Our approach of identifying environmental predictors for EO content, chemotype or presence of high amounts of specific compounds can help to select regions for sampling plant material with the desired chemical profile for direct use or for breeding.

INTRODUCTION

All over the world, plants face different local climatic regimes as well as different edaphic factors. To predict how different environmental factors affect species dispersal, the abundance of populations and chemotypes as well as the content of specific compounds can be a valuable tool to understand plant variation in chemical features. It can also facilitate prospecting plants with high amounts of specific compounds for nutrition, pharmaceutical or agricultural use. In most cases, plant essential oils (EOs) are characterized by a strong aroma, which is mainly produced by secondary metabolites. EO compounds are coupled with environmental acclimatization and play vital biological roles. Several factors, such as environmental and edaphic conditions, geographical regions, season of collection, harvesting time, genotype, and ecotype influence the quantitative and qualitative composition of EO (Milos et al., 2001; Zgheib et al., 2016; Morshedloo et al., 2018). For example, in *Matricaria chamomilla* L. climatic conditions, altitude, soil properties, and irrigation influence the phytochemical composition and antioxidant activity of EO (Formisano et al., 2015).

Zataria multiflora Boiss. (Lamiaceae) is an aromatic and perennial shrub growing wild in Iran (**Figure 1A**), Pakistan, and Afghanistan. This aromatic plant is known by the Persian name of Avishan Shirazi which is also entitled Sattar or Zattar, meaning thyme. *Z. multiflora* can be identified by the orbicular, densely gland-dotted, grey-green ovate leaves, and the thickly white hairy round buds in the leaf axils. Its inflorescence is verticillate, and the flowers are very small and white (Simbar et al., 2008). *Z. multiflora* has shown pharmacological (antimicrobial, antinociceptive, spasmolytic, and anti-inflammatory) properties, is utilized in traditional folk remedies for its antiseptic, analgesic, carminative, anthelmintic, and antidiarrheal properties, and it is also a condiment (Iranian Herbal Pharmacopoeia Committee, 2002; Moazeni et al., 2014; Khazdair et al., 2018; Mohajeri et al., 2018). Currently, some pharmaceutical forms of this plant, such as syrups, oral drops, soft capsules, and vaginal creams are produced (Sajed et al., 2013; Mahboubi, 2019).

The EO of *Z. multiflora* is rich in phenolic oxygenated monoterpenes. The main chemical constituents are carvacrol, thymol, linalool, and *p*-cymene (Hadian et al., 2011a; Saedi Dezaki et al., 2016; Mahmoudvand et al., 2017). Although there are some studies based on *Z. multiflora* EO constituents (Saleem et al., 2004; Niczad et al., 2019), there is hardly any information on the environmental factors affecting EO content and composition. *Z. multiflora* is not only harvested for local

markets but is also one of the valuable species for industry, so this plant is under severe threat from overharvesting. Thus, a deep perception of its phytochemical and environmental characteristics in its natural habitats is crucial to foretell its behavior under man-made cultivation.

Today, the standard method for EO analysis is gas chromatography coupled with different detection techniques like mass spectrometry. In the last two decades, numerous vibrational spectroscopy methods including mid-infrared (IR), near-infrared (NIR), and Raman spectroscopy have been described as a useful tool to examine the plant secondary metabolites which are commonly applied in the chemical fingerprinting of plants (Schulz et al., 2004; Schulz et al., 2005; Gudi et al., 2014). However, up to now, no studies have been performed utilizing this capable approach to differentiate and characterize various *Z. multiflora* chemotypes.

The aim of this study was to evaluate how different environmental factors affect species dispersal with respect to EO production, chemotype as well as the content of specific compounds of *Z. multiflora* population (**Figures 1A, B**). Besides, we aimed to evaluate whether geographic, edaphic, and climatic information can predict the presence of specific compounds. Furthermore, we tested rapid analytical methods based on NIRS coupled with GC/GC-MS methods for characterization and classification metabolite profiling of *Z. multiflora* populations.

MATERIALS AND METHODS

Study Area
To determine the effects of geography, climate, and edaphic conditions on EO yield and composition of *Z. multiflora*, plant materials were collected in 2018 in 14 natural habitats across five provinces from the center to the south of Iran including their major growing areas Isfahan, Kerman, Yazd, Fars, and Hormozgan provinces (**Figure 1A**).

Plant Material and Chemicals
Plant samples were collected in June 2018 at the flowering stage. At each region, 6 to 11 individual shrubs were collected depending on the population size with a minimum distance of 100 m. Voucher specimens (no. MPH-1799) were authenticated and deposited in the Herbarium of Medicinal Plants and Drugs Research Institute (MPH), Shahid Beheshti University, Tehran, Iran. Geographical data and altitude for each sampling area were recorded using GPS (**Table 1**). Besides, climate data for five years were taken from metrological stations closest to the habitats

FIGURE 1 | Collection sites **(A)** and overview on geographic, climatic, and edaphic factors **(B)** affecting *Zataria multiflora* populations from Iran.
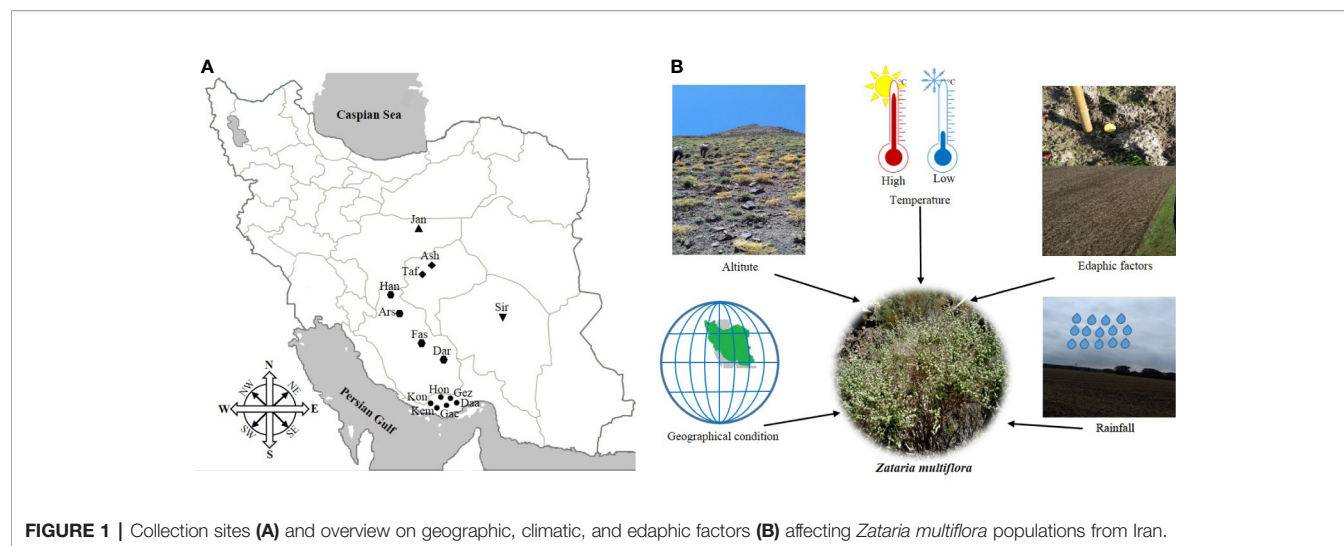
TABLE 1 | General information on natural habitats of *Zataria multiflora* populations.

Population name	Code	Province	Location	Latitude (N)	Longitude (E)	Altitude (m)
Jandaq	Jan	Esfahan	Jandaq toward mesr desert	33° 57' 44''	54° 31' 02''	1235
Ashkezar	Ash	Yazd	Zarband village	31° 48' 49''	54° 00' 26''	1946
Taft	Taf	Yazd	Darreh-ye Gahan mountains	31° 42' 26''	54° 10'	1697
Siriz	Sir	Kerman	Hamsij village	30° 55' 43''	55° 57' 01''	1763
Fasa	Fas	Fars	Kohankouye village	28° 59' 27''	53° 42' 25''	1516
Arsenjan	Ars	Fars	Tange laykhare mountains	29° 53' 49''	53° 16' 20''	1865
Haneshk	Han	Fars	Haneshk village, Safashahr	30° 49' 16''	53° 18' 19''	1898
Darab	Dar	Fars	Tange Talar Jangi mountains	28° 44' 27''	54° 34' 41''	1276
Gezeh	Gez	Hormozgan	Cheshmeh-ye seyyed	27° 06' 35''	54° 04' 46''	731
Hongooyeh	Hon	Hormozgan	Darreh-ye Baraveh	27° 06' 19''	54° 04' 07''	820
Daarbast	Daa	Hormozgan	Daarbast	26° 58' 02''	54° 01' 59''	1009
Gachooyeh	Gac	Hormozgan	Gachooyeh	26° 58' 28''	53° 58' 06''	1055
Kemeshk	Kem	Hormozgan	Kemeshk	27° 03' 13''	53° 50' 41''	937
Konar Siah	Kon	Hormozgan	Konar Siah	27° 09' 05''	53° 57' 04''	981

(**Table 2**). Carvacrol, linalool, *p*-cymene, and γ-terpinene were purchased from Sigma-Aldrich-Fluka (Germany), and thymol and α-pinene from Roth (Germany).

Soil Analysis

Soil samples from the surface layer (0 to 30 cm depth) were taken from five randomly selected plots in each sampling site. The five soil samples were combined into a single 500 g sample that was dried at room temperature (20–25°C) and sieved to 2 mm. A duplicate soil sample was sieved through a 2 mm filter once again for determination of soil chemical characteristics including the soil texture (percentage content of sand, silt, and clay), the amount of abundant nutrients (N, P, K, Ca, Al, and Fe), pH value, and organic matter. The total heavy metal and nutrient contents of soil samples were determined after pressure dissolution with 69% supra pure nitric acid (according to A2.4.3.1, VDLUFA, 1991) by ICP-AES (iCAP™ 7600 Duo, Thermo Fischer Scientific). Contents of total carbon and total nitrogen were determined with CNS elemental analyzer (Vario EL Cube, Elementar Analysesysteme GmbH). Pedological base parameters (soil particle size, pH value, C/N) were collected for

characterization. The particle size determination of soil texture was performed according to DIN 19683-2 (1997).

Isolation of the Essential Oils

The aerial plant parts were dried at room temperature (20–25°C) in the shade, then the leaves of each plant were separated and 10 g of each plant sample were ground manually. The EO of each sampled plant (10 g of leaves) was isolated by hydro-distillation for 2 h utilizing a clevenger-type system (Pavela et al., 2018). The distilled oils were dried over anhydrous sodium sulfate and stored at 4°C in sealed glass vials for analysis. The yield of the essential oil was calculated based on the dry weight of the plant material.

GC-FID and GC/MS Analyses

EOs were analyzed by GC–FID using an Agilent gas chromatograph 6890N, equipped with a HP-5 column (30 m × 0. 25 mm i.d., with a film thickness of 0.5 μm). The oven temperature was programmed at 50°C for 2 min, then from 50°C to 320°C at 5°C min^{-1}, and held at 320°C for 6 min. Both injector and detector temperatures were 250°C. Hydrogen was

TABLE 2 | Edaphic factors and climatic characteristics in natural habitats of *Zataria multiflora*.

Region	Rainfall [mm year⁻¹]	M-Temp [°C]	Soil texture	pH	OM	N	P	K	Ca	F	Al
					[%]				[mg kg⁻¹]		
Jandaq	55.86	21.50	Silty sand	7.70	4.43	0.06	394.30	5581.40	69376.20	16424.20	17012.60
Ashkezar	40.46	21.10	Loamy sand	7.80	5.17	0.04	409.70	7171.40	155569.40	14486.10	20489.70
Taft	49.76	20.30	Sandy loamy silt	7.70	5.74	0.04	354.90	7925.10	112373.80	19637.10	20137.30
Siriz	107.80	20.20	Sandy loam	7.70	4.44	0.03	416.90	6863.10	75182.10	20339.10	20521.30
Fasa	278.48	20.30	Clayey loam	7.60	6.97	0.06	329.20	8441.60	96355.40	23471.60	27656.70
Arsenjan	215.22	20.30	Sandy loamy silt	7.70	7.54	0.18	346.20	6151.10	117935.10	20298.30	20936.10
Haneshk	180.06	14.90	Loamy sand	7.60	4.00	0.09	342.60	12427.10	12756.10	24538.40	34280.10
Darab	276.38	24.30	Silty loam	7.60	10.0	0.23	312.60	5306.80	116218.30	16820.20	18424.30
Gezeh	302.72	28.80	Sandy loam	7.80	7.02	0.05	293.40	3609.80	174266.60	11979.30	12212.20
Hongooyeh	302.72	28.80	Silty loamy sand	7.90	7.54	0.03	355.90	3640.50	139953.40	14064.00	12378.00
Daarbast	302.72	28.80	Sandy loam	7.80	6.59	0.03	232.80	1993.80	197820.30	8190.60	6945.90
Gachooyeh	302.72	28.80	Loamy sand	7.80	7.02	0.03	222.50	2368.80	195214.40	9272.60	7986.60
Konar Siah	302.72	28.80	Loamy sand	7.80	7.02	0.05	275.80	2272.10	178396.30	7442.80	6327.50
Kemeshk	302.72	28.80	Loamy sand	7.80	7.02	0.04	249.20	2320.40	186805.30	8357.60	7157.10

M-Temp, Mean annual temperature; OM, organic matter.

used as carrier gas with a constant flow rate of 1 ml min⁻¹, and 1 µl of the diluted EOs (1/500 v/v in isooctane) was injected automatically (Gerstel MPS) in a splitless mode. Nitrogen was used as make-up gas, which was set at a flow of 45 ml min⁻¹.

Mass spectrometry of the EOs was performed using an Agilent MSD 5975B/GC 6890N, equipped with a 30 m × 0.25 mm i.d., 0.5 µm, HP-5MS column. The injector temperature was 250°C, and the initial GC oven temperature was 50°C, held for 2 min, then raised to 320°C at 5°C min⁻¹ and held for 6 min. Helium was used as carrier gas with a flow rate of 1 ml min⁻¹. One µl of the diluted EOs (1/500 v/v in isooctane) was injected automatically (Gerstel MPS) in a splitless mode. Injector and detector temperatures were set at 250°C. The EI⁺-MS operating parameters were as follows: ionization energy, 70 eV and ion source temperature, 230°C. The quadrupole mass spectrometer was scanned over 35 to 350 *m/z*. The runtime and solvent delay were set at 60 and 5 min, respectively (4.45 scans/s). Carvacrol, thymol, linalool, *p*-cymene, γ-terpinene, and α-pinene were used as standard. 6-Methyl-5-hepten-2-one was used as internal standard and was added to the dilution before the analysis. The oil components were identified by comparison of mass spectra and retention indices with those recorded in the Adams (Adams, 2014), NIST mass spectral databases SRD 69 (NIST Chemistry WebBook, 2002), standard constituents, and the previously published data. The retention indices of individual components were calculated using a series of n-alkanes (C8-C40) (Sigma-Aldrich-Fluka, Germany) (1/100 in n-Pentan). The relative percentage composition of individual compounds was computed from the GC peak areas obtained without using correction factors.

NIR Spectroscopy and Chemometrics

Before isolation of EO, vibrational spectroscopy was performed directly on the homogenized plant material. NIRS analyses were carried out on a Fourier-Transform (FT)-NIR spectrometer (Multi-Purpose Analyser MPA, Bruker Optics GmbH, Germany). Spectra were recorded in the wavenumber range of 4,000 to 12,000 cm⁻¹ with a spectral resolution of 8 cm⁻¹. Approximately 7 g of dried

leaves were put in a glass Petri dish and spectra were collected during rotation of the dish using the integrating sphere for measuring in diffuse reflection. Spectra were acquired at 30 s. Each sample was analyzed with threefold repetition. The raw spectra were centered and corrected for scattering effects and baseline shifting using WMSC of the OPUS 6.5 software (Bruker Optics). Only averaged spectra of the three replicates were used for the later chemometric analysis.

Statistical Analysis

Statistical analysis was performed using hierarchical cluster analysis (HCA) with SPSS version 16 to classify and cluster the populations of *Z. multiflora* based on the squared Euclidean distances. Pearson's correlation coefficients were estimated among the EO content, major components, and edaphic factors using SPSS (SPSS, Chicago, IL, USA) software package from version 16. The calculation of means, standard deviations (SD) and t-test were used to express the significance of differences (P < 0.05) using SAS 9.1 program (SAS Inc. USA).

For chemometrics (based on NIR), HCA was performed to evaluate the diversity of the samples. Characteristic spectral ranges were identified by comparison with spectra appropriate reference standards and HCA. Calibration models were built by 10-fold cross-validation using a partial least squares (PLS) algorithm. Therefore, GC data of each plant and averaged plant wise spectra of the population were correlated.

Furthermore, we set up SEMs for each region using partial least squares (PLS) regression using Warp PLS 6.0 (Kock and Lynn, 2012). The PLS regression was chosen over covariance based approaches because it suited our small sample size and, compared to covariance structure analysis, can accommodate both reflective and formative scales more easily. Moreover, PLS does not require any *a priori* distributional assumptions (Chin and Newsted, 1999). We present individual standardized path coefficients (β), partial model fit scores (R²), and overall model P values calculated by resampling estimations coupled with Bonferroni like corrections (Kock, 2010). To validate the models three model-fit indices [average path coefficient (APC),

average R-squared (ARS), and average variance inflation factor (AVIF)] were calculated for each region. For model fit, it is recommended that P values for APC and ARS are both lower than 0.05 (i.e., significance at the 0.05 level). The AVIF index controls for multicollinearity and should be below 5 (Kock, 2010). In the SEM analysis we set paths from geographic factors (latitude, longitude, altitude), climatic factors (rainfall, temperature), soil texture (relative proportion of clay, silt, and sand), constituents (N, P, K, Al, Ca, Fe), and pH value directly to EO content and compounds; furthermore, we included the possible effects of the geographic factors on climatic and soil factors.

RESULTS

Phytochemical Analysis of Essential Oil

The EOs were obtained and analyzed by hydro-distillation and GC-FID/GC-MS respectively. There was a significant difference in EO content among the studied populations. The EO content ranged from 2.75 (for population Siriz) to 5.89% in dry matter (DM), (for population Konar Siah) (**Figure 2**). Fifty-six compounds were identified with significant differences between the populations (**Table 3**). The oils mainly consisted of carvacrol (10.56–73.31%), thymol (3.51–48.12%), linalool (0.90–55.38%), *p*-cymene (1.66–13.96%), γ-terpinene (0.99–6.28%), α-pinene (0.93–4.01%), carvacrol methyl ether (0.39–3.71%), myrcene (0.94–2.77%), *E*-caryophyllene (1.09–2.37%), and α-terpinene (0.39–1.61%).

The Pearson correlations indicated positive and negative significant correlations between phytochemical compounds. Carvacrol had been positively correlated with carvacrol acetate (r = 0.70), carvacrol methyl ether (r = 0.54), and negatively correlated with linalool (r = −0.69), thymol (r = −0.64) and limonene (r = −0.79) while thymol was in significant negative correlation with carvacrol (r = −0.64). Furthermore, linalool had a significant positive correlation with *E*-β-ocimene (r = 0.99), myrcene (r = 0.97), limonene (r = 0.72), *Z*-β-ocimene (r = 0.69) and a negative correlation with α-terpinene (r = −0.72), γ-terpinene (r = −0.70), carvacrol (r = −0.69), *p*-cymene (r = −0.61), carvacrol acetate (r = −0.56) and carvacrol methyl ether (r = −0.53).

To determine the degree of phytochemical variation, HCA based on the phytochemical profiles was performed (**Figure 3**). According to the major components, three chemotypes can be distinguished thus populations of *Z. multiflora* were divided into three main clusters. Cluster I consists of two populations (Siriz and Haneshk) characterized by higher content of linalool. Cluster II contains two populations (Fasa and Darab) which are characterized by higher amounts of thymol, carvacrol, *p*-cymene, and linalool. Cluster III contains ten populations including Jandaq, Ashkezar, Taft, Arsenjan, Gezeh, Hongooyeh, Daarbast, Gachooyeh, Konar Siah, and Kemeshk characterized by lower quantities of α-pinene, myrcene, α-terpinene, linalool, and carvacrol methyl ether and higher amounts of carvacrol, thymol, *p*-cymene, and γ-terpinene.

Environmental Characteristics

Geographical, climatic, and edaphic characteristics of *Z. multiflora* natural habitats are exhibited in **Tables 1** and **2**. Our results indicate that this species grows in areas characterized by a mean rainfall of 40.46 to 302.72 mm year^{-1} and mean annual temperature of 14.90°C to 28.80°C. The altitude ranges from 731 to 1946 m. The percentage of organic
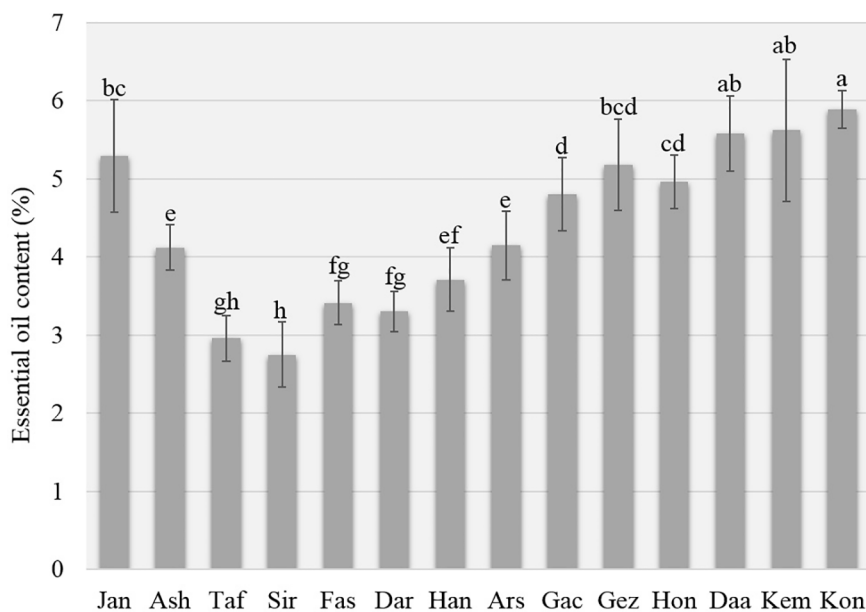

FIGURE 2 | Essential oil content of *Zataria multiflora* populations.

TABLE 3 | Variation of the phytochemical compositions (%) among the studied populations of *Zataria multiflora*.

N	Components	RI[a]	RI[b]	Populations					
				Jandaq	Ashkezar	Taft	Siriz	Fasa	Methods
1	α-Thujene	933	929	0.43 ± 0.06	0.18 ± 0.10	0.41 ± 0.03	0.14 ± 0.05	0.30 ± 0.03	RI, MS
2	α-Pinene	940	939	0.93 ± 0.28	1.82 ± 0.23	2.01 ± 0.14	1.85 ± 0.33	1.64 ± 0.18	RI, MS
3	Camphene	955	954	0.10 ± 0.03	0.12 ± 0.01	0.12 ± 0.01	0.10 ± 0.01	0.10 ± 0.01	RI, MS
4	2,4(10)-Thujadiene	960	957	–	–	–	tr	–	RI, MS
5	1,3-Octanol	969	973	0.07 ± 0.01	0.07 ± 0.01	0.07 ± 0.01	0.09 ± 0.01	0.08 ± 0.01	RI, MS
6	Sabinene	979	977	0.11 ± 0.01	0.07 ± 0.01	0.11 ± 0.01	0.09 ± 0.02	0.08 ± 0.01	RI, MS
7	β-Pinene	983	978	0.23 ± 0.03	0.28 ± 0.03	0.37 ± 0.01	0.34 ± 0.06	0.33 ± 0.01	RI, MS
8	Myrcene	995	992	1.35 ± 0.51	1.15 ± 0.16	1.39 ± 0.14	2.77 ± 0.51	1.45 ± 0.17	RI, MS
9	3-Octanol	997	995	0.05 ± 0.01	0.06 ± 0.01	0.10 ± 0.05	0.21 ± 0.04	0.10 ± 0.06	RI, MS
10	α-Phellandrene	1010	1005	0.15 ± 0.04	0.17 ± 0.01	0.19 ± 0.01	0.29 ± 0.03	0.21 ± 0.01	RI, MS
11	δ-3-Carene	1015	1011	0.02 ± 0.01	0.04 ± 0.02	0.02 ± 0.01	0.03 ± 0.01	–	RI, MS
12	α-Terpinene	1021	1019	0.92 ± 0.08	1.00 ± 0.07	1.38 ± 0.26	0.39 ± 0.14	1.61 ± 0.09	RI, MS
13	p-Cymene	1030	1025	5.88 ± 1.38	5.69 ± 0.31	7.29 ± 1.17	1.66 ± 0.95	7.21 ± 0.42	RI, MS
14	Limonene	1034	1030	0.37 ± 0.10	0.49 ± 0.07	0.54 ± 0.12	0.81 ± 0.07	0.61 ± 0.03	RI, MS
15	1,8-Cineole	1037	1035	0.24 ± 0.10	0.05 ± 0.01	0.08 ± 0.01	0.46 ± 0.10	0.03 ± 0.01	RI, MS
16	Z-β-Ocimene	1039	1039	0.28 ± 0.08	0.41 ± 0.04	0.44 ± 0.10	0.69 ± 0.21	0.37 ± 0.05	RI, MS
17	E-β-Ocimene	1050	1050	0.15 ± 0.10	0.07 ± 0.01	0.12 ± 0.04	1.31 ± 0.12	0.28 ± 0.05	RI, MS
18	γ-Terpinene	1063	1062	3.65 ± 0.72	3.14 ± 0.26	5.24 ± 1.12	0.99 ± 0.57	6.28 ± 0.40	RI, MS
19	Z-Sabinene hydrate	1071	1070	0.30 ± 0.01	0.14 ± 0.06	0.27 ± 0.03	0.08 ± 0.02	0.18 ± 0.01	RI, MS
20	Z-Linalool oxide	1076	1074	0.04 ± 0.01	0.02 ± 0.01	0.05 ± 0.01	1.67 ± 0.57	0.15 ± 0.03	RI, MS
21	Terpinolene	1092	1089	0.22 ± 0.09	0.21 ± 0.03	0.27 ± 0.10	1.62 ± 0.42	0.35 ± 0.03	RI, MS
22	Linalool	1103	1100	1.19 ± 0.23	1.62 ± 0.60	3.35 ± 2.20	55.38 ± 6.4	9.59 ± 1.83	RI, MS
23	E-γ-Caryophyllene	1106	1106	0.11 ± 0.04	0.04 ± 0.01	0.06 ± 0.04	1.37 ± 0.17	0.27 ± 0.05	RI, MS
24	1-Octenyl-3-acetate	1111	1113	–	–	–	0.20 ± 0.10	–	RI, MS
25	p-Menth-2-en-1-ol	1124	1122	–	–	–	0.32 ± 0.18	–	RI, MS
26	allo-Ocimene	1131	1132	–	–	–	0.16 ± 0.11	0.11 ± 0.02	RI, MS
27	1,3,8-p-Menthatriene	1133		–	–	–	0.12 ± 0.01	–	MS
28	Borneol	1172	1171	tr	–	0.02 ± 0.01	0.63 ± 0.18	0.21 ± 0.04	RI, MS
29	Z-Linalool oxide (pyranoid)	1177	1173	0.10 ± 0.01	0.02 ± 0.01	0.05 ± 0.02	0.10 ± 0.01	0.08 ± 0.01	RI, MS
30	4-Terpineol	1183	1179	0.09 ± 0.01	0.14 ± 0.03	0.14 ± 0.01	0.25 ± 0.03	0.15 ± 0.01	RI, MS
31	p-Cymenol-8	1188	1184	0.46 ± 0.04	0.46 ± 0.02	0.49 ± 0.04	0.23 ± 0.03	0.43 ± 0.02	RI, MS
32	α-Terpineol	1195	1190	0.44 ± 0.13	0.56 ± 0.05	0.56 ± 0.11	0.68 ± 0.05	0.45 ± 0.02	RI, MS
33	Z-Dihydro carvone	1202	1200	0.15 ± 0.01	0.17 ± 0.01	0.10 ± 0.03	0.15 ± 0.06	0.06 ± 0.04	RI, MS
34	Nerol	1234	1228	0.07 ± 0.02	0.06 ± 0.01	–	0.10 ± 0.02	–	RI, MS
35	Thymol methyl ether	1238	1237	–	0.08 ± 0.03	0.35 ± 0.19	0.20 ± 0.08	0.88 ± 0.06	RI, MS
36	Carvacrol methyl ether	1248	1241	1.04 ± 0.37	1.50 ± 0.40	1.00 ± 0.21	0.39 ± 0.10	0.49 ± 0.02	RI, MS
37	Geraniol	1260	1263	0.04 ± 0.01	–	0.06 ± 0.03	1.00 ± 0.43	0.11 ± 0.02	RI, MS
38	Geranial	1280	1273	–	–	–	tr	–	RI, MS
39	p-Thymol	1286		0.07 ± 0.03	0.08 ± 0.01	0.10 ± 0.02	0.02 ± 0.01	0.10 ± 0.01	MS
40	Thymol	1295	1295	3.51 ± 3.27	9.94 ± 2.03	25.32 ± 12.7	6.17 ± 3.00	48.12 ± 2.9	RI, MS
41	Carvacrol	1309	1305	73.31 ± 4.3	65.14 ± 1.8	42.23 ± 15.3	10.56 ± 3.9	12.42 ± 2.8	RI, MS
42	Thymol acetate	1358	1359	0.02 ± 0.01	0.10 ± 0.03	0.49 ± 0.27	0.10 ± 0.03	0.87 ± 0.07	RI, MS
43	Carvacrol acetate	1377	1368	0.80 ± 0.21	0.80 ± 0.13	0.60 ± 0.19	0.20 ± 0.07	0.16 ± 0.01	RI, MS
44	β-Bourbonene	1384	1378	–	–	–	0.09 ± 0.04	–	RI, MS
45	E-Caryophyllene	1435	1427	1.09 ± 0.22	1.61 ± 0.21	2.16 ± 0.42	1.56 ± 0.60	1.85 ± 0.20	RI, MS
46	Aromadendrene	1455	1436	0.29 ± 0.06	0.40 ± 0.07	0.37 ± 0.08	0.14 ± 0.10	0.19 ± 0.05	RI, MS
47	α-Humullene	1469	1452	0.10 ± 0.01	0.11 ± 0.01	0.14 ± 0.02	0.14 ± 0.04	0.13 ± 0.01	RI, MS
48	9-epi-(E)-Caryophyllene	1477	1474	–	–	–	–	–	RI, MS
49	E-β-Guaiene	1504	1498	–	–	–	0.02 ± 0.01	–	RI, MS
50	Viridiflorene	1510	1505	0.24 ± 0.10	0.26 ± 0.05	0.28 ± 0.08	0.19 ± 0.10	0.18 ± 0.02	RI, MS
51	Spathulenol	1594	1578	0.38 ± 0.09	0.35 ± 0.03	0.36 ± 0.06	0.59 ± 0.17	0.39 ± 0.02	RI, MS
52	Isoaromadendrene epoxide	1597		0.08 ± 0.01	0.11 ± 0.01	0.12 ± 0.02	0.47 ± 0.09	0.14 ± 0.02	MS
53	Caryophyllene oxide	1601	1599	0.18 ± 0.03	0.20 ± 0.01	0.20 ± 0.05	0.76 ± 0.11	0.21 ± 0.02	RI, MS
54	Caryophylla-4(12),8(13)-dien-5 beta-ol	1654	1663	0.07 ± 0.01	0.08 ± 0.01	0.09 ± 0.01	0.32 ± 0.03	0.11 ± 0.01	RI, MS
55	14-Hydroxy-9-epi-(E)-caryophyllene	1673	1669	0.03 ± 0.02	0.07 ± 0.02	0.08 ± 0.01	0.34 ± 0.06	0.09 ± 0.01	RI, MS
56	Khusinol	1686	1674	0.05 ± 0.03	0.13 ± 0.01	0.08 ± 0.01	0.27 ± 0.03	0.09 ± 0.01	RI, MS
	Monoterpene hydrocarbons			14.79	14.84	19.90	13.36	20.93	
	Oxygenated monoterpens			81.87	80.88	75.26	78.89	74.48	
	Sesquiterpene hydrocarbons			1.83	2.42	3.01	3.51	2.62	
	Oxygenated sesquiterpenes			0.80	0.94	0.93	2.75	1.03	
	Essential oil content (%)			5.29 ± 0.72	4.12 ± 0.29	2.96 ± 0.29	2.75 ± 0.42	3.41 ± 0.28	

(Continued)

TABLE 3 | Continued

N	Components	RI[a]	RI[b]	Populations					
				Gachooyeh	Kemeshk	Gezeh	Hongooyeh	Arsenjan	Methods
1	α-Thujene	933	929	0.29 ± 0.06	0.27 ± 0.11	0.47 ± 0.13	0.41 ± 0.09	0.33 ± 0.03	RI, MS
2	α-Pinene	940	939	2.57 ± 0.74	2.59 ± 0.73	4.01 ± 0.97	3.23 ± 1.15	1.43 ± 0.20	RI, MS
3	Camphene	955	954	0.14 ± 0.02	0.14 ± 0.02	0.11 ± 0.04	0.17 ± 0.04	0.10 ± 0.01	RI, MS
4	2,4(10)-Thujadiene	960	957	–	–	–	–	–	RI, MS
5	1,3-Octanol	969	973	0.07 ± 0.01	0.07 ± 0.01	0.07 ± 0.01	0.08 ± 0.01	0.07 ± 0.01	RI, MS
6	Sabinene	979	977	0.09 ± 0.01	0.09 ± 0.01	0.11 ± 0.01	0.11 ± 0.01	0.10 ± 0.01	RI, MS
7	β-Pinene	983	978	0.47 ± 0.05	0.48 ± 0.06	0.64 ± 0.11	0.57 ± 0.17	0.32 ± 0.02	RI, MS
8	Myrcene	995	992	1.11 ± 0.13	1.04 ± 0.18	1.12 ± 0.17	1.26 ± 0.13	1.32 ± 0.15	RI, MS
9	3-Octanol	997	995	0.13 ± 0.08	0.02 ± 0.01	0.10 ± 0.10	0.15 ± 0.13	0.04 ± 0.02	RI, MS
10	α-Phellandrene	1010	1005	0.16 ± 0.01	0.17 ± 0.01	0.17 ± 0.01	0.17 ± 0.01	0.19 ± 0.02	RI, MS
11	δ-3-Carene	1015	1011		0.02 ± 0.01	0.03 ± 0.02	–	0.03 ± 0.02	RI, MS
12	α-Terpinene	1021	1019	1.20 ± 0.14	1.34 ± 0.44	1.17 ± 0.08	1.32 ± 0.20	0.92 ± 0.05	RI, MS
13	p-Cymene	1030	1025	6.46 ± 2.47	5.50 ± 1.54	9.44 ± 1.18	10.27 ± 0.7	5.93 ± 0.48	RI, MS
14	Limonene	1034	1030	0.42 ± 0.07	0.41 ± 0.16	0.58 ± 0.21	0.67 ± 0.09	0.42 ± 0.09	RI, MS
15	1,8-Cineole	1037	1035	0.10 ± 0.06	0.11 ± 0.06	0.08 ± 0.05	0.02 ± 0.01	0.12 ± 0.07	RI, MS
16	Z-β-Ocimene	1039	1039	0.40 ± 0.05	0.32 ± 0.13	0.46 ± 0.25	0.55 ± 0.13	0.41 ± 0.08	RI, MS
17	E-β-Ocimene	1050	1050	0.04 ± 0.02	0.06 ± 0.02	0.07 ± 0.01	0.08 ± 0.01	0.20 ± 0.06	RI, MS
18	γ-Terpinene	1063	1062	5.45 ± 0.77	6.16 ± 2.30	5.08 ± 0.34	5.57 ± 0.99	3.49 ± 0.23	RI, MS
19	Z-Sabinene hydrate	1071	1070	0.21 ± 0.01	0.11 ± 0.04	0.21 ± 0.02	0.19 ± 0.02	0.26 ± 0.01	RI, MS
20	Z-Linalool oxide	1076	1074	–	–	–	tr	0.09 ± 0.05	RI, MS
21	Terpinolene	1092	1089	0.18 ± 0.01	0.18 ± 0.02	0.20 ± 0.02	0.21 ± 0.02	0.28 ± 0.04	RI, MS
22	Linalool	1103	1100	0.90 ± 0.17	1.01 ± 0.55	1.13 ± 0.22	1.28 ± 0.54	8.88 ± 3.17	RI, MS
23	E-γ-Caryophyllene	1106	1106	0.06 ± 0.02	0.05 ± 0.03	0.07 ± 0.01	0.06 ± 0.02	0.11 ± 0.04	RI, MS
24	1-Octenyl-3-acetate	1111	1113	–	0.02 ± 0.01	tr	0.02 ± 0.01	–	RI, MS
25	p-Menth-2-en-1-ol	1124	1122	–	–	–	–	–	RI, MS
26	Allo-Ocimene	1131	1132	–	–	–	–	0.09 ± 0.04	RI, MS
27	1,3,8-p-Menthatriene	1133		–	–	–	–	–	MS
28	Borneol	1172	1171	–	0.02 ± 0.01	–	–	0.24 ± 0.11	RI, MS
29	Z-Linalool oxide (pyranoid)	1177	1173	–	tr	tr	0.02 ± 0.01	tr	RI, MS
30	4-Terpineol	1183	1179	0.20 ± 0.02	0.20 ± 0.07	0.18 ± 0.02	0.20 ± 0.01	0.17 ± 0.02	RI, MS
31	p-Cymenol-8	1188	1184	0.48 ± 0.04	0.44 ± 0.02	0.50 ± 0.04	0.54 ± 0.03	0.44 ± 0.02	RI, MS
32	α-Terpineol	1195	1190	0.54 ± 0.06	0.47 ± 0.18	0.63 ± 0.30	0.73 ± 0.17	0.47 ± 0.07	RI, MS
33	Z-Dihydro carvone	1202	1200	0.14 ± 0.01	0.12 ± 0.06	0.16 ± 0.02	0.14 ± 0.03	0.13 ± 0.02	RI, MS
34	Nerol	1234	1228	–	tr	0.03 ± 0.02	–	0.06 ± 0.01	RI, MS
35	Thymol methyl ether	1238	1237	0.08 ± 0.01	0.05 ± 0.01	0.13 ± 0.07	0.15 ± 0.06	0.13 ± 0.07	RI, MS
36	Carvacrol methyl ether	1248	1241	1.94 ± 0.78	1.87 ± 1.19	3.58 ± 0.88	2.90 ± 1.67	1.84 ± 0.39	RI, MS
37	Geraniol	1260	1263	0.02 ± 0.02	–	–	0.03 ± 0.02	tr	RI, MS
38	Geranial	1280	1273	–	–	–	–	0.02 ± 0.01	RI, MS
39	p-Thymol	1286		0.10 ± 0.05	0.18 ± 0.10	0.15 ± 0.03	0.17 ± 0.01	0.07 ± 0.01	MS
40	Thymol	1295	1295	10.24 ± 6.6	7.34 ± 4.00	8.72 ± 1.97	14.67 ± 7.8	12.28 ± 2.2	RI, MS
41	Carvacrol	1309	1305	60.26 ± 10.2	64.22 ± 7.5	54.17 ± 6.3	48.16 ± 6.7	55.10 ± 3.1	RI, MS
42	Thymol acetate	1358	1359	0.09 ± 0.07	0.05 ± 0.03	0.19 ± 0.13	0.21 ± 0.06	0.08 ± 0.02	RI, MS
43	Carvacrol acetate	1377	1368	0.66 ± 0.13	0.58 ± 0.39	1.39 ± 0.10	0.92 ± 0.45	0.55 ± 0.06	RI, MS
44	β-Bourbonene	1384	1378	–	–	–	–	–	RI, MS
45	E-Caryophyllene	1435	1427	1.95 ± 0.40	2.37 ± 0.40	1.85 ± 0.30	1.64 ± 0.29	1.38 ± 0.25	RI, MS
46	Aromadendrene	1455	1436	0.43 ± 0.09	0.29 ± 0.04	0.49 ± 0.05	0.40 ± 0.08	0.20 ± 0.05	RI, MS
47	α-Humullene	1469	1452	0.15 ± 0.02	0.15 ± 0.02	0.16 ± 0.01	0.15 ± 0.02	0.10 ± 0.01	RI, MS
48	9-epi-(E)-Caryophyllene	1477	1474	0.02 ± 0.02	–	0.05 ± 0.02	0.02 ± 0.02	–	RI, MS
49	E-β-Guaiene	1504	1498	–	–	–	–	–	RI, MS
50	Viridiflorene	1510	1505	0.31 ± 0.07	0.21 ± 0.01	0.35 ± 0.03	0.28 ± 0.08	0.15 ± 0.03	RI, MS
51	Spathulenol	1594	1578	0.44 ± 0.13	0.27 ± 0.05	0.63 ± 0.04	0.59 ± 0.11	0.28 ± 0.02	RI, MS
52	Isoaromadendrene epoxide	1597		0.11 ± 0.02	0.11 ± 0.01	0.14 ± 0.01	0.15 ± 0.03	0.10 ± 0.02	MS
53	Caryophyllene oxide	1601	1599	0.21 ± 0.06	0.17 ± 0.03	0.30 ± 0.05	0.28 ± 0.06	0.19 ± 0.03	RI, MS
54	Caryophylla-4(12),8(13)-dien-5 beta-ol	1654	1663	0.09 ± 0.02	0.07 ± 0.01	0.11 ± 0.04	0.12 ± 0.02	0.08 ± 0.02	RI, MS
55	14-Hydroxy-9-epi-(E)-caryophyllene	1673	1669	0.06 ± 0.04	0.07 ± 0.01	0.11 ± 0.01	0.11 ± 0.02	0.06 ± 0.02	RI, MS
56	Khusinol	1686	1674	0.08 ± 0.02	0.07 ± 0.01	0.11 ± 0.01	0.11 ± 0.02	0.07 ± 0.01	RI, MS
	Monoterpene hydrocarbons			18.98	18.77	23.66	24.59	15.56	
	Oxygenated monoterpens			76.17	76.77	71.25	70.35	80.93	
	Sesquiterpene hydrocarbons			2.92	3.09	2.97	2.55	1.94	
	Oxygenated sesquiterpenes			0.99	0.76	1.40	1.35	0.78	
	Essential oil content (%)			4.8 ± 0.47	5.62 ± 0.91	5.18 ± 0.58	4.96 ± 0.34	4.15 ± 0.44	

(Continued)

TABLE 3 | Continued

N	Components	RI[a]	RI[b]	Populations				
				Daarbast	Konar Siah	Darab	Haneshk	Methods
1	α-Thujene	933	929	0.40 ± 0.24	0.43 ± 0.25	0.36 ± 0.03	0.28 ± 0.17	RI, MS
2	α-Pinene	940	939	3.89 ± 2.00	3.31 ± 2.50	2.62 ± 0.28	1.54 ± 0.95	RI, MS
3	Camphene	955	954	0.19 ± 0.07	0.14 ± 0.07	0.15 ± 0.01	0.10 ± 0.03	RI, MS
4	2,4(10)-Thujadiene	960	957	–	–	–	–	RI, MS
5	1,3-Octanol	969	973	0.07 ± 0.01	0.15 ± 0.01	0.16 ± 0.24	0.08 ± 0.02	RI, MS
6	Sabinene	979	977	0.11 ± 0.03	0.21 ± 0.10	0.10 ± 0.01	0.10 ± 0.02	RI, MS
7	β-Pinene	983	978	0.65 ± 0.28	0.56 ± 0.30	0.51 ± 0.04	0.30 ± 0.13	RI, MS
8	Myrcene	995	992	1.10 ± 0.27	1.15 ± 0.29	0.94 ± 0.11	2.34 ± 0.76	RI, MS
9	3-Octanol	997	995	0.18 ± 0.10	0.13 ± 0.10	0.04 ± 0.01	0.16 ± 0.09	RI, MS
10	α-Phellandrene	1010	1005	0.18 ± 0.01	0.18 ± 0.01	0.13 ± 0.01	0.34 ± 0.11	RI, MS
11	δ-3-Carene	1015	1011	0.03 ± 0.01	0.02 ± 0.01	–	–	RI, MS
12	α-Terpinene	1021	1019	1.15 ± 0.15	1.25 ± 0.25	1.04 ± 0.06	0.81 ± 0.41	RI, MS
13	p-Cymene	1030	1025	6.34 ± 2.69	6.42 ± 2.21	13.96 ± 1.1	4.33 ± 2.76	RI, MS
14	Limonene	1034	1030	0.58 ± 0.27	0.58 ± 0.27	0.61 ± 0.09	0.76 ± 0.13	RI, MS
15	1,8-Cineole	1037	1035	0.06 ± 0.01	0.06 ± 0.01	0.07 ± 0.05	0.19 ± 0.16	RI, MS
16	Z-β-Ocimene	1039	1039	0.52 ± 0.30	0.59 ± 0.33	0.45 ± 0.04	0.68 ± 0.30	RI, MS
17	E-β-Ocimene	1050	1050	0.08 ± 0.02	0.09 ± 0.03	0.04 ± 0.02	0.94 ± 0.50	RI, MS
18	γ-Terpinene	1063	1062	4.73 ± 0.97	4.79 ± 1.03	2.93 ± 0.24	2.60 ± 1.76	RI, MS
19	Z-Sabinene hydrate	1071	1070	0.20 ± 0.04	0.25 ± 0.05	0.21 ± 0.01	0.33 ± 0.27	RI, MS
20	Z-Linalool oxide	1076	1074	0.10 ± 0.01	0.04 ± 0.01	0.04 ± 0.02	0.45 ± 0.30	RI, MS
21	Terpinolene	1092	1089	0.21 ± 0.04	0.23 ± 0.06	0.21 ± 0.03	0.78 ± 0.28	RI, MS
22	Linalool	1103	1100	1.25 ± 0.70	1.73 ± 1.10	1.69 ± 0.97	37.65 ± 20.6	RI, MS
23	E-γ-Caryophyllene	1106	1106	0.06 ± 0.02	0.09 ± 0.06	0.10 ± 0.04	0.93 ± 0.50	RI, MS
24	1-Octenyl-3-acetate	1111	1113	0.03 ± 0.01	0.07 ± 0.06	–	0.11 ± 0.07	RI, MS
25	p-Menth-2-en-1-ol	1124	1122	0.04 ± 0.03	tr	–	0.16 ± 0.10	RI, MS
26	allo-Ocimene	1131	1132	–	–	–	0.38 ± 0.26	RI, MS
27	1,3,8-p-Menthatriene	1133		–	–	–	0.10 ± 0.08	MS
28	Borneol	1172	1171	–	–	–	0.99 ± 0.68	RI, MS
29	Z-Linalool oxide (pyranoid)	1177	1173	–	tr	0.12 ± 0.02	0.13 ± 0.06	RI, MS
30	4-Terpineol	1183	1179	0.19 ± 0.04	0.16 ± 0.05	0.18 ± 0.02	0.18 ± 0.05	RI, MS
31	p-Cymenol-8	1188	1184	0.48 ± 0.07	0.48 ± 0.05	0.59 ± 0.03	0.29 ± 0.09	RI, MS
32	α-Terpineol	1195	1190	0.72 ± 0.37	0.81 ± 0.40	0.58 ± 0.04	0.60 ± 0.09	RI, MS
33	Z-Dihydro carvone	1202	1200	0.19 ± 0.04	0.17 ± 0.08	0.03 ± 0.02	0.21 ± 0.11	RI, MS
34	Nerol	1234	1228	0.02 ± 0.01	–	0.02 ± 0.01	0.06 ± 0.04	RI, MS
35	Thymol methyl ether	1238	1237	0.08 ± 0.01	0.10 ± 0.08	1.52 ± 0.12	0.59 ± 0.50	RI, MS
36	Carvacrol methyl ether	1248	1241	3.71 ± 3.00	3.23 ± 3.00	1.62 ± 0.38	0.92 ± 0.72	RI, MS
37	Geraniol	1260	1263	tr	0.02 ± 0.01	tr	0.60 ± 0.28	RI, MS
38	Geranial	1280	1273	–	–	–	–	RI, MS
39	p-Thymol	1286		0.09 ± 0.05	0.07 ± 0.01	0.22 ± 0.03	0.05 ± 0.04	MS
40	Thymol	1295	1295	5.26 ± 3.85	8.65 ± 8.30	41.61 ± 4.14	17.55 ± 10.9	RI, MS
41	Carvacrol	1309	1305	61.41 ± 13.3	58.29 ± 13.3	21.81 ± 4.91	15.74 ± 12.9	RI, MS
42	Thymol acetate	1358	1359	0.10 ± 0.10	0.17 ± 0.13	0.75 ± 0.09	0.21 ± 0.13	RI, MS
43	Carvacrol acetate	1377	1368	1.12 ± 0.61	1.45 ± 1.23	0.31 ± 0.09	0.20 ± 0.09	RI, MS
44	β-Bourbonene	1384	1378	–	–	–	0.04 ± 0.03	RI, MS
45	E-Caryophyllene	1435	1427	1.93 ± 0.62	1.59 ± 0.36	1.59 ± 0.38	1.71 ± 0.62	RI, MS
46	Aromadendrene	1455	1436	0.47 ± 0.07	0.38 ± 0.16	0.24 ± 0.09	0.19 ± 0.11	RI, MS
47	α-Humulene	1469	1452	0.14 ± 0.02	0.12 ± 0.03	0.11 ± 0.02	0.16 ± 0.04	RI, MS
48	9-epi-(E)-Caryophyllene	1477	1474	0.03 ± 0.01	0.03 ± 0.03	–	tr	RI, MS
49	E-β-Guaiene	1504	1498	0.02 ± 0.01	tr	–	–	RI, MS
50	Viridiflorene	1510	1505	0.38 ± 0.13	0.32 ± 0.18	0.14 ± 0.05	0.36 ± 0.15	RI, MS
51	Spathulenol	1594	1578	0.48 ± 0.24	0.44 ± 0.24	0.35 ± 0.05	0.77 ± 0.31	RI, MS
52	Isoaromadendrene epoxide	1597		0.10 ± 0.03	0.09 ± 0.04	0.21 ± 0.02	0.27 ± 0.10	MS
53	Caryophyllene oxide	1601	1599	0.21 ± 0.04	0.20 ± 0.07	0.35 ± 0.03	0.45 ± 0.13	RI, MS
54	Caryophylla-4(12),8(13)-dien-5 beta-ol	1654	1663	0.09 ± 0.02	0.07 ± 0.05	0.12 ± 0.01	0.21 ± 0.06	RI, MS
55	14-Hydroxy-9-epi-(E)-caryophyllene	1673	1669	0.07 ± 0.03	0.05 ± 0.04	0.13 ± 0.01	0.16 ± 0.06	RI, MS
56	Khusinol	1686	1674	0.08 ± 0.02	0.06 ± 0.01	0.12 ± 0.01	0.17 ± 0.05	RI, MS
	Monoterpene hydrocarbons			20.16	19.95	24.05	16.38	
	Oxygenated monoterpens			75.05	75.75	71.37	77.21	
	Sesquiterpene hydrocarbons			3.03	2.53	2.18	3.39	

(Continued)

TABLE 3 | Continued

N	Components	RI[a]	RI[b]	Populations				
				Daarbast	Konar Siah	Darab	Haneshk	Methods
	Oxygenated sesquiterpenes			1.03	0.91	1.28	2.03	
	Essential oil content (%)			5.58 ± 0.48	5.89 ± 0.24	3.3 ± 0.26	3.71 ± 0.41	

tr, trace < 0.02%.

a: RI, linear retention indices on HP-5MS column, experimentally determined using homologue series of n-alkanes.

b: Relative retention indices taken from Adams and NIST.

Methods: MS, by comparison of the mass spectrum with those of the computer mass libraries Adams and NIST.

matter (OM) ranged from 4% to 10% (Haneshk and Darab regions, respectively). The soil of regions were rich in calcium (Ca), iron (Fe), potassium (K) and aluminum (Al) whereas nitrogen (N) and phosphor (P) were present in lower levels. Furthermore, *Z. multiflora* grows on soils with alkaline pH (7.60 to 7.90).

The volatile constituents were influenced by edaphic factors (**Table 4**). Carvacrol was significantly positively correlated with pH, Ca, and temperature [0.69 ($p < 0.01$), 0.62 and 0.54 ($p < 0.05$) respectively] and there was a highly negative correlation between carvacrol and Al, Fe, and K. The correlation analysis indicated that linalool was considerably positively correlated with Al, Fe, and K ($p < 0.01$). No statistically significant correlations were detected among N and EO content and phytochemical constituents.

The SEM approach was used to dissect the contribution of environmental factors on EO and EO constituent content. Significant SEMs for EO [APC = 0.641 (P < 0.001), ARS = 0.571 (P=0.002), AVIF = 1.001], thymol [APC = 0.874 (P < 0.001), ARS = 0.770 (P < 0.001), AVIF = 1.435], carvacrol [APC = 0.602 (P = 0.001), ARS = 0.560 (P = 0.002), AVIF = 1.536] and linalool [APC = 0.489 (P = 0.005), ARS = 0.655 (P=0.008), AVIF = 1.019] were obtained. The portion of clay and

phosphor had a direct negative influence on EO content. The altitude had a positive effect on phosphor content while latitude had a negative effect on clay content in the soil (**Figure 4A**). Thymol content was positively affected by clay amount in the soil and indirect negatively via the negative effect of latitude on clay (**Figure 4B**). Carvacrol was directly positively influenced by silt content and pH-value in the soil, which was positively depended on the amount of sand in the soil (**Figure 4C**). Latitude had a negative effect on soil silt and a positive one on the soil sand portion. The linalool content was affected on the one hand, directly by longitude (positively) and on the other hand by silt (negatively) while silt content itself was negatively affected by latitude (**Figure 4D**).

Quantitative Analysis of EO Composition by NIRS

The dried leaves of specimens of *Z. multiflora* from different regions were analyzed by near infrared spectroscopy and hierarchical cluster analysis (HCA, Wards Algorithm). The NIR spectra of *Z. multiflora* were characterized by combination, first and second overtone vibrations in the range of 4,000 to 12,000 cm^{-1}. HCA was used to group samples according to their spectral appearance determined through their chemical profile. **Figure 5** presents the appropriate HCA plot showing the separation of *Z. multiflora* populations into different clusters. In contrast to GC analysis, NIRS combines spectral features of chemically similar structures. Hence, carvacrol, thymol, and *p*-cymene, all characterized by an isopropyl- and methyl-substituted aromatic ring system, show all nearly identical NIRS absorption patterns. Therefore, for NIRS not only the quantity of individual EO components are relevant, but the amount of structurally related substances. As shown in **Figure 5** HCA resulted on highest level of heterogeneity in the clustering of samples according to the ratio of aromatic EO compounds (thymol + carvacrol + *p*-cymene) to aliphatic, isolated C=C structures (linalool). On the next level, types with a high content of aromatic structures are divided into sub-clusters with high amounts of carvacrol (cluster IIIB), high thymol, and high linalool or high *p*-cymene (cluster IIIA) or high carvacrol and high *p*-cymene (cluster II).

Chemometrics of superintended pattern identification based on PLS-DA of GC combined with NIR spectroscopy was endeavored to categorize fourteen populations of *Z. multiflora*. Quantification models for the EO content and for major compounds were developed by 10-fold cross-validation
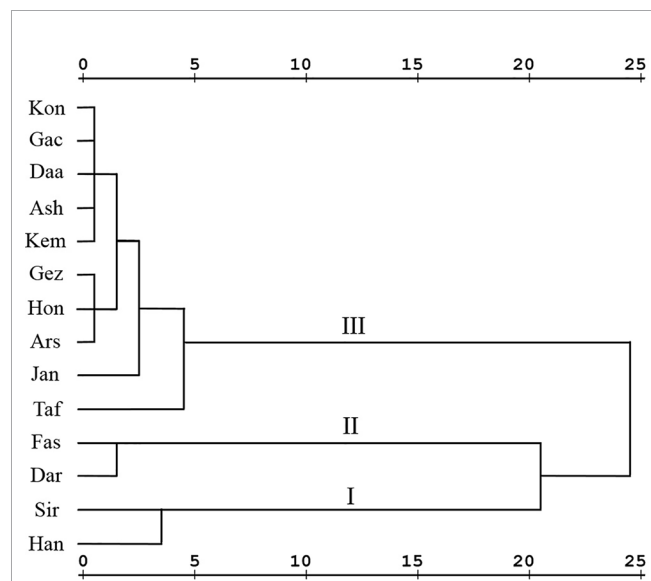

FIGURE 3 | Hierarchical cluster analysis of *Zataria multiflora* populations based on phytochemical composition.
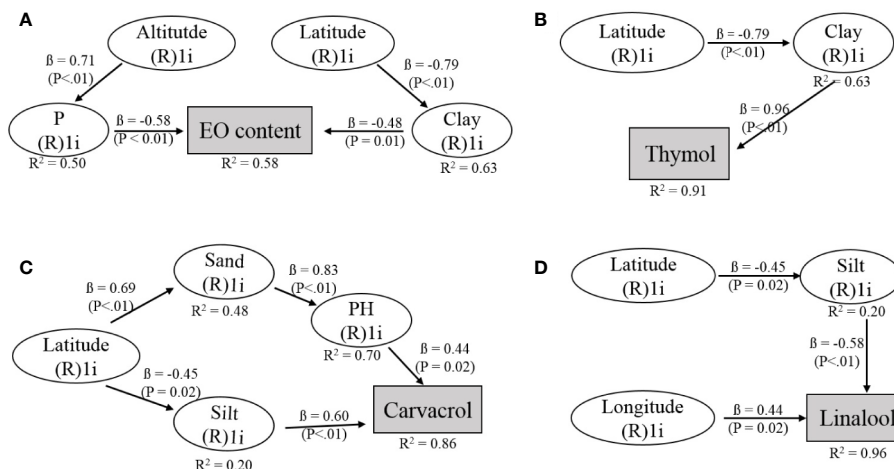
FIGURE 4 | Hypothetical structural equation models (SEMs) to describe the relationships between geographical and edaphic factors and **(A)** EO, **(B)** thymol, **(C)** carvacrol, and **(D)** linalool content of *Zataria multiflora*. The climatic factors, temperature, and rainfall were included in the full model but did not explain EO or EO constituent content. R^2: coefficient of determination indicating the variability explained for each variable. ß- values indicate the path coefficients, P: significance level for relationship.

procedure according to literature (Krähmer et al., 2013). Therefore, averaged spectra for each plant were correlated with GC reference data for carvacrol, thymol, and linalool as well as EO content. For all constituents, appropriate prediction models were achieved. **Figure 6** shows the results of cross-validation according to plant wise averaged NIR spectra from all populations. Generally, coefficients of determination (R^2) were higher than 0.82 for individual components and EO content. As shown in **Figure 6A**, NIRS offers a fast tool for estimation of EO content with a coefficient of determination $R^2 = 0.85$ and a root mean square error of prediction (RMESP) below 10% of mean EO content (the mean of EO content over all samples used in the model, according to **Figure 6** something about 4 to 5 ml/100 g) (RMSEP = 0.431%). Furthermore, for major EO components, prediction quality was best for linalool ($R^2 = 0.97$) followed by $R^2 = 0.87$ and $R^2 = 0.82$ for carvacrol and thymol (**Figures 6B–D**), respectively.

DISCUSSION

This study investigated the effect of different environmental factors on EO production, the content of specific EO compounds as well as on chemotype of different *Z. multiflora* populations. The EO values (up to 5.89% dry weight) detected in 14 populations in Iran were higher than those reported previously in the literature including 1.2% to 3.4% (Hadian et al., 2011a), 2.91% to 4% (Sadeghi et al., 2015), and 1.93% to 2.22% (Golkar et al., 2020). The EO content can be affected by geological, climatic, and edaphic characteristics as well as harvesting time. Saei-Dehkordi et al. (2010) described that the largest quantity of the EO content of *Z. multiflora* was collected in mid-May with 1.57% (v/w). Thus, knowledge on the season, phenological stage, and harvesting time during the day is necessary to obtain high quantities of EO content. Of the chemical

constituents detected, carvacrol, thymol, linalool, *p*-cymene, *γ*-terpinene, and *α*-pinene were found as the main compounds of *Z. multiflora*. In other studies, the highest diversity was shown for the monoterpenes, including carvacrol, thymol, linalool, and *p*-cymene (Shafiee and Javidnia, 1997; Abkenar et al., 2008; Mahboubi and Bidgoli, 2010; Ziaee et al., 2018). Carvacrol, the major compound of the Jandagh population, has been previously reported as one of the most important components of EO among various members of the Lamiaceae family (Ebrahimi et al., 2008; Hadian et al., 2011b; Stefanaki et al., 2018; Santos et al., 2019). The main component of Darab and Fasa populations was thymol (41.61% and 48.12% respectively), which is an isomer of carvacrol. Saei-Dehkordi et al. (2010) and Sharififar et al. (2007) had depicted thymol as the most abundant component in the essential oil profile of *Z. multiflora* from different areas in Iran. Contrariwise, two other studies showed carvacrol as the main constituent of *Z. multiflora* (Basti et al., 2007; Khosravi et al., 2009). Moreover, EO of *Z. multiflora* contains other important monoterpene constituents like linalool, *p*-cymene, *γ*-terpinene, and *α*-pinene. Siriz and Haneshk populations were rich in linalool (55.38% and 37.65% respectively) and *p*-cymene was one of the main components of Darab population (13.96%).

The positive and negative correlations between EO components indicate the presence of three different chemotypes: thymol, carvacrol, and linalool. Furthermore, they indicate which compounds are interlinked in a chain of monoterpene synthesis with certain branches in the predicted enzymatic pathway: while geranyl-diphosphate is the precursor of non-phenolic linalool and phenolic thymol and carvacrol, the latter are connected *via p*-cymene (Thompson, 2005). In agreement to our results, similar correlations between individual EO components were found in *Artemisia dracunculus*, where methyl chavicol as the main constituent of *A. dracunculus* was positively correlated with

**TABLE 4 |** Pearson correlation coefficients between EO content, major components and edaphic factors.

Edaphic factors	EO	Linalool	Thymol	Carvacrol	*p*-Cymene
Altitude	−0.73**	0.53*	0.20	−0.41	−0.45
Temperature	0.73**	−0.59*	−0.30	0.54*	0.37
OM	0.13	−0.55*	0.45	0.02	0.80**
pH	0.67**	−0.42	−0.64*	0.69**	0.01
Sand	0.22	0.41	−0.63*	0.15	−0.64*
Silt	−0.12	−0.56*	0.50	0.06	0.64*
Clay	−0.30	0.14	0.53*	−0.52	0.27
N	−0.31	−0.04	0.47	−0.27	0.49
P	−0.56*	0.44	0.04	−0.28	−0.23
K	−0.73**	0.57*	0.41	−0.63*	−0.25
Ca	0.62*	−0.65*	−0.29	0.62*	0.23
Fe	−0.78**	0.60*	0.50	−0.72**	−0.15
Al	−0.75**	0.55*	0.49	−0.68**	−0.17

*Significance: *P < 0.05; **P < 0.01.*

terpinolene and methyl eugenol, and negatively correlated with α-pinene, limonene, (Z)-β-ocimene, and (E)-β-ocimene (Karimi et al., 2015). Hierarchical cluster analysis based on phytochemical components was proven to be a helpful tool to classify medicinal and aromatic plants accessions. For instance, cluster analysis on *Verbascum songaricum* resulted into nine groups (Selseleh et al., 2019) and for lemon balm populations three different chemotypes could be identified (Pouyanfar et al., 2018). Also grouping based on EO constituents of four *Vitex specimens* revealed different clusters (de Sena Filho et al., 2017). In the present study, the components of the EO measured at full

flowering stage underpin the presence of the three chemotypes (carvacrol, thymol, linalool).

Rapid and reliable identification of medicinal plant species and chemotypes concerning authenticity and quality is crucial for pharmaceutical and food processing. Spectroscopy techniques as fast and easy handling technologies are nowadays widely applied directly on plant material for qualitative and semi-quantitative characterization. Different studies describe the application of NIRS, IRS, and Raman for differentiation of chemotypes and prediction of EO composition in various medicinal and aromatic plants (Seidler-Lozykowska et al., 2010; Gudi et al., 2014; Farag et al.,
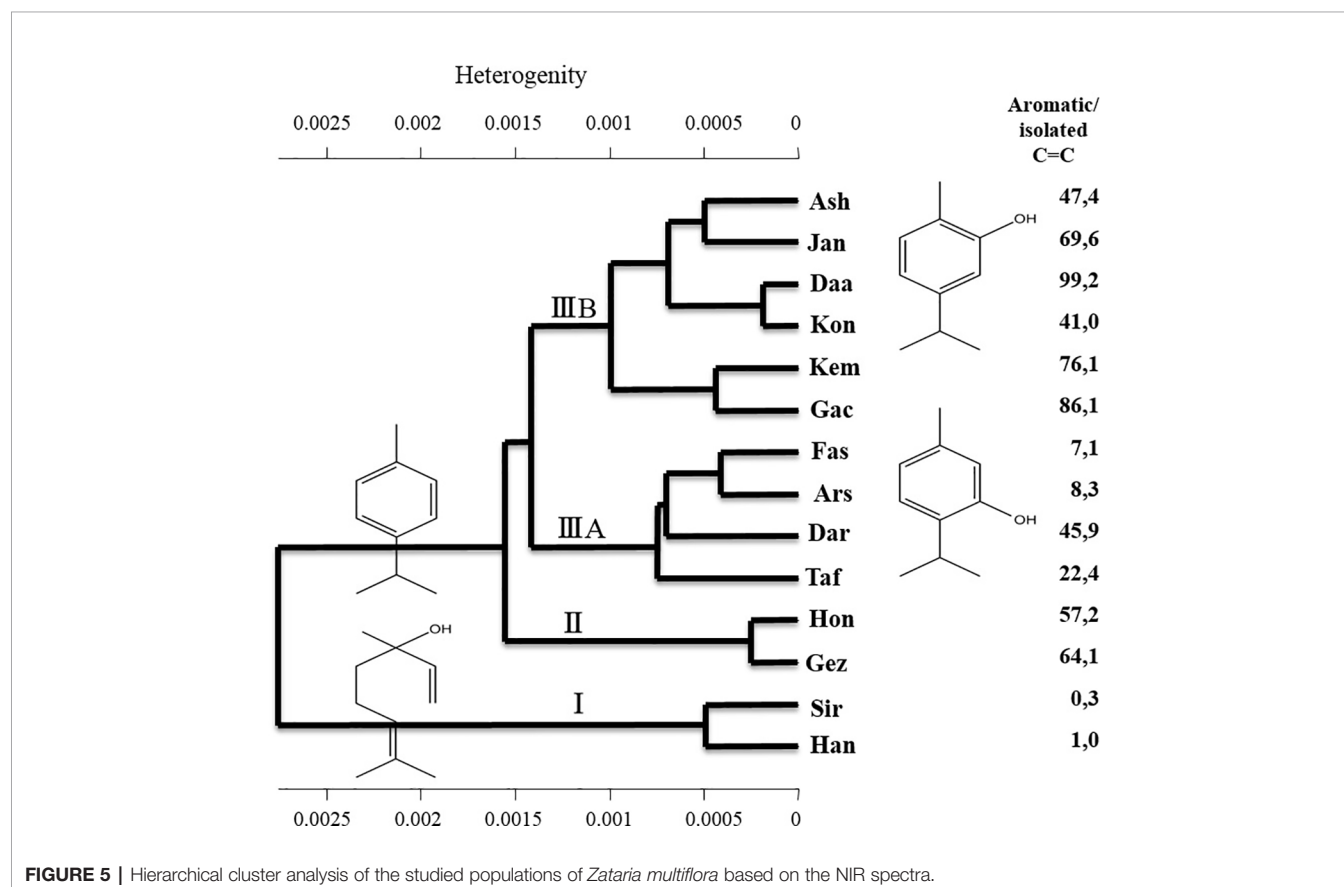
FIGURE 5 | Hierarchical cluster analysis of the studied populations of *Zataria multiflora* based on the NIR spectra.

2018). For *Z. multiflora* the presented quantification models are not accurate for exact determination at current state, since, e.g., for linalool, samples are very inhomogeneous distributed over the investigated range of concentration. Nevertheless, in combination with HCA, near infrared spectroscopy offers a fast method for chemotyping and EO estimation already on plant material. An improved prediction of EO content and main components with regard to cross-validation concerning averaged ATR–FTIR spectra can also be achieved for constituents with lower concentrations (Gudi et al., 2015). The high correlation between NIRS and GC data allows application of NIRS for authenticity and quality control directly on the plant material for the flavor and fragrance as well as pharmaceutical industries. NIR spectroscopy can be used to classify plants according to their chemotype as well as predict the content of valuable components such as carvacrol, thymol, and linalool as well as other terpenes, rapidly and accurately.

The effect of soil parameters and climatic condition on plant perfomance and EO content has been shown for many plant species. For example, *Kelussia odoratissima* Mozaff grows in dark soil, rich in mineral content (Raiesi et al., 2013) and growth habitats of *Thymus pulegioides* were characterized by high amount of Al, Ca, Fe, K, and Si, however, by low amount of P and Mn (Vaičiulytè et al., 2017). Mexican oregano populations grown in soil with high nitrogen and iron content, lower soil water availability, and higher pH values showed a higher EO yield (Martínez-Natarén et al., 2012). It is widely accepted that environmental conditions affect plant EO content and its components (Ormeño et al., 2008; Mansour et al., 2010). Several studies have revealed that the predominance of carvacrol or thymol in different Lamiaceae species is related to environmental factors (Boira and Blanquer, 1998; Economou et al., 2011). In *Thymus vulgaris* such phenolic chemotypes cope better with summer drought, while non-phenolic (e.g., linalool) chemotypes cope better with early-winter freezing temperatures (Thompson et al., 2007). In our study the Pearson correlations revealed that altitude, K, Fe, and Al were significantly ($p < 0.01$) negatively correlated with EO content (**Table 3**). In agreement to our results, the lowest altitudes showed higher EO yield in *Lavandula angustifolia* (Demasi et al., 2018) and *Satureja rechingeri* (Hadian et al., 2014). Also, a correlation between higher EO yields at decreasing altitudes was found in *Origanum vulgare* (Giuliani et al., 2013). Notwithstanding the effect of geographical condition, EO content and EO constituents can be affected by edaphic factors and climatic conditions, for example, the soil type affects *Origanum syriacum* chemotype (El-Alam et al., 2019). In our study EO yield showed a highly significant positive correlation with temperature, pH value, and Ca. Former studies highlighted the same behavior in other aromatic plants and suggest that the wide variation in the chemical composition of the EO can be ascribed to habitat influences in *Origanum compactum* (Aboukhalid et al., 2017) and *Origanum vulgare* L. (De Falco et al., 2013). The influence of environmental conditions on EO of *Origanum vulgare* ssp. showed a negative correlation with altitude and a positive correlation with soil temperature and air temperature (Tuttolomondo et al., 2014). SEMs were applied to impute relationships between the different factors and revealed indirect geographic and direct edaphic effects on EO content and compounds, while climate factors do not have an influence. Chemotype and high amount of specific compounds can thus be predicted when looking for populations with specific features. Biotic factors like co-occuring vegetation (Wäschke et al., 2015) or herbivore activity (Dicke et al., 2009) can additionally influence the metabolome profile of plants and shall be considered in future studies.
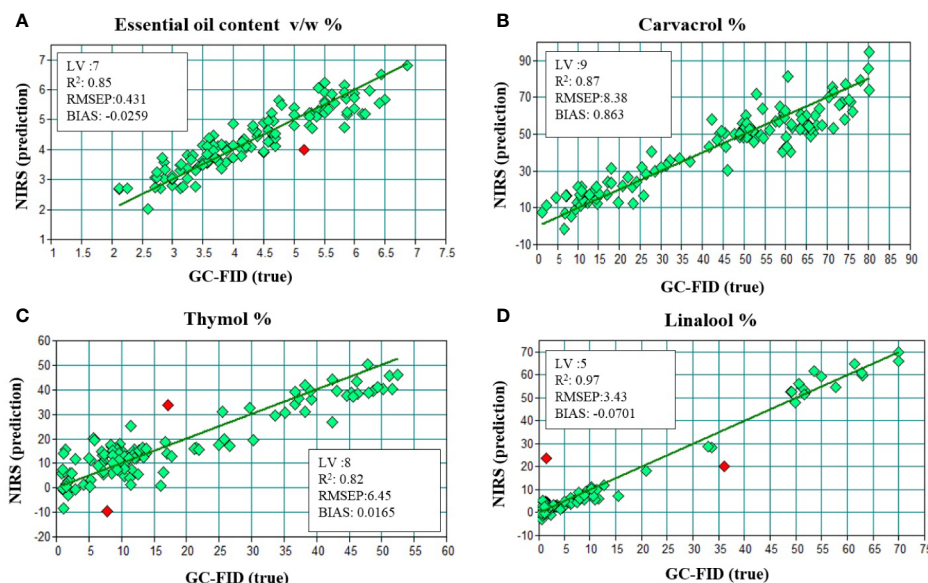

FIGURE 6 | Results of 10-fold cross-validation of NIR and GC data for the **(A)** EO content, **(B)** carvacrol, **(C)** thymol, **(D)** linalool by correlation of averaged spectra for each population.

CONCLUSIONS

Medicinal and aromatic plants play important roles all over the world because of their wide application due to pharmacological, therapeutic, industrial, and agricultural properties. The varying climate and environmental growth conditions lead to a huge phytochemical diversity of these resources. *Zataria multiflora* is a valuable medicinal plant with various pharmaceutical properties and has potential as source of compounds with agricultural relevance as plant protection agents. Ingredients such as carvacrol, thymol, and linalool are responsible for the respective effects and show a high variability among the investigated populations. Environmental conditions are affecting the EO content and its components. Hence, existing variability in the chemical profile of studied populations allow selection of populations with distinct scent or bioactive components for use in pertinent industries and breeding purposes. Our approach of identifying environmental predictors for EO content, chemotype or presence of high amounts of specific compounds can help to identify regions for sampling plant material with the desired chemical profile. Based on mobile NIRS devices, fast classification of yet undescribed populations and individual plants together with an EO profiling can be performed directly in the field.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding authors.

REFERENCES

AUTHOR CONTRIBUTIONS

TM, HS, and JH conceived and designed the project; AlK performed all sampling, extraction, and chemical analyses, except soil analysis which was performed by NH. Statistical analyses were performed by AlK, TM, and AnK. AlK and TM wrote the article with contributions from all other authors.

Abkenar, S. D., Yamini, Y., Shemirani, F., and Assadi, Y. (2008). Headspace solid phase microextraction using a porous-layer activated charcoal coating fused silica fiber for identification of volatile organic compounds emitted by *Zataria multiflora* Boiss. *Chem. Anal-Warsaw* 53, 277–287.

Aboukhalid, K., Al Faiz, C., Douaik, A., Bakha, M., Kursa, K., Agacka-Mołdoch, M., et al. (2017). Influence of environmental factors on essential oil variability in *Origanum compactum* Bent. *Chem. Biodivers.* 14, e1700158. doi: 10.1002/cbdv.201700158

Adams, R. P. (2014). *Identification of essential oil components by gas chromatography/mass spectrometry* Vol. 456 (Carol Stream, IL: Allured publishing corporation).

Basti, A. A., Misaghi, A., and Khaschabi, D. (2007). Growth response and modelling of the effects of *Zataria multiflora* Boiss. essential oil, pH and temperature on *Salmonella typhimurium* and *Staphylococcus aureus*. *LWT Food Sci. Technol.* 40, 973–981. doi: 10.1016/j.lwt.2006.07.007

Boira, H., and Blanquer, A. (1998). Environmental factors affecting chemical variability of essential oils in *Thymus piperella* L. *Biochem. Syst. Ecol.* 26, 811–822. PII: S0305-1978(98)00047-7. doi: 10.1016/S0305-1978(98)00047-7

Chin, W. W., and Newsted, P. R. (1999). Structural equation modeling analysis with small samples using partial least squares. *Stat. Strateg. Small Sample Res.* 1 (1), 307–341.

De Falco, E., Roscigno, G., Iodice, C., and Senatore, F. (2013). Phytomorphological and Essential Oil Characterization *in situ* and ex situ of wild biotypes of *Oregano* collected in the Campania Region (Southern Italy). *Chem. Biodivers.* 10, 2078–2090. doi: 10.1002/cbdv.201300185

de Sena Filho, J. G., Barreto, I. C., Soares Filho, A. O., Nogueira, P. C., Teodoro, A. V., Cruz da Silva, A. V., et al. (2017). Volatile metabolomic composition of *Vitex* species: chemodiversity insights and acaricidal activity. *Front. Plant Sci.* 8, 1931. doi: 10.3389/fpls.2017.01931

Demasi, S., Caser, M., Lonati, M., Cioni, P. L., Pistelli, L., Najar, B., et al. (2018). Latitude and altitude influence secondary metabolite production in peripheral alpine populations of the Mediterranean species *Lavandula angustifolia* Mill. *Front. Plant Sci.* 9, 983. doi: 10.3389/fpls.2018.00983

Dicke, M., Van Loon, J. J., and Soler, R. (2009). Chemical complexity of volatiles from plants induced by multiple attack. *Nat. Chem. Biol.* 5, 317. doi: 10.1038/nchembio.169

Ebrahimi, S. N., Hadian, J., Mirjalili, M., Sonboli, A., and Yousefzadi, M. (2008). Essential oil composition and antibacterial activity of *Thymus caramanicus* at different phenological stages. *Food Chem.* 110, 927–931. doi: 10.1016/j.foodchem.2008.02.083

Economou, G., Panagopoulos, G., Tarantilis, P., Kalivas, D., Kotoulas, V., Travlos, I. S., et al. (2011). Variability in essential oil content and composition of *Origanum hirtum* L., *Origanum onites* L., *Coridothymus capitatus* (L.) and *Satureja thymbra* L. populations from the Greek island Ikaria. *Ind. Crops Prod.* 33, 236–241. doi: 10.1016/j.indcrop.2010.10.021

El-Alam, I., Zgheib, R., Iriti, M., El Beyrouthy, M., Hattouny, P., Verdin, A., et al. (2019). *Origanum syriacum* Essential Oil Chemical Polymorphism According to Soil Type. *Foods* 8, 90. doi: 10.3390/foods8030090

Farag, N. F., El-Ahmady, S. H., Abdelrahman, E. H., Naumann, A., Schulz, H., Azzam, S. M., et al. (2018). Characterization of essential oils from Myrtaceae species using ATR-IR vibrational spectroscopy coupled to chemometrics. *Ind. Crops Prod.* 124, 870–877. doi: 10.1016/j.indcrop.2018.07.066

Formisano, C., Delfine, S., Oliviero, F., Tenore, G. C., Rigano, D., and Senatore, F. (2015). Correlation among environmental factors, chemical composition and antioxidative properties of essential oil and extracts of chamomile (*Matricaria chamomilla* L.) collected in Molise (South-central Italy). *Ind. Crops Prod.* 63, 256–263. doi: 10.1016/j.indcrop.2014.09.042

Giuliani, C., Maggi, F., Papa, F., and Maleci Bini, L. (2013). Congruence of phytochemical and morphological profiles along an altitudinal gradient in *Origanum vulgare* ssp. *vulgare* from Venetian Region (NE Italy). *Chem. Biodivers.* 10 (4), 569–583. doi: 10.1002/cbdv.201300019

Golkar, P., Mosavat, N., and Jalali, S. A. H. (2020). Essential oils, chemical constituents, antioxidant, antibacterial and *in vitro* cytotoxic activity of different *Thymus* species and *Zataria multiflora* collected from Iran. *S. Afr. J. Bot.* 130, 250–258. doi: 10.1016/j.sajb.2019.12.005

Gudi, G., Krähmer, A., Krüger, H., Hennig, L., and Schulz, H. (2014). Discrimination of fennel chemotypes applying IR and Raman spectroscopy: Discovery of a new γ-asarone chemotype. *J. Agric. Food. Chem.* 62, 3537–3547. doi: 10.1021/jf405752x

Gudi, G., Krähmer, A., Krüger, H., and Schulz, H. (2015). Attenuated total reflectance–Fourier transform infrared spectroscopy on intact dried leaves of sage (*Salvia officinalis* L.): accelerated chemotaxonomic discrimination and analysis of essential oil composition. *J. Agric. Food. Chem.* 63, 8743–8750. doi: 10.1021/acs.jafc.5b03852

Hadian, J., Ebrahimi, S. N., Mirjalili, M. H., Azizi, A., Ranjbar, H., and Friedt, W. (2011a). Chemical and genetic diversity of *Zataria multiflora* Boiss. accessions growing wild in Iran. *Chem. Biodivers.* 8, 176–188. doi: 10.1002/cbdv.201000070

Hadian, J., Hossein Mirjalili, M., Reza Kanani, M., Salehnia, A., and Ganjipoor, P. (2011b). Phytochemical and morphological characterization of *Satureja khuzistanica* Jamzad populations from Iran. *Chem. Biodivers.* 8 (5), 902–915. doi: 10.1002/cbdv.201000249

Hadian, J., Esmaeili, H., Nadjafi, F., and Khadivi-Khub, A. (2014). Essential oil characterization of *Satureja rechingeri* in Iran. *Ind. Crops Prod.* 61, 403–409. doi: 10.1016/j.indcrop.2014.07.034

Iranian Herbal Pharmacopoeia Committee (2002). *Iranian Herbal Pharmacopoeia*. (Tehran: Ministry of Health and Medical Education of Iran), 2, 40711.

Karimi, A., Hadian, J., Farzaneh, M., and Khadivi-Khub, A. (2015). Phenotypic diversity and volatile composition of Iranian *Artemisia dracunculus*. *Ind. Crops Prod.* 65, 315–323. doi: 10.1016/j.indcrop.2014.12.003

Khazdair, M. R., Ghorani, V., Alavinezhad, A., and Boskabady, M. H. (2018). Pharmacological effects of *Zataria multiflora* Boiss L. and its constituents focus on their anti-inflammatory, antioxidant, and immunomodulatory effects. *Fund. Clin. Pharmacol.* 32, 26–50. doi: 10.1111/fcp.12331

Khosravi, A. R., Shokri, H., Tootian, Z., Alizadeh, M., and Yahyaraeyat, R. (2009). Comparative efficacies of *Zataria multiflora* essential oil and itraconazole against disseminated *Candida albicans* infection in BALB/c mice. *Braz. J. Microbiol.* 40, 439–445. doi: 10.1590/S1517-83822009000300003

Kock, N., and Lynn, G. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *JAIS* 13 (7), 1–40. doi: 10.17705/1jais.00302

Kock, N. (2010). Using WarpPLS in e-collaboration studies: An overview of five main analysis steps. *IJeC* 6, 1–11. doi: 10.4018/jec.2010100101

Krähmer, A., Gudi, G., Weiher, N., Gierus, M., Schütze, W., and Schulz, H. (2013). Characterization and quantification of secondary metabolite profiles in leaves of red and white clover species by NIR and ATR-IR spectroscopy. *Vib. Spectrosc.* 68, 96–103. doi: 10.1016/j.vibspec.2013.05.012

Mahboubi, M., and Bidgoli, F. G. (2010). Antistaphylococcal activity of *Zataria multiflora* essential oil and its synergy with vancomycin. *Phytomedicine* 17, 548–550. doi: 10.1016/j.phymed.2009.11.004

Mahboubi, M. (2019). Therapeutic potential of *Zataria multiflora* Boiss. in treatment of irritable bowel syndrome (IBS). *J. Diet. Suppl.* 16, 119–128. doi: 10.1080/19390211.2017.1409852

Mahmoudvand, H., Mirbadie, S. R., Sadooghian, S., Harandi, M. F., Jahanbakhsh, S., and Saedi Dezaki, E. (2017). Chemical composition and scolicidal activity of *Zataria multiflora* Boiss essential oil. *J. Essent. Oil Res.* 29, 42–47. doi: 10.1080/10412905.2016.1201546

Mansour, A., Enayat, K., Neda, M. S., and Behzad, A. (2010). Antibacterial effect and physicochemical properties of essential oil of *Zataria multiflora* Boiss. *Asian Pac. J. Trop. Med.* 3 (6), 439–442. doi: 10.1016/S1995-7645(10)60105-8

Martínez-Natarén, D. A., Parra-Tabla, V., Dzib, G., Acosta-Arriola, V., Canul-Puc, K. A., and Calvo-Irabién, L. M. (2012). Essential oil yield variation within and among wild populations of Mexican oregano (*Lippia graveolens* HBK-Verbenaceae), and its relation to climatic and edaphic conditions. *J. Essent. Oil-Bear. Plants* 15, 589–601. doi: 10.1080/0972060X.2012.10644093

Milos, M., Radonic, A., Bezic, N., and Dunkic, V. (2001). Localities and seasonal variations in the chemical composition of essential oils of *Satureja montana* L. and *S. cuneifolia* Ten. *Flavour Fragr. J.* 16, 157–160. doi: 10.1002/ffj.965

Moazeni, M., Larki, S., Oryan, A., and Saharkhiz, M. J. (2014). Preventive and therapeutic effects of *Zataria multiflora* methanolic extract on hydatid cyst: An *in vivo* study. *Vet. Parasitol.* 205, 107–112. doi: 10.1016/j.vetpar.2014.07.006

Mohajeri, F. A., Misaghi, A., Gheisari, H., Basti, A. A., Amiri, A., Ghalebi, S. R., et al. (2018). The effect of *Zataria multiflora* Boiss Essential oil on the growth and citrinin production of *Penicillium citrinum* in culture media and cheese. *Food Chem. Toxicol.* 118, 691–694. doi: 10.1016/j.fct.2018.06.021

Morshedloo, M. R., Salami, S. A., Nazeri, V., Maggi, F., and Craker, L. (2018). Essential oil profile of oregano (*Origanum vulgare* L.) populations grown under similar soil and climate conditions. *Ind. Crops Prod.* 119, 183–190. doi: 10.1016/j.indcrop.2018.03.049

Niczad, A., Sharafzadeh, S., Alizadeh, A., Amiri, B., and Bazrafshan, F. (2019). Variability in essential oil constituent, phenolic content, antioxidant and antimicrobial activities of different ecotypes of *Zataria multiflora* Boiss. from Iran. *J. Essent. Oil Bear. Pl.* 22, 1435–1449. doi: 10.1080/0972060X.2020.1713221

NIST Chemistry WebBook, 69 (2002). *NIST Standard Reference Database Number 69*. P. J. Linstrom and W. G. Mallard Eds. (National Institute of Standards and Technology, Gaithersburg, USA). Available at: http://webbook.nist.gov.

Ormeño, E., Baldy, V., Ballini, C., and Fernandez, C. (2008). Production and diversity of volatile terpenes from plants on calcareous and siliceous soils: effect of soil nutrients. *J. Chem. Ecol.* 34, 1219. doi: 10.1007/s10886-008-9515-2.

Pavela, R., Żabka, M., Vrchotová, N., and Tříska, J. (2018). Effect of foliar nutrition on the essential oil yield of Thyme (*Thymus vulgaris* L.). *Ind. Crops Prod.* 112, 762–765. doi: 10.1016/j.indcrop.2018.05.048

Pouyanfar, E., Hadian, J., Akbarzade, M., Hatami, M., Kanani, M. R., and Ghorbanpour, M. (2018). Analysis of phytochemical and morphological variability in different wild-and agro-ecotypic populations of *Melissa officinalis* L. growing in northern habitats of Iran. *Ind. Crops Prod.* 112, 262–273. doi: 10.1016/j.indcrop.2017.12.008

Raiesi, S., Nadjafi, F., Hadian, J., Kanani, M. R., and Ayyari, M. (2013). Autecological and phytochemical studies of Kelussia odoratissima Mozaff. an endangered ethnomedicinal plant of Iran. *JBAPN* 3, 285–294. doi: 10.1080/22311866.2013.782748

Sadeghi, H., Robati, Z., and Saharkhiz, M. J. (2015). Variability in *Zataria multiflora* Bioss. essential oil of twelve populations from Fars province, Iran. *Ind. Crops Prod.* 67, 221–226. doi: 10.1016/j.indcrop.2015.01.021

Saedi Dezaki, E., Mahmoudvand, H., Sharififar, F., Fallahi, S., Monzote, L., and Ezatkhah, F. (2016). Chemical composition along with anti-leishmanial and cytotoxic activity of *Zataria multiflora*. *Pharm. Biol.* 54, 752–758. doi: 10.3109/13880209.2015.1079223

Saei-Dehkordi, S. S., Tajik, H., Moradi, M., and Khalighi-Sigaroodi, F. (2010). Chemical composition of essential oils in *Zataria multiflora* Boiss. from different parts of Iran and their radical scavenging and antimicrobial activity. *Food Chem. Toxicol.* 48, 1562–1567. doi: 10.1016/j.fct.2010.03.025

Sajed, H., Sahebkar, A., and Iranshahi, M. (2013). *Zataria multiflora* Boiss. (Shirazi thyme) - an ancient condiment with modern pharmaceutical uses. *J. Ethnopharmacol.* 145, 686–698. doi: 10.1016/j.jep.2012.12.018

Saleem, M., Nazli, R., Afza, N., Sami, A., and Shaiq Ali, M. (2004). Biological significance of essential oil of *Zataria multiflora* Boiss. *Nat. Prod. Res.* 18, 493–497. doi: 10.1080/14786410310001608064

Santos, J. D., Coelho, E., Silva, R., Passos, C. P., Teixeira, P., Henriques, I., et al. (2019). Chemical composition and antimicrobial activity of *Satureja montana* byproducts essential oils. *Ind. Crops Prod.* 137, 541–548. doi: 10.1016/j.indcrop.2019.05.058

Schulz, H., Baranska, M., Belz, H. H., Rösch, P., Strehle, M. A., and Popp, J. (2004). Chemotaxonomic characterisation of essential oil plants by vibrational spectroscopy measurements. *Vib. Spectrosc.* 35, 81–86. doi: 10.1016/j.vibspec.2003.12.014

Schulz, H., Özkan, G., Baranska, M., Krüger, H., and Özcan, M. (2005). Characterisation of essential oil plants from Turkey by IR and Raman spectroscopy. *Vib. Spectrosc.* 39, 249–256. doi: 10.1016/j.vibspec.2005.04.009

Seidler-Lozykowska, K., Baranska, M., Baranski, R., and Krol, D. (2010). Raman analysis of caraway (*Carum carvi* L.) single fruits. Evaluation of essential oil content and its composition. *J. Agric. Food. Chem.* 58, 5271–5275. doi: 10.1021/jf100298z

Selseleh, M., Hadian, J., Ebrahimi, S. N., Sonboli, A., Georgiev, M.II, and Mirjalili, M. H. (2019). Metabolic diversity and genetic association between wild

populations of *Verbascum songaricum* (Scrophulariaceae). *Ind. Crops Prod.* 137, 112–125. doi: 10.1016/j.indcrop.2019.03.069

Shafiee, A., and Javidnia, K. (1997). Composition of essential oil of *Zataria multiflora*. *Planta Med.* 63, 371–372. doi: 10.1055/s-2006-957707

Sharififar, F., Moshafi, M. H., Mansouri, S. H., Khodashenas, M., and Khoshnoodi, M. (2007). In vitro evaluation of antibacterial and antioxidant activities of the essential oil and methanol extract of endemic *Zataria multiflora* Boiss. *Food Control* 18, 800–805. doi: 10.1016/j.foodcont.2006.04.002

Simbar, M., Azarbad, Z., Mojab, F., and Majd, H. A. (2008). A comparative study of the therapeutic effects of the *Zataria multiflora* vaginal cream and metronidazole vaginal gel on bacterial vaginosis. *Phytomedicine* 15, 1025–1031. doi: 10.1016/j.phymed.2008.08.004

Stefanaki, A., Cook, C. M., Lanaras, T., and Kokkini, S. (2018). Essential oil variation of *Thymbra spicata* L. (Lamiaceae), an East Mediterranean "oregano" herb. *Biochem. Syst. Ecol.* 80, 63–69. doi: 10.1016/j.bse.2018.06.006

Thompson, J. D., Gauthier, P., Amiot, J., Ehlers, B. K., Collin, C., Fossat, J., et al. (2007). Ongoing adaptation to Mediterranean climate extremes in a chemically polymorphic plant. *Ecol. Monogr.* 77, 421–439. doi: 10.1890/06-1973.1

Thompson, J. D. (2005). *Plant evolution in the Mediterranean* (New York, United States: Oxford University Press Inc.).

Tuttolomondo, T., Leto, C., Leone, R., Licata, M., Virga, G., Ruberto, G., et al. (2014). Essential oil characteristics of wild Sicilian oregano populations in relation to environmental conditions. *J. Essent. Oil Res.* 26, 210–220. doi: 10.1080/10412905.2014.882278

Vaičiulytė, V., Ložienė, K., Taraškevičius, R., and Butkienė, R. (2017). Variation of essential oil composition of *Thymus pulegioides* in relation to soil chemistry. *Ind. Crops Prod.* 95, 422–433. doi: 10.1016/j.indcrop.2016.10.052

Wäschke, N., Hancock, C., Hilker, M., Obermaier, E., and Meiners, T. (2015). Does vegetation complexity affect host plant chemistry, and thus multitrophic interactions, in a human-altered landscape? *Oecologia* 179, 281–292. doi: 10.1007/s00442-015-3347-x

Zgheib, R., Chaillou, S., Ouaini, N., Kassouf, A., Rutledge, D., El Azzi, D., et al. (2016). Chemometric tools to highlight the variability of the chemical composition and yield of Lebanese *Origanum syriacum* L. essential oil. *Chem. Biodivers.* 13, 1326–1347. doi: 10.1002/cbdv.201600061

Ziaee, E., Razmjooei, M., Shad, E., and Eskandari, M. H. (2018). Antibacterial mechanisms of *Zataria multiflora* Boiss. essential oil against *Lactobacillus curvatus*. *LWT* 87, 406–412. doi: 10.1016/j.lwt.2017.08.089

Principles and Applications of Vibrational Spectroscopic Imaging in Plant Science: A Review

Krzysztof B. Beć[1]*, Justyna Grabska[1], Günther K. Bonn[1,2], Michael Popp[3] and Christian W. Huck[1]*

[1] CCB-Center for Chemistry and Biomedicine, Institute of Analytical Chemistry and Radiochemistry, Leopold-Franzens University, Innsbruck, Austria, [2] ADSI, Austrian Drug Screening Institute, Innsbruck, Austria, [3] Michael Popp Research Institute for New Phyto Entities, University of Innsbruck, Innsbruck, Austria

Detailed knowledge about plant chemical constituents and their distributions from organ level to sub-cellular level is of critical interest to basic and applied sciences. Spectral imaging techniques offer unparalleled advantages in that regard. The core advantage of these technologies is that they acquire spatially distributed semi-quantitative information of high specificity towards chemical constituents of plants. This forms invaluable asset in the studies on plant biochemical and structural features. In certain applications, non-invasive analysis is possible. The information harvested through spectral imaging can be used for exploration of plant biochemistry, physiology, metabolism, classification, and phenotyping among others, with significant gains for basic and applied research. This article aims to present a general perspective about vibrational spectral imaging/micro-spectroscopy in the context of plant research. Within the scope of this review are infrared (IR), near-infrared (NIR) and Raman imaging techniques. To better expose the potential and limitations of these techniques, fluorescence imaging is briefly overviewed as a method relatively less flexible but particularly powerful for the investigation of photosynthesis. Included is a brief introduction to the physical, instrumental, and data-analytical background essential for the applications of imaging techniques. The applications are discussed on the basis of recent literature.

Keywords: hyperspectral, multispectral, imaging, near-infrared, FT-IR, Raman, plant, vibrational spectroscopy

INTRODUCTION

From the point-of-view of physicochemical methods of analysis, plants form a challenging subject. As complex, microstructured forms with multi-constituent chemical composition, their comprehensive studies require sensitive and chemically selective methods. On the other hand, conventional sample preparations may result in non-representative results, e.g., chemical clearing or drying as prerequisites for plant sample interrogation. It is preferable to retain the capability of examining living forms in their native state. Having these general remarks in mind, vibrational spectral imaging techniques offer a superb potential. These methods have high chemical specificity, enabling them to construct a chemical image of the sample, within which the distribution of compounds is available even at cellular and sub-cellular level. This may be accomplished in a non-destructive way, and in certain cases with no sample

preparation. In numerous cases, examinations *in vivo* are feasible. In other cases, fractions can be isolated from the plant material for further spectroscopic analysis. Both pathways are suitable for obtaining spectral images that feature characteristic key bands of individual components. These bands yield information on the chemical composition of the plant sample, e.g., structural constituents, or primary and secondary metabolites. The identified chemicals may be used as the markers, further interpreted to discriminate different species, or chemotypes among the same species. Insights into plant microstructures, physiology and biochemistry are available. With use of data-analytical methods, information on various properties of a plant sample may be unraveled and presented in the form of an easily accessible image; often, quantitative or semi-quantitative data on the chemical content may be obtained. These advantages have been well recognized in plant science, with spectral imaging becoming an increasingly popular investigation tool (Elsayad, 2019).

The present review aims to overview vibrational spectral imaging in the field of plant-related research. The methods in scope include infrared (IR), near-infrared (NIR) and Raman imaging (Schulz et al., 2014). Additionally, fluorescence imaging is briefly overviewed; it is based on a different physical principle, yet gained profound use in plant science, e.g., in the studies of photosynthesis. As it delivers information of complementary character, it seems advantageous to compare fluorescence and vibrational imaging, with aim to better expose the advantages and limitations of each of these techniques. The purpose of this work is to dissect the methods and applications in an interpretative way, while assessing their relative usefulness at various directions of research and analysis, in which plant-related samples are the common denominator. Brief introductions to the background phenomena, instrumentation, image generation and data analytical methods, spectra interpretation, and related information are included, while the interested reader is pointed to the referenced literature for more exhaustive information. The discussion of these fundamental topics is directed towards the better understanding of the final performance and applicability of reviewed techniques in plant science. The majority of reviewed applications are based on literature published over the past few years, with selected exceptions that present relevant information or have initiated significant lanes of research.

FUNDAMENTALS AND PRINCIPLES OF VIBRATIONAL SPECTROSCOPY AND IMAGING

Basic Information Related to Spectra Origin and Interpretation

In vibrational (e.g., infrared, IR; near-infrared, NIR; Raman) spectroscopy the interaction with electromagnetic radiation triggers the irradiated molecules between their quantum vibrational states v. The vibrational (i.e., internal) degrees of freedom of molecules correspond to oscillating changes in bond lengths and angles between these bonds, or in other words

stretching and deformation vibrations (modes), respectively. In Raman and IR spectra, the most meaningful are bands resulting from fundamental transitions, in contrast NIR spectra re populated by overtones and combination bands. IR and NIR are absorption spectroscopies, the appearance of a band in the corresponding spectra results from an act of photon absorption. In contrast, Raman spectroscopy probes the vibrational excitations of molecules, although the mechanisms of the interaction with electromagnetic radiation is entirely different (Hammes, 2005). It involves Raman effect, which is an inelastic scattering of photons (Gardiner and Graves, 1989; Hammes, 2005). The symmetry of some vibrations prevents the absorption of photon, leading to so-called (either IR- or Raman-) inactive modes. The probability of absorption, connected with the band intensity (i.e., spectral intensity), is directly proportional to the extent of the dipole moment change, which is the selection rule in this case. Therefore, IR and NIR spectroscopies are particularly sensitive toward the vibrations of polar functional groups. Raman band intensities are dictated by different criteria (Gardiner and Graves, 1989; Hammes, 2005). Raman effect is intrinsically weak, therefore, in order to measure a useful Raman signal, a relatively strong source of monochromatic light is necessary to yield high amount of incident photons. Practical considerations make laser emitting in the visible, near-infrared, or visible/near-ultraviolet range most useful in this role. However, such source is prone to stimulate fluorescence in certain types of samples. This occurs in chlorophyll-rich samples, e.g., plant tissues and related materials. In contrast, fluorescence is a form of luminescence, i.e., spontaneous emission of radiation by a fluorescent molecule (fluorophore) after photon-induced excitation. The spectroscopy based on this phenomenon (fluorimetry or spectrofluorometry) probes electronic and vibrational energy levels of molecules (Lichtman and Conchello, 2005; Kokawa et al., 2015). To stimulate fluorescence, the fluorophore first needs to be electronically excited, e.g., by using UV radiation with wavelengths matching the electronic transitions of the fluorophore. The fluorescent response depends on and is specific to both the electronic and vibrational structure of the fluorophore. Numerous biomolecules found in plants are strong fluorophores, e.g., the ubiquitous chlorophyll or various bio-active compounds found in medicinal plants such as quinine with maximum fluorescence in Vis region.

IR and Raman spectra feature high level of chemical specificity, as the spectral bands are relatively sharp and their positions and intensities can be reliably ascribed to the specific chemical functional groups. This enables to identify the chemical compounds typically present in plant samples (**Table 1**). Exhaustive correlation tables for specific IR and Raman peaks relevant to plant-related samples can be found in literature, e.g., in Schulz and Barańska (2007), or Lin-Vien et al. (1991). NIR spectra preserve to an extent this advantageous feature with absorption bands appearing at predictable wavenumbers (**Table 2**). Yet, in practice, one encounters several effects that prevent straightforward interpretation of NIR spectra. Detailed discussions of this topic can be found in literature (Ozaki et al., 2018; Beć and Huck, 2019). In sharp contrast, fluorescence spectra depend on the excitation and emission properties of fluorophores and are generally complex with

TABLE 1 | Characteristic fundamental bands in infrared (IR) and Raman spectra of chemical compounds commonly found in plants.

Wavenumber in cm^{-1}	Vibrational mode assignment and the associated most characteristic compounds
3,500–3,200	O-H and N-H stretch: carbohydrates, proteins, alcohols and phenolic compounds
2,960–2,950	CH$_3$ asymmetric stretching: mainly lipid with a little contribution from protein, carbohydrate, and nucleic acid
2,930–2,920	CH$_2$ asymmetric stretching: mainly lipid with a little contribution from protein, carbohydrate, and nucleic acid
2,875–2,870	CH$_3$ symmetric stretching: mainly proteins with a little contributions from lipid, carbohydrate, and nucleic acid
2,860–2,840	CH$_2$ symmetric stretching: mainly lipids with a little contributions from protein, carbohydrate, and nucleic acid
1,745–1,730	Saturated ester C=O stretch: phospholipid, cholesterol, ester, hemicellulose, pectin, lignin, suberin/cutin esters
1,650–1,630	Amide I (C=O stretch): protein, pectin, water associated cellulose or lignin, alkaloids
1,630–1,620	C=C stretch: phenolic compound
1,610–1,590	C=O aromatic stretch: lignin, alkaloid
1,560–1,540	Amide II (C=N and N-H stretch): mainly protein
1,515–1,505	C=C aromatic stretch: lignin
1,460–1,455	Amide III (aromatic hydrocarbons): mainly protein
1,455–1,440	C-H asym bending of CH$_2$ and CH$_3$: cell wall polysaccharide, lipid and protein
1,430–1,420	O-H bend: cell wall polysaccharide, alcohol and carboxylic acid
1,380–1,370	C–H sym bending of CH$_2$ and CH$_3$: cell wall polysaccharide, lipid and protein
1,375–1,365	C–H bend: cellulose and hemicellulose
1,250–1,240	C=O stretch: pectic substances, lignin, hemicellulose, suberin/cutin esters
1,235	Amide IV (C=N and N–H stretching): mainly protein
1,235–1,230	C–O stretch: lignin, xylan
1,205–1,200	O–H in plane bend: cellulose
1,170–1,160	C–O–C asym stretch: cutin
1,160–1,150	Symmetric bonding of aliphatic CH$_2$, OH, or C–O stretch of various groups: cell wall polysaccharide
1,145–1,140	C–O–C asym stretch: cellulose (β-1.4 glucan)
1,110–1,105	C–O–C sym stretch: cutin
1,105–1,100	Antisymmetric in-phase: pectic substance
1,085–1,075	C–O deformation: secondary alcohol, aliphatic ester
1,075–1,070	C–O ring stretch: rhamnogalactorunan, b-galactan
1,065–1,060	C–O stretch: cell wall polysaccharides (glucomannan)
1,045–1,030	O–H and C–OH stretch: cell wall polysaccharides (arabinan, cellulose)
990–980	C–O stretch: cutin
900–890	C–H deformation: arabinan
895–890	C–O valence vibration: galactan
875–870	C–O stretch: β–D–fructose

Adapted from *Türker-Kaya and Huck (2017)* under CC-BY 4.0 license.

TABLE 2 | Characteristic near-infrared (NIR) bands of chemical compounds commonly found in plants.

Wavenumber in cm^{-1}	Wavelength in nm	Vibrational mode assignment and the associated most characteristic compounds
8,251	1,212	3 C–H str.: carbohydrates
7,375	1,356	2 C–H str. + C–H def.: carbohydrates
7,168	1,395	2 C–H str. + C–H def.: carbohydrates
6,983	1,432	2 N–H str.: proteins
6,748	1,482	2 O–H str.: carbohydrates
6,662	1,501	2 N–H str.: carbohydrates
6,494	1,540	2 O–H str. (intermol. H-bond): starch
6,394	1,564	2 N–H str.: proteins
6,196	1,614	2 C–H str.: carbohydrates
6,053	1,652	2 C–H str.: carbohydrates
5,896	1,696	2 C–H str.: carbohydrates
5,627	1,777	2 C–H str.: plant fiber composed of cellulose, lignin and other carbohydrates
5,507	1,816	O–H str. + 2 C–O str.: plant fiber composed of cellulose, lignin and other carbohydrates
5,120	1,953	3 C–O str.: carbohydrates
4,878	2,050	N–H sym. str. + amide II: proteins
4,824	2,073	O–H str. + O–H def.: alcohols
4,643	2,154	Amide I + amide III: proteins
4,439	2,253	O–H str. + O–H def.: starch
4,363	2,292	N–H str. + CO str.: proteins

Adapted from *Türker-Kaya and Huck (2017)* under CC-BY 4.0 license.

overlapped band structures forming several distinct intensity maxima (Albani, 2004). Fluorescence profile also depends on the matrix properties and the rigidity of the medium and the measurement conditions as well. For those reasons, the approach to the interpretation of fluorescence is much less systematic than in the previously introduced techniques (Albani, 2004).

Techniques of Spectroscopic Imaging and Microspectroscopy
Basics of the Spectral Image Acquisition
Spectral imaging is a spatially resolved technique able to acquire the spectrum from a specified point at the sample surface. In addition to the spectral wavelength-dependent (Λ) also spatial (x, y)

information is collected from the sample. By adjusting the x, y position, acquisitions of the spectra from multiple points on the sample surface can be performed, assembling a spectral map of the sample (**Figure 1**). Certain techniques also have the ability to acquire the information from beneath the sample surface (z plane). There are essential differences among the generations of spectral imaging instrumentation; primarily, multispectral and hyperspectral techniques need to be separated. In hyperspectral approach, a large number of wavelengths are measured, yielding high spectral resolution. In contrast, in multispectral imaging only a limited number of usually very broad channels are measured whilst simultaneously recording the image. These wavelengths can be tuned for a particular application, which leads to cheaper instrumentation and simpler data processing, but without the flexibility of hyperspectral imaging systems.

The simplest implementation of a spatially resolved spectroscopy is microspectroscopy, in which a reflective optical microscope is combined with a spectrometer. For example, in an IR microspectrometer (i.e., IR microscope) the beam is focused onto a controlled point at the sample surface; this enables acquiring an IR spectrum from an extremely small sample area down to ca. 3 μm of diameter under certain conditions (Colarusso et al., 1999; Larkin, 2018). Such instrumentation uses an optical microscope system for supervision of this process. The IR spectrum may be acquired using different sample presentation modes (**Figure 2**); transmission, diffuse reflection or attenuated total reflection (ATR; i.e., total internal reflection). A chemical image of the sample may be assembled by automated process of registering point-by-point spectra from the intended area of the sample. Most commonly used are the systems utilizing diffuse reflection principle; general schematic of such instrumentation is presented in **Figure 3**. The Fourier
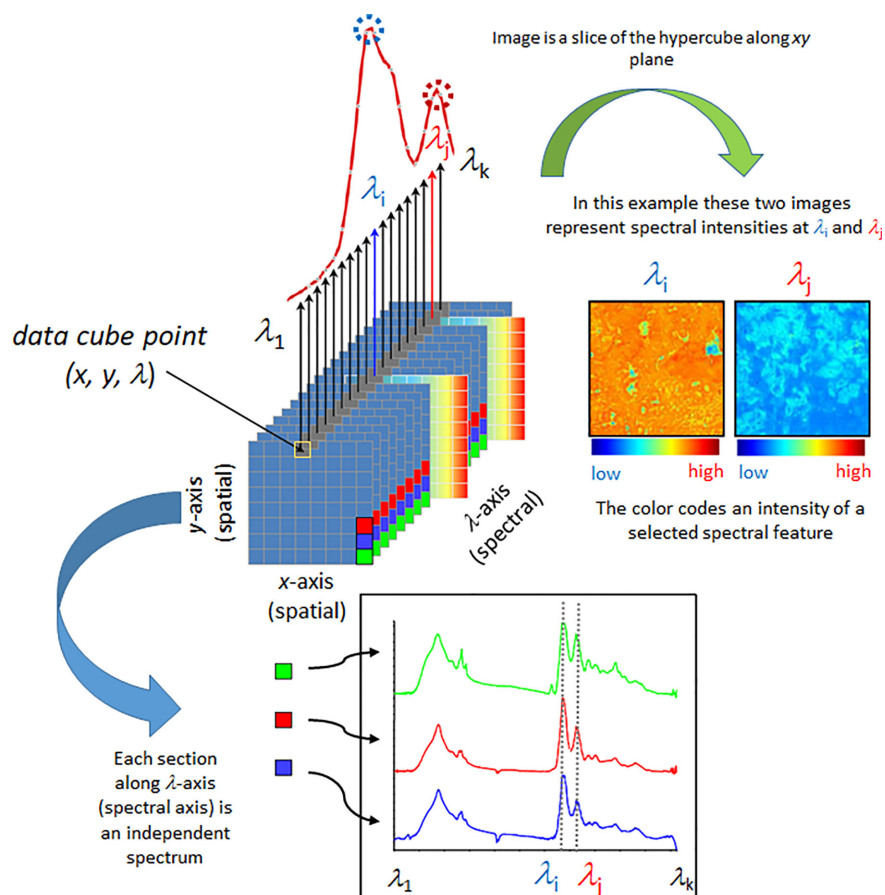
FIGURE 1 | Simplified scheme presenting fundamental principles of a spectral data hypercube and visualization based on the most straightforward spectral information available, spectral intensity. Reproduced with permission from Elsevier (Beć et al., 2020).
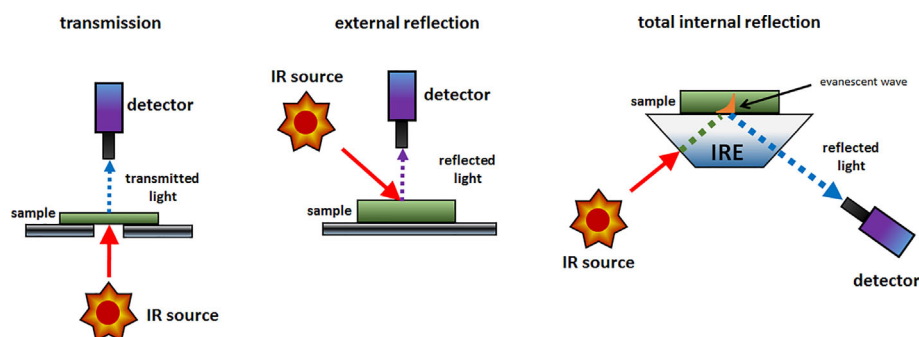

FIGURE 2 | The typical sample presentation modes used in spectroscopic imaging instrumentation.

transform (FT) instruments are preferred as their circular apertures make them better suited for integration with a microscope. Available are FT-IR microscopes utilizing conventional transmission measurements and ATR (or 'micro-ATR') modes as well (Larkin, 2018). Transmission mode microspectroscopic systems, although simple, face considerable

limitations. Mostly, the sample thickness needs to be low enough to prevent a complete absorption of the radiation. This typically limits sample thickness to few μm, while for thicker samples only wavenumber regions featuring relatively weak bands may be acquired reliably. This often implies sample preparation, e.g., slicing, and thus excludes non-destructive way of analysis.
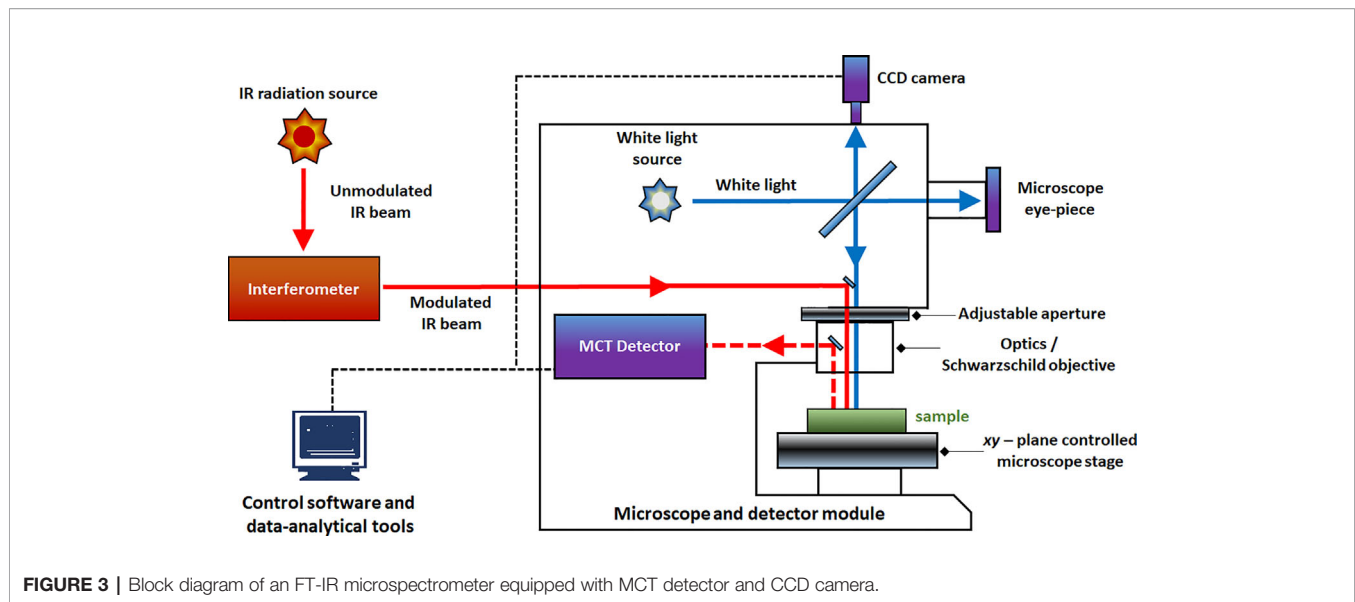
FIGURE 3 | Block diagram of an FT-IR microspectrometer equipped with MCT detector and CCD camera.

Further complications arise if the sample surface features irregularities; its smoothness and flatness are necessary to minimize detrimental optical effects along the optical path of IR radiation (Larkin, 2018). In certain cases, this can be mitigated by encapsulating the sample between IR-transparent optical windows. In ATR approach, there is no such limitation, due the typical penetration depth of the evanescent radiation in μm range (Chan and Kazarian, 2003). However, other kinds of challenges are faced, e.g., the properties of sample-IRE contact surface strongly affect the measured spectra. To an extent, similar rules apply to measurements performed in NIR region. However, ATR approach is no longer feasible here, as the penetration depth is insufficient to yield useful spectral intensity values from weak NIR absorption. Transmission spectra can be reliably obtained from thicker samples. Moreover, in NIR reflection mode, the sample is sensed not only at its surface but deeper into its volume; this can be considered as either an advantage or disadvantage, depending on the aim of the measurement.

Further evolution of spatially-resolved spectroscopy is hyperspectral imaging (HSI). This powerful tool is based on instrumentation capable of simultaneous multi-point spectra acquisition from the sample (Kidder et al., 2006). Various scanning techniques exist in HSI, e.g., spatial scanning, spectral scanning, spatiospectral scanning, and non-scanning (Lu and Fei, 2014; Gowen et al., 2015). HSI measurement involves collection of a large amount of data in short time. Therefore, efficient architecture for data storage and processing are essential. It is commonly accepted to use the hypercube (**Figure 1**), a three-dimensional data structure that is assembled from both spatial and spectral information (x, y, Λ). Since hypercube is a higher-dimension structure, a reduction in dimensionality is necessary to construct a two-dimensional image; despite its critical importance for the informational value in HSI techniques, this is an immensely complex problem and only basics will be presented here. The simplest

way of presenting an image can be accomplished through slicing the hypercube along a given wavelength Λ (e.g., the one selected as meaningful for representing the concentration of a given chemical), with each pixel coded into a false-color according to the spectral intensity, i.e., peak maximum (**Figure 1**). Other spectral features may be used in a similar way; e.g., relative intensity (spectral intensity at a given wavelength in relation to spectral intensity at another selected wavelength), integral intensity (spectrum integrated between selected wavelengths), etc. State-of-the-art image assembly is done through the methods of chemometrics, where pixel colors correspond to quantitatively resolved information on the sample property (Williams and Norris, 2001; Ozaki et al., 2006; Rinnan et al., 2009; Vidal and Amigo, 2012). Note, algorithms integrating spatial and spectral information, as well as combining information from more than one pixel are in use (Gu et al., 2008).

Briefly outlined should be the performance and quality parameters of spectral imaging, as these are crucial for practical applications. Through the introduction of the spatial dimension, additional performance parameters of an instrument need to be defined. The spatial resolution limit is related to the diffraction light limit, light scattering, and focal shifts due to high refractive index samples (Kazarian and Chan, 2010; Larkin, 2018). The diffraction-limited spatial resolution follows Rayleigh criterion, given in Eq. 1.

$$r = \frac{0.61\lambda}{NA} \quad (1)$$

where NA denotes numerical aperture of the optical system and r is half the minimal distance at which two adjacent objects may be distinguished in the acquired image. Note, as seen in Eq. 1, the diffraction-limited spatial resolution is wavelength dependent. This implies physically-dictated difference in the spatial resolution limits between different spectral imaging modalities. Compared with the conventional point-spectroscopy, the

expansion into the spatial dimension and acquisition of spectra from a very small area introduce additional challenges. Optimizations of optical throughput, source efficiency, detector sensitivity, etc, are highly important. On the example of an IR microscope, reflective optics elements yielding high-throughput transmission and focusing optical elements in Cassegrain-type configuration to mitigate optical aberrations are used. Apart of the optics, the image contrast is mostly determined by the source brightness and the type and configuration of the IR detector. Single-point detectors, linear array detectors, or two-dimensional focal plane array detectors are used. Most common IR instruments use conventional radiation sources, e.g., well-known globar (thermal light source based on silicon carbide). Noteworthy, high brightness sources are employed in new generation of instrumentation for IR imaging, e.g., tunable diode lasers, quantum cascade lasers (QCLs), or synchrotrons. High brightness sources vastly enhance spectral quality in terms of the signal-to-noise parameter (SNR or S/N; defined as the level of a desired signal to the level of background noise) and spatial resolution (Vongsvivut et al., 2019); however, such instruments are not yet widely spread and still mostly used for biomedical research. Noteworthy, macro-ATR imaging based on an inverted prism is a particularly potent technique, as it mitigates the limitations of a microscope and enables effective imaging of large sample areas (Kazarian and Chan, 2010). However, poor SNR and difficulties in aligning chemical and visual images are still an issue in ATR-IR imaging. Raman spectroscopy demonstrates advantages in these two parameters, as it combines high diffraction-limited spatial resolution with the possibility of focusing the laser on a very small spot. On the other hand, undersampling may become an issue if the spacing between acquisition points is larger than the laser spot. Unlike IR or NIR spectroscopy, Raman instruments based on FT principle are not the best option for imaging, as these require optical construction that limits the spatial resolution. Practical differences in applicability of these two techniques mostly result from the physical and chemical properties of investigated sample. A brief overview of this issue, as seen from the perspective of plant-related investigations, is presented in the *Advantages and Limitations of Spectral Imaging for Examination of Plant Tissues, Products and Related Materials.*

Spectra and Image Processing and Analysis

Pretreatment, processing and analysis of spectral data have key importance in generating useful image and for the understanding of the encoded chemical information. This topic is immensely complex, and only fundamentals will be briefly overviewed here. For further details, the interested reader is pointed to exhaustive literature (Norris, 1983; Massart et al., 2014; Salzer and Siesler, 2014). Spectral pretreatments are applied to suppress random variations, normalize the spectra against measurement conditions and to enhance the chemically-relevant information; these pretreatments include baseline corrections, normalizations, derivatives and smoothing (Salzer and Siesler, 2014). Image generation covers coding the spectral information into colored pixel to present it in the form useful for analysis by human. Images may be generated through simple univariate approach, in which pixels represent a single characteristic or attribute such as spectral

intensity at a given wavelength; these methods require little processing power and are suitable for rapid generation of large images. However, multivariate analysis (MVA) approaches give far greater potential in elucidating chemical information as they take advantage from the excessive dimensionality of the imaging data. The most common methods include classification methods used to discriminate and group the samples depending on identified spectral variability; popular algorithms are principal component analysis (PCA), linear discrimination analysis (LDA) and support vector machine classification (SVM or SVMC). Cluster analysis methods (e.g., hierarchical cluster analysis, HCA) are often used for elucidating the spatial distribution of the spectral features (Salzer and Siesler, 2014). Plant samples constitute from multiple chemicals; in some cases, if there are only few major components, it may be attempted to decompose (deconvolute) the spectra of each of them. Unmixing approaches such as multivariate curve resolution (MCR methods), e.g., MCR alternative least squares (MCR-ALS) should be mentioned here. Quantitative correlation of the spectral data with reference quantities (e.g., concentration of a given chemical) may be performed with multivariate regression (Burger and Geladi, 2006). Partial least squares (PLSR), multiple linear (MLR), and principal component (PCR) regression algorithms are noteworthy. Once an image is generated, additional processing is available. For instance, cross-pixel information can be used for further gains. Discussions of the relevant topics covering image processing, reduction of data dimensionality and image fusion are available in literature (Burger and Gowen, 2011; ElMasry and Nakauchi, 2016; de Juan et al., 2019).

The Advantages and Limitations of Spectral Imaging for Examination of Plant Tissues, Products and Related Materials

Spectral imaging is an extremely potent tool in plant-related field of research, given its capability for performing *in-situ* non-intrusive compositional and functional analysis in the form of a surface and sub-surface map of the sample (Türker-Kaya and Huck, 2017). Additionally, point-specific quantitative information on the sample is available, e.g., the content and distribution of a certain molecule of interest can be obtained. Given the structural and chemical complexity of plant organs, these capabilities add up to form an outstanding exploratory potential common for all of the techniques reviewed here. Nevertheless, different approaches to imaging that result from the differences in the physicochemical foundations of the spectral techniques, design and engineering principles of the instrumentation, or sampling methods among other factors, separate the applicability of spectral imaging modalities in various plant-related kinds of research. Depending on the sample type, measurement conditions and aims of the analysis, different spectral imaging techniques may be preferable. **Table 3** summarizes the primary parameter ranges of the reviewed techniques, typical values of the working spectral region, spectral and spatial resolution.

IR imaging offers high chemical specificity and relative ease of tracing chemical constituents in the sample. It has reasonably high sensitivity and the instrumentation in its basic form (i.e., IR microscope) is relatively simple. However, IR spectra are easily

Spectral technique	Spectral parameters			Spatial resolution
	Spectral region(wavelength)	Spectral region(wavenumbers)	Resolution	
Vis/NIR	400–1,000 nm	25,000–10,000 cm^{-1}	1.5–5 nm	0.25 – 1.336 µm
IR	3,280–12,500 nm	3,050–800 cm^{-1}	4 cm^{-1}	3.2–10 µm
Raman	2,855–33,333 nm	3,500–300 cm^{-1}	2.6–4 cm^{-1}	0.25–1 µm
Fluorescence	500–800 nm	12,500–20,000 cm^{-1}	3–5.6 nm	0.2–0.45 µm

obscured by the presence of water in the sample, and hence, IR external reflection or transmission techniques are limited in studies *in situ* and in examinations of not dried plant material. Additionally, IR measurements are sensitive to the presence of atmospheric gases [ro-vibrational structure of $H_2O(g)$ and CO_2 (g)], which may become problematic in the setups where the radiation beam travels through open air. Therefore, external measurement conditions need to be controlled with greater care than in other techniques. Moreover, IR imaging instrumentation is rather restricted in spatial resolution because it operates over relatively long wavelengths (**Table 3**). Further, near the limit of the optical resolution the effective optical throughput drops significantly and the SNR parameter decreases. Compared with IR, the loosened diffraction limit resulting from a shorter wavelength of NIR radiation yields a relatively higher contrast of spectral images. Shorter NIR wavelengths enable imaging instrumentation to achieve better spatial resolution. In addition, high spectral resolution is not as much stressed, as NIR bands are broad and loss of chemical information in low-resolution is manageable. This, combined with the availability of rapid scanning detectors, makes NIR imaging instrumentation simpler. NIR and IR imaging techniques are similar to an extent and even the instrumentation capable of working in both regions is available on the market. The key practical differences in applicability of these two techniques root in their physical background. Because of extensive band overlapping, NIR spectra tend to be much more difficult for direct interpretation and chemical specificity of NIR images may vary, depending on the chemical composition of the sample, with enhanced signature of some components (X-H chemical groups). Water has very strong absorption bands in IR region, which overlaps with a number of characteristic bands of organic compounds (**Table 1**). Therefore, IR imaging is not optimal for examining moist samples, while NIR is relatively less affected. Distinct difference between the typical NIR and IR absorption coefficients results in rather deep probing of the sample by NIR radiation (several millimeters beneath surface), while IR tends to probe the surface (penetration depth in µm range).

The characteristics of Raman spectroscopy lead to certain advantages in imaging application. Laser light can easily be focused on a small spot yielding high spatial resolution, e.g., in confocal Raman microscope. Therefore, very small sample volumes can be studied (< 1 µm in diameter, < 10 µm in depth). Theoretical resolution level resulting from diffraction limit are favorable, e.g., a 532 nm laser with a 0.90/100x objective corresponds to a spatial resolution of 0.36 µm; in practice, due to complex optical effects the resolution is ca. 0.5 µm. Confocal

instruments offer unique depth resolution with the possibility to probe the sample beneath its surface. In such case, image acquisition at controlled depth dimension z is available; hence, the collected hyperspectral data has (x, y, z, λ) structure. However, interrogation of plant sample meets additional limitation here, as sharp difference between the refractive indices of the cell walls and cytoplasms, and abundance of pigments and fluorescent molecules disrupts the transmission of the light in the sample. Effectively, under typical conditions the penetration depth of light in plant tissue is limited to ca. 30 µm, which corresponds to the distance of only a few layers of cells. Moreover, the listed phenomena can be detrimental for image quality (Feijó and Moreno, 2004; Paddock and Eliceiri, 2014). For dried samples the danger of thermal decomposition due to energy delivered by the excitation laser at a sample spot should be mentioned. Balance needs to be found between a sufficent intenisty of the Raman signal and heat damage induced in the sample; this factor can be controlled by optimizing the exposure time, laser power, shape and size of the laser spot at the sample surface (Vítek et al., 2020). Moist samples are easily studied, as the spectral information is not obscured by water signal and living plants can be investigated. However, a considerable disadvantage results from stimulated autofluorescence emission from chlorophyll, as its signal tends to obscure all other molecules in 'green' plant samples. One of few exceptions are carotenoids, abundant pigments in numerous plants, for which resonance condition can conveniently be achieved, e.g., their v_1(C=C) band becomes strongly enhanced in Raman spectrum. By performing two independent measurements simultaneously using two excitation lasers with different wavelengths, fluorescence and Raman spectra can be discriminated; however, this approach requires more complex instrumentation (Cooper et al., 2013). Wide-field illumination Raman is a multi-spectral imaging approach, in which only selected Raman shifts (wavelengths) are measured; therefore, the collected chemical information is scarcer. Such instrumentation is simpler, and relatively fragile specimen may be examined as the energy of excitation laser that reaches the sample is dispersed over the scanned sample area (wide field-of-view). The problem of thermal damage induced to sample by the excitation laser has prime importance for studies of delicate plant specimens. Even relatively mild heating at the laser spot occurring in Raman confocal microscope can affect the structure of biomolecules, in particular proteins. On the other hand, the exposition parameters are highly important for yielding quality images. The optimization of the measurement conditions is a continuously developed topic. Recently, Hauswald et al. (2019) dissected the existing strategies for sample illumination and proposed a novel approach that improves Raman imaging quality

versus the thermal illumination limit. The study evaluated the practical applicability of point- and line-confocal microscopes as well as widefield-, light sheet-, and light line illumination, based on the developed models describing the fundamental physical limits of Raman spectroscopy with respect to SNR, sample load and maximum imaging speed. These accomplishments may help to develop new concepts of Raman microscopy, by extending its applicability for the three-dimensional measurement of biological samples including large and sensitive specimens (Hauswald et al., 2019).

Fluorescence imaging occupies a particular spot across the field of plant-oriented studies. On the one hand, the abundance of chlorophyll as a strong fluorophore obscures most of the signal of other chemical constituents in the imaging of plant-related samples. Fluorescence spectra are less specific than IR, Raman, or even NIR spectra. Thus, in this case the ability to obtain chemical profile of the sample is largely inferior to that of the previously discussed techniques. Chemical clearing from chlorophyll and pigments, and subsequent staining with fluorescent die, enables imaging of internal structures through highlighting cell walls. On the other hand, that feature of chlorophyll makes fluorescence imaging applicable to living plants as a sensitive tool capable of monitoring plant's metabolism (chlorophyll fluorescence imaging, CFI). It is a highly accurate indicator of the photosynthetic efficiency, which makes fluorescence imaging a very potent tool in plant science research (Warner et al., 2014), as it unrivaled unique insights into plant phenotypic variation (Meijón et al., 2014), gene expression patterns (Truernit et al., 2008), or plant-microbe interactions (Dagdas et al., 2012). Since the instrumental principle is similar to that used in Raman imaging, with laser excitation, this technique features similar optical advantages. Moreover, three-dimensional fluorescence imaging is available. Proteins are sensitive fluorophores as well, and their subcellular distribution alongside nonproteinaceous cellular constituents can be visualized with this technique.

To mitigate the limitations in light penetration distance through micro-structured, heterogeneous medium such as a plant sample, mechanical sectioning or clearing with chemical agents is routinely applied. These conventional approaches enable imaging of internal plant structures. However, they remove the non-intrusive character of analysis, need to be performed carefully to avoid introducing artificial damages and changes in the sample and, sectioning in particular, are labor and time intensive. Alternative clearing methods with fewer drawbacks have been developed and are discussed in *Fluorescence Imaging*.

APPLICATIONS OF SPECTROSCOPIC IMAGING IN PLANT-ORIENTED STUDIES

IR Imaging
Investigations of Microstructural Features
IR microspectroscopy is a well-established powerful tool used for investigations of the microstructure, chemical composition and functionality of plants at a subcellular level. Despite a relatively

unfavorable level of diffraction limit, it remains within the reach of high-performing IR instrumentation to resolve cellular and sub-cellular structures in plant samples. On the other hand, relatively high chemical specificity of IR spectra gives such studies the necessary fingerprinting capability. For example, Warren et al. (2015) demonstrated their successful application of IR imaging system in resolving individual cells and cell walls in the images acquired from common wheat (*Triticum aestivum*) kernels and *Arabidopsis* sp. leaves. Furthermore, large structures within cells, such as starch granules and protein bodies, were clearly identified. This required sufficient contrast, which was achieved through PCA overlays and correlation analysis applied to hyperspectral data cubes to generate images. Unsupervised PCA algorithm was sufficient to generate a clear image of the sample microstructure (example provided in **Figure 4**), while the correlation analysis enabled confirming the identity of different anatomical structures based on the spectra of isolated components. The proposed approach allowed distinguishing gelatinized and native starch within the cells. Further, the loss of starch during wheat digestion, as well as the accumulation of starch in leaves during a diurnal period could be clearly evidenced in the generated images (Warren et al., 2015). Correspondences between chemical, microstructural and mechanical properties of cell walls were investigated by IR imaging as well, e.g., the properties related to microplasticity were studied by Largo-Gosens et al. (2014).

Cell walls are highly characteristic microstructural features of plant tissues. Noteworthy, cell walls attracted much attention in IR imaging studies relatively early (McCann et al., 1992; Suutarinen et al., 1998). This technique demonstrated particular potential for elucidating structural and functional properties of cell walls. IR spectra in the region of 1200-950 cm^{-1} contain several characteristic peaks of the cell wall constituents (Schulz et al., 2014). For example, Monti et al. (2013) reported that IR absorption at the following wavenumbers can be correlated with the molecular components of cell walls; 1,740, 1,595, 1,440, 1,150, 1,105, and 975 cm^{-1} (pectins); 1260, 1230 and 1075 cm^{-1} (hemicellulose); 1025 cm^{-1} (cellulose); additionally, refer to **Table 1** for group frequencies. This advantageous characteristic of IR spectral techniques can be used with high success to investigate the chemical structure of cell walls (Carpita et al., 2001; Barron et al., 2005; Dokken and Davis, 2007; Gou et al., 2008; Saulnier et al., 2009; Gorzsas et al., 2011; Zhong et al., 2011; Monti et al., 2013; Pesquet et al., 2013). Furthermore, through patterned shifts and intensity variations the absorption bands can deliver information on the chemical environment and intermolecular interactions of the molecules within the cell walls. This feature enables monitoring the alterations in cell wall components in different tissues in connection with physiological changes or throughout plant development and growth. For instance, endosperm textures of hard and soft wheat (*Triticum aestivum*) were studied by Barron et al. (2005). The presence of higher amounts of a water-extractable arabinoxylan in the peripheral endosperm of soft grains was unveiled. Cell elongation, e.g., during growth, affects the molecular structure of the wall. This phenomenon as investigated, e.g., by Carpita et al.
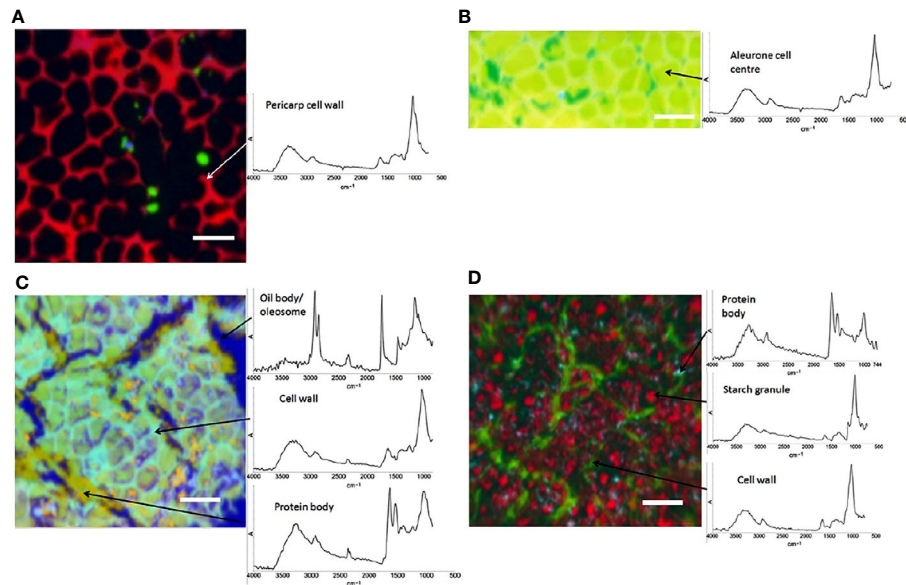
FIGURE 4 | Images of hand-dissected wheat kernel tissues (Consort), shown as false-color principal component analysis (PCA) images, with example spectra provided for regions of interest: **(A)** pericarp-testa; **(B)** aleurone; **(C)** germ; and **(D)** endosperm. Scale bars: 50 µm. Reproduced in compliance with CC BY 4.0 license from Warren et al., 2015.

(2001) in maize (*Zea mays*), or in grains as reported by Saulnier et al. (2009). During embryo generation, it was evidenced that chemical composition of cell walls varies; an increase in cellulose (identified at 900 and 1,320 cm⁻¹) and a decrease in pectin (identified at 1,014, 1,094, 1,152, 1,238, and 1,741 cm⁻¹) was observed by Zhong et al. (2011). These features may be useful markers for imaging studies (Gou et al., 2008). Biochemical changes in cell walls occurring in leaves were observed by Gou et al. (2008) and concluded to correlate with the leaf maturity. Authors observed acetyl esterification of the cell walls in black cotton-wood (*Populus trichocarpa*). For young leaves accumulation of *p*-coumarate was characteristic, while its content decreased in mature leaves. Post mortem, changes could also be monitored by IR microspectroscopy, e.g., lignification of treachery elements of common zinnia (*Zinnia elegans*), in which case the characteristic wavenumbers 1510 and 1595 cm⁻¹ were established as the markers of the chemical changes (Pesquet et al., 2013). Recently, novel insights into cell wall chemistry at cellular level have been obtained by Cuello et al. (2020) using ATR-IR microspectroscopy. They succeeded in non-destructive microphenotyping of the three types of popular wood; normal wood of staked trees, tension and opposite wood of artificially tilted trees. Cell wall composition could be dissected with respect to the cell wall multi-layered structure, gelatinous extra-layer (G-layer), S2 and S3 layers (S-layers). These continuously developing studies evidence the potential or IR imaging to monitor the temporal and spatial patterns in biochemistry, physiology, and microstructural features of plants during their development. IR imaging has become fairly matured tool in investigations of cell walls (Schulz et al., 2014), with research activity being shifted towards applications of Raman imaging technique (Prats Mateu

et al., 2020). Nonetheless, recent attention is given to increasingly accessible IR imaging instrumentation utilizing high brightness radiation sources. The limitation in spatial resolution of the conventional IR microspectroscopy can be significantly lifted by employing synchrotron radiation (SR) source. The improved spatial resolution and SNR compared with conventional IR imaging substantially enhances the potential for examination of microstructures in plant samples, unveiling subcellular details unachievable with traditional approach. In this field of application, SR-IR imaging shares some of the advantages of the techniques based on Raman spectroscopy. As reported by Butler et al., 2017, SR-IR imaging performs notably better in the analysis of living plant tissues, as the detrimental effect introduced by the presence of water in the sample can be minimized using this technique. It eliminates the need for time-consuming sample preparation (i.e., tissue fixing) and directly improves the research potential by enabling studies of plant tissues in native state. This potential was utilized by Butler et al. (2017) as demonstrated by their ability to detect calcium (Ca) deficiencies in *C. communis* leaf samples.

Spatial Distribution of Chemical Composition in Plants

The chemical specificity of IR spectroscopy makes the corresponding imaging techniques applicable directly to study the spatial distribution of biomolecules in plant tissues. The most essential gain here is the determination of distributions of chemical compounds in plant tissues quantitatively. IR imaging was demonstrated relatively early to be a potent tool in this role. For instance, Huck-Pezzei et al. (2012) could successfully identify the characteristic wavenumbers and use them to monitor the

distribution of several classes of major biochemical constituents in tissues of St. John's wort (*Hypericum perforatum*); lipids (1,740 cm^{-1}), phospholipids (1,240 cm^{-1}), proteins (1,630 and 1,550 cm^{-1}), carbohydrates (1,185 to 930 cm^{-1}), and nucleic acids (1,080 cm^{-1}). Furthermore, distribution of these molecules was successfully determined in epidermis, phloem, protoxylem, sclerenchyma, and xylem tissue (**Figure 5**). As evidenced by comparative analyses of their data, in the study by Huck-Pezzei et al. (2012) the spectral data processing and interpretation was essential to resolve reliable and useful information on the spatial distribution of chemical information (**Figure 6**). In sharp contrast to an optical image or univariate spectral analysis, clustering techniques including hierarchical cluster analysis (HCA), k-means clustering (KMC), and fuzzy C-means clustering (FCM) are capable to significantly improve one's ability to interpret IR imaging data collected from plant specimens. In the discussed case, clustering techniques algorithms increased the information content of the IR images dramatically and enabled differentiating morphological and molecular patterns of different tissues. Huck-Pezzei et al. (2012) could semi-quantitatively determine the distribution of chemical ingredients such as lipids, phospholipids, proteins, carbohydrates, nucleic acids, and cellulose in the images of different tissue types.

Similar chemical profiling and characterization of the spatial distribution of several constituents present in *Ginko bilboa* leaves was performed by Chen et al. (2013). Interestingly, that study involved ATR-IR and NIR imaging techniques and demonstrated well the practical difference between both approaches in the analysis of plant tissue. The superior potential of the former one to provide chemical fingerprinting of the sample was noted. Through the analysis of the characteristic IR bands, distribution of, e.g., proteins, saccharides, esters, glycosides, ketones, oxalates, aromatic, unsaturated and long chain aliphatic compounds could be performed. However, the limitation of the elucidated information to the sample surface in ATR-IR technique made use of the deeper sampling characteristic of NIR imaging highly useful in that case. The latter technique enabled rapid exploration of the distribution of the primary chemical constituents in a whole leaf blade. Authors concluded that the combination of both imaging techniques yield the highest exploratory potential. It was also stressed, that ATR-IR approach does not require sample preparation (i.e., microtoming) necessary for IR transmission measurements, which is an essential advantage in plant tissue analysis, as no chemical reagents are used that can change the native composition of the tissue, nor cutting that can mechanically distort the structure of the tissue or cause migration of chemical constituents.
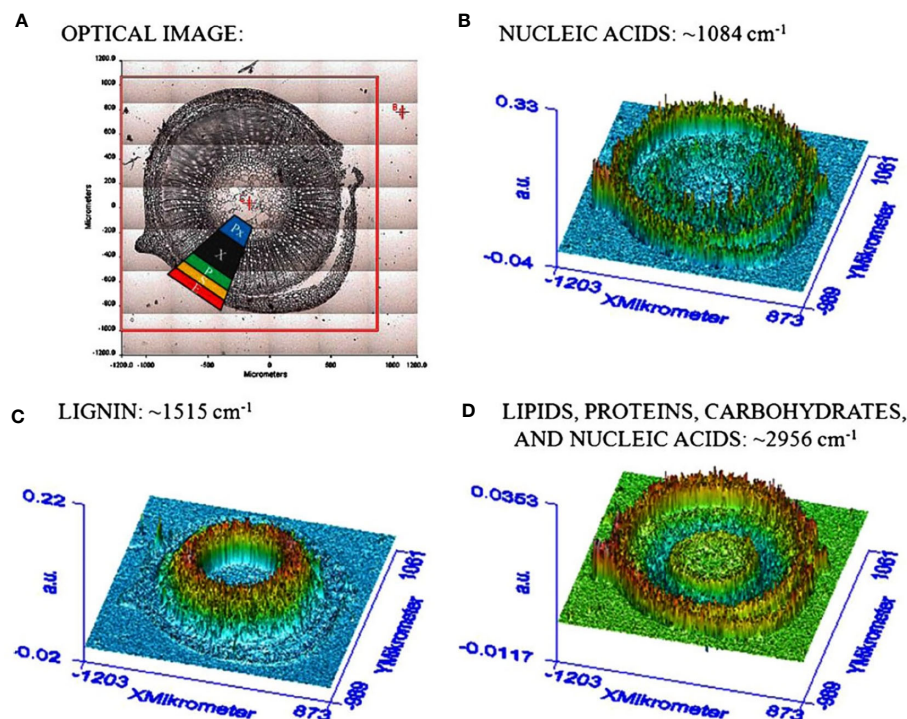

FIGURE 5 | (A) Section through the caulis of St. John's wort (*H. perforatum*) with marked regions (2) obtained by Huck-Pezzei et al., 2012. (e) epidermis, (s) sclerenchyma, (p) phloem, (Px) protoxylem, and (x) xylem. **(B)** FTIR imaging result shown in false color representation. Colors reflect intensities of the selected absorption at 1084 cm^{-1}, which is commonly attributed to nucleic acids. **(C)** FTIR imaging result shown in false color representation. Colors reflect intensities of the selected absorption at 1,515 cm^{-1}, which is commonly attributed to lignin. **(D)** FTIR imaging result in false color representation. Colors reflect intensities of the selected absorption at 2,956 cm^{-1}, which is commonly attributed to lipids, proteins, carbohydrates, and nucleic acids. Reproduced with permission (Springer) from Huck-Pezzei et al., 2012.
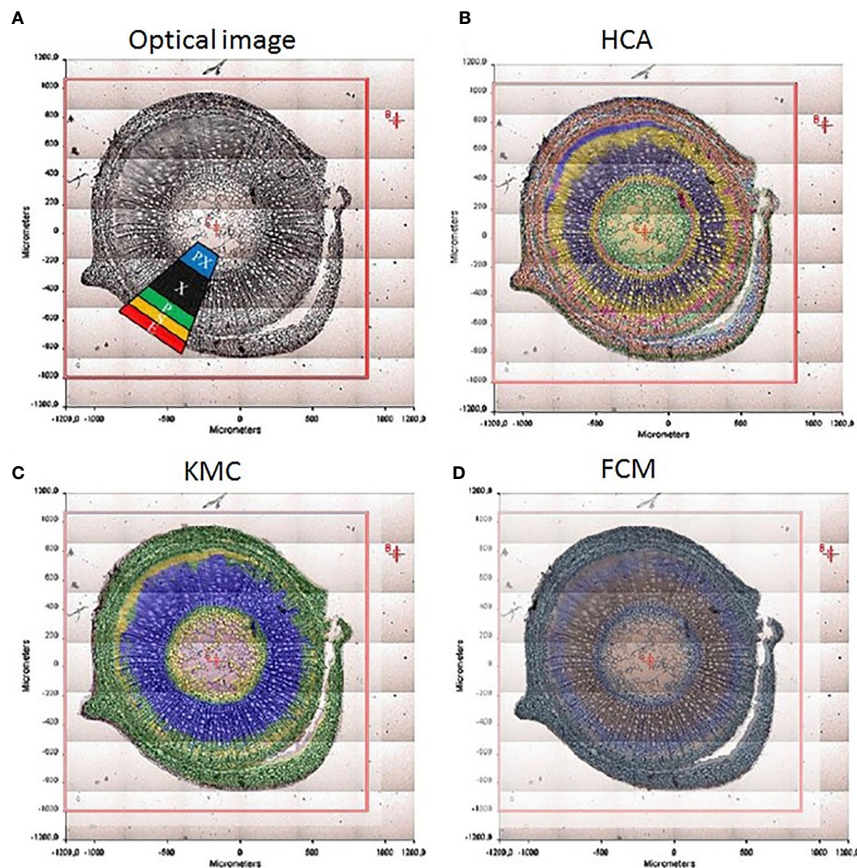
FIGURE 6 | The significance of spectral image processing method for elucidating the chemical information from plant specimen. **(A)** section through the caulis of St. John's wort (*H. perforatum*) with marked regions; **(B)** hierarchical cluster analysis; **(C)** k-means clustering image; **(D)** spectroscopic image of the fuzzy c-means clustering. Reproduced with permission (Springer) from Huck-Pezzei et al., 2012.

The concept of combined use of ATR-IR and NIR imaging spectroscopy was continued by Chen et al. (2015). That study focused on enhancing the spectra analysis methods towards improved fingerprinting capability. In addition to several MVA methods (PCA and independent component analysis-alternating least squares, ICA-ALS; partial least squares target, PLST), two-dimensional correlation spectroscopy (2D-COS) was applied as well. Detailed distribution maps of eugenol and calcium oxalate in tissue sample of calyx tubes (from dried bud) of clove (*Eugenia caryophyllata*) demonstrated the potential of elucidating chemical information from spectral images featuring substantial band overlapping. Noteworthy, it was shown by Chen et al. (2015) that simultaneous application and analysis of both techniques improves interpretability of NIR spectral images. Similar potential was demonstrated in exploring the chemical morphology of areca (*Areca catechu*) nut (Chen et al., 2016).

Investigation In Vivo of the Properties of Biomolecules

Further, high sensitivity and selectivity of IR spectroscopy towards different chemical compounds enables determination

the physicochemical properties of biomolecules *in vivo*. Determination of protein structure is well established, as amide I and amide II bands (**Table 1**) are particularly characteristic markers of that feature (Kumar et al., 2016). Protein structure is sensitive to local environment and can be used to sense the physiological state of the organism. On the other hand, protein structure correlates to nutrition quality of certain crops, and is meaningful for livestock digestive behavior and nutrient availability, which adds up to the topic's significance in agriculture-related research. Relevant examples include the exploration of barley protein by Yu (2006), wheat protein by Bonwell et al. (2008); attention should be given to the applications of synchrotron-based IR imaging technique (e.g., by Yu et al., 2009 and Xin et al., 2013). Highly-sensitive synchrotron radiation-based Fourier transform infrared microspectroscopy (SRFTIRM) was applied by Yu et al. (2009) to investigate protein structures in agricultural plant forage. The distribution of those structures influences nutritional value of protein, as e.g., beta-sheet proteins have limited access to gastrointestinal digestive enzymes. HCA and PCA analyses identified protein alpha-helices, beta-sheets and other structures

such as beta-turns and random coils in SRFTIRM imaging data collected from maize specimens (Yu et al., 2009). Synchrotron IR imaging system operating with high spatial resolution (10x10 μm) was also employed by Xin et al. (2013) to examine *in vivo* protein chemical characteristics and secondary structure and carbohydrate internal structure, with respect to chemical differences in wheat specimens. Normal and frost damaged wheat was examined, with aim to dissect structural variation and frost-induced damage, as these factors have critical importance for the nutritional value of wheat. IR fingerprints of protein and carbohydrates could be identified in the generated images; protein amide I and II bands (ca. 1,774–1,475 cm^{-1}), structural carbohydrates (SCHO, ca. 1,498–1,176 cm^{-1}), cellulosic compounds (CELC, ca. 1,295–1,176 cm^{-1}), total carbohydrates (CHO, ca. 1,191–906 cm^{-1}), and non-structural carbohydrates (NSCHO, ca. 954–809 cm^{-1}). Evidences of frost-induced damage were gathered based on the spectral variations among the studied wheat grain specimens. Frost damaged wheat revealed suppressed amide I and II bands, as well as the bands due to carbohydrate-related functional groups, including SCHO, CHO, and NSCHO. The intensity ratio of protein bands and some of carbohydrate bands was also observed. The study suggested that chemical changes and structural variations in wheat and other grains influenced by climate conditions, such as frost damage, might be a major reason for the decreases in nutritive values, nutrients availability and milling and baking quality in wheat grains. These conclusions are significant for exploring the effects of cultivation conditions and external factors on the protein matrix of agricultural plants, and their resulting nutritional values.

Characteristic IR wavenumbers were successfully used by Gonzalez-Torres et al. (2017) to identify proteins (amide I peak at 1630 cm^{-1}) and carbohydrates (C-O peak at 1030 cm^{-1}) distribution in algae single cells and algal organic matter (AOM) extracts of *M. aeruginosa* and *C. vulgaris*. Elucidation of the composition and distribution of these biomolecules gave insight into the chemical interactions that drive physical floc properties. *C. vulgaris* formed larger flocs characterized by homogenous distribution of proteins and polysaccharides across its area, while smaller flocs of *M. aeruginosa* were observed to develop localized areas of increased protein concentration. The latter features tended to be present near the edge of regions absent of biomolecules, where the coagulant was expected to appear. The interactions between the investigated biomolecules were considered. High Pearson's correlation between carbohydrates and amides was determined in the imaging data of *M. aeruginosa*. Either the presence in the cell surfaces of peptidoglycans characterized by amide groups linked to carbohydrates or the interaction between peptides and carbohydrates at the level of macromolecular pool were suggested as the feasible origins of that observation. In light of the earlier literature, Gonzalez-Torres et al. (2017) hypothesized that proteins not conjugated with carbohydrates form complexes with coagulant in *M. aeruginosa* flocs. Such binding was not observed in the case of *C. vulgaris*, and this was proposed as the main driver in forming stronger, more uniform floc by this species. IR imaging study yielded valuable insight on how the inter actions between biomolecules and the distribution of biopolymer, proteins, and carbohydrates influences cell micro-aggregation in different algae species.

Biochemistry of Adaptive/Defensive Mechanisms in Plants

IR imaging is useful in monitoring biochemical changes underlying the adaptive and defensive mechanisms employed by plants in response to external perturbations, e.g., pathogens. Synchrotron IR imaging combined with atomic force microscope infrared spectroscopy (AFM-IR) was recently employed for investigation of the formation of extractive-rich heartwood (HW) in live trees at cellular level (Piqueras et al., 2020). The role of this natural mechanism is to increase tree's resilience against fungal degradation; however, little was known before about the deposition pathways and the distribution of extractives. In the discussed study, imaging data was acquired from Kurile larch species (*Larix gmelinii* var. *Japonica*) across the HW formation zone sampled through transverse and tangential micro-sections of wood. MCR-ALS deconvolution algorithm was used for unmixing purposes. IR spectral signatures were successfully resolved, with major spectral changes occurring in the transition zone from sapwood to HW. A decrease in the absorbance at ca. 1,660 cm^{-1} and an increase of the absorbance at ca. 1,640 cm^{-1} were identified. Several possibilities were suggested for interpreting this pattern, with type II (Juglans-type) process suggested as the most viable, where an underlying accumulation of phenolic precursors in the sapwood rays precedes extractives oxidation and condensation in the transition zone from which they spread to the neighboring HW cells (Piqueras et al., 2020). Noteworthy, IR imaging provided valuable insight into the importance of the local environment on the response mechanisms at single-cell level (Op De Beeck et al., 2020). While that study concerned fungi (Basidiomycete *Paxillus involutus*), it may be expected that plant cells feature a similar behavior dependent on the microenvironment. This was concluded based on the fungi decomposition activity that was observed to be regulated by the local conditions.

NIR Imaging
Water Distribution in Plant Body and in Soil

Water is an essential substance for functioning of any plant, and the knowledge on its distribution in proximity of a plant brings key benefits for several disciplines of science. It is meaningful for basic understanding of plant physiology, but also substantially aids our ability to counter drought in practical agricultural applications by improving crop water uptake capacity and maximizing the yield. As explained in *The Advantages and Limitations of Spectral Imaging for Examination of Plant Tissues, Products and Related Materials*, the characteristics of NIR spectroscopy make it particularly suited for analysis of the presence of water in sample. In this role, it is far superior to the popular multispectral Vis (RGB) imaging, which is not sensitive towards water signal. On the other hand, NIR light penetrates deeper and is not prone to complete absorption, unlike in IR techniques. These advantages of NIR-HSI approach could have

been successfully employed by Arnold et al. (2016) for examination of the distribution of water in roots of living plants and soil. Compared with approaches based on Vis region, the study confirmed a supreme level of chemical information unraveled in an NIR image of drought-resistance roots. Substantially improved image contrast that directly enabled segmenting roots from soil, discrimination of the essential plant features by their unique spectral characteristics, acquisition of additional information, e.g., on root structure, were concluded as the gains offered by NIR-HSI experiment. The capability to study the water distribution within the body of a plant offers key benefits for plant science, and it may be expected that feasible methodologies towards such goal will continue to develop.

Recent research activity at this direction demonstrated the potential of NIR-HSI to perform spatially-resolved quantitative analysis and visualization of moisture content distribution in tea leaves (Sun J. et al., 2019). A sophisticated suit of data-analytical methods was employed for this purpose, successive projections algorithm (SPA) coupled with stepwise regression (SPA-SR), and competitive adaptive reweighted sampling (CARS) coupled with stepwise regression (CARS-SR). Further, the study involved several spectral image preprocessing methods coupled with feature selection algorithms. Twenty combined treatments were evaluated to yield the best prediction models based on multiple linear regression (MLR). The highest performance of MLR model was obtained when the spectral images were pretreated using Savitzky-Golay and multiplicative scatter correction (SG-MSC) combined with CARS-SR. This approach retrieved best the distribution of moisture content in tea leaves. The established suit of methods for obtaining water distribution maps in the leaf tissue offered new insights essential for improving plant irrigation methods.

While NIR spectroscopy is intrinsically better suited than IR for examining water distribution in biological tissue, it should be mentioned, imaging technique based on terahertz (THz) spectroscopy is a capable tool for this purpose as well (Gente et al., 2013; Gente et al., 2018). Recently, THz time-domain imaging provided three-dimensional mapping of water distribution in a leaf tissue of succulent plant *Agave victoriae-reginae* (Singh et al., 2020).

Analysis *In Vivo* of Macro- and Micronutrients

In plant science, the vast majority of spectral imaging studies were aimed for high throughput plant phenotyping, in which the pursued information corresponded mostly to specimen morphological, such as size or growth, and physiological features (for instance, chlorophyll Photosystem II). HSI technique is capable, however, to determine the spatial distribution of chemical contents within plant specimen as well. For example, Yu et al. (2014) developed a method for determining the spatial distribution of total nitrogen content (TNC) in pepper plant based on Vis/NIR imaging. As a non-intrusive analysis, it has a critical advantage over the conventional destructive methods, such as Dumas combustion method, as it enables *in vivo* studies. The above discussed example concerned an analysis of a single chemical property in

a plant. However, it has been demonstrated by Pandey et al. (2017), that through employing HSI system operating over a broad spectral region covering Vis and NIR (18,182–5,882 cm⁻¹; 550–1,700 nm) it is feasible to acquire *in vivo* quantitative information on various chemical properties. The discussed study examined maize and soybean plants, each in 60 samples, towards water content and macronutrient concentrations. The latter included nitrogen (N), phosphorus (P), potassium (K), magnesium (Mg), calcium (Ca), and sulphur (S), and micronutrients sodium (Na), iron (Fe), manganese (Mn), boron (B), copper (Cu), and zinc (Zn). The specimens were exposed to biotic stress, either through water deficit or through nutrient limitation, to introduce variable properties in the sample set. With exception of Na and B, all other concentrations could have been successfully quantified by PLS models, although the prediction performance for micronutrients was comparatively lower, which seems reasonable giving their lower concentration profiles. This accomplishment demonstrated the potential of Vis/NIR imaging to perform high-throughput nutrient quantification in living plants.

A considerable attention was given to the analysis of nitrate in plant tissue. Nitrates are essential for plant physiology; however, excessive fertilization may lead to an unhealthy high nitrate content in vegetables. Yang et al. (2017) applied NIR-HSI technique to gain better understanding of the spatial distribution of nitrates in spinach leaves. Images were collected *in vivo* and subjected to PLS regression analysis to yield quantitative per-pixel information on nitrate content. The approach was demonstrated to be capable of providing high-resolution maps of nitrate content in different parts of spinach leaves. Detailed display of the nitrate distribution in the petiole, vein, and blade was possible. Noteworthy, the study revealed dynamic changes in the nitrate content in intact leaf samples under different storage conditions. Therefore, the method shows value as a rapid, high-resolution non-destructive tool for quantitative analysis of the nitrate content, and distribution in vegetables.

Development of quantitative analytical methods based on NIR-HSI for elucidating nutrient distribution and dynamics in plants remains an active field of research. Wang et al. (2020) used HSI instrumentation operating in 908–1735 nm wavelength region (11,013–5,763 cm⁻¹) to monitor P and K contents in tea plant leaves. The study involved 87 leaf samples from five different cultivars. A considerable attention was given to the selection of the best performing spectral image pretreatment methods, with prime aim to eliminate the detrimental effect of noise in raw spectral data. Subsequently, the accuracy of the prediction of P and K contents through several MVA prediction methods were compared. The best results were obtained with the use of standard normal variate (SNV) for spectra pretreatment and successive projections algorithm coupled with multiple linear regression (SPA-MLR) to predict the contents of P and K content in tea leaves.

Quality Assessment of Plant Products

Quality control plays a critical role for medicines derived from plants, as the chemical composition of natural drugs is prone to a

much greater variation than conventional pharmaceutical products. In this field, spectral imaging offers substantial improvements over the classical methods. This potential has been demonstrated, e.g., by Sandasi et al. (2014) by their method developed for authentication of *Echinacea* based medicines appearing on the pharmaceutical market. *Echinacea* species are often included in various formulations to treat upper respiratory tract infections. The study involved three species, *E. angustifolia*, *E. pallida*, and *E. purpurea*, acquired from local market in South Africa. By employing NIR-HSI technique operating in 1,0870–3,978 cm^{-1} (920–2,514 nm) region, with aid of PCA it was possible to clearly discriminate between the three *Echinacea* species and further differentiated the roots of different specimens (**Figure 7**). The method accurately predicted the raw material content in several commercially available products and identified products that did not contain crude *Echinacea* material as well.

Herbal teas often constitute of blended plant species and it is critical to be able to determine these constituents in order to assess the tea quality. In sharp contrast to destructive, manpower- and time-intensive conventional routes of analysis (e.g., chromatography), NIR-HSI offers rapid and non-destructive alternative, with addition of rich chemical information and its spatial distribution. It was demonstrated in literature, that NIR-HSI is a potent technique for controlling quality of herbal teas, e.g., as reported by Djokam et al. (2017) on their successful discrimination between tea blends of different brands with varying quality and origin. The quality parameters could have been determined for intact tea bags, which is an essential improvement over the destructive methods of analysis. State-of-the-art in this area is exemplified, e.g., in the study on herbal tea blends (*Sceletium tortuosum* and *Cyclopia genistoides*) by Sandasi et al. (2018). The authors acquired hyperspectral images of the raw material and tea blends in 10870-3978 cm^{-1} (920-2514 nm) region. The images were subsequently analyzed with PCA, which revealed a distinct (54.2%) chemical variation between the raw materials. Next, a partial least squares-

discriminant analysis (PLS-DA) model was constructed with prediction power of 95.8%. Applied for pixel classification, it enabled to visualize the distribution of the constituents *S. tortuosum* and *C. genistoides* and quantitatively predict their contents within the blend tea samples. It was pointed out that HSI instrumentation with a better sensitivity could potentially further improve the quantitative analysis in this and similar cases.

Kava (*Piper methysticum*) is a widely popularized pharmaceutical product containing root extracts. However, kava roots may contain toxins that led to several reported cases of liver intoxication. In consequence, several countries banned import of kava materials and products in the past, and strict quality control procedures were introduced afterwards. The existing methods for this purpose are lacking in high-throughput capacity, and therefore, efforts are made to develop new, more capable strategies. Segone et al. (2020) adopted NIR-HSI system operating in 10870-3978 cm^{-1} (920-2514 nm) region for rapid assessment of six kavalactone biomarkers (methysticin, dihydromethysticin, yangonin, desmethoxyyangonin, dihydrokavain, and flavokawains A and B) content in kava extracts (roots, peeled stems, and stump peelings). The feasibility study was supported by single-point NIR and MIR spectroscopy, as well as HPLC as the reference analytical method. One of the methods' intended purposes was to reliably discriminate kava roots from non-roots samples. PCA algorithm was used to identify spectral variance correlated best with the chemical differences between these two types of sample. The compounds identified as the primarily responsible for the chemical differences between the plant parts were kavain, methysticin, and yangonin. NIR-HSI technqiue achieved good classification performance, with the prediction accuracy extending beyond 90%. The developed approach was reported to be suitable for automation and for continuous operation, fully suitable for practical application as a high-throughput quality control tool.

As the result of imperfect storage conditions, mycotoxigenic fungi may appear in foodstuff produced from plants, e.g., cereals. A significant health risk may result from intoxication by
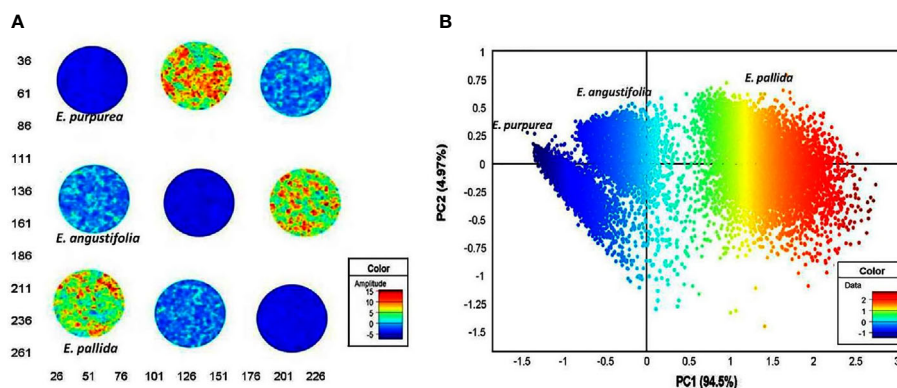

FIGURE 7 | Principal component analysis (PCA) score image (t1) of *Echinacea* sp. leaf powders based on color amplitudes **(A)**. The corresponding score plot (PC1 vs. PC3) shows minimal separation of the pixel clusters **(B)**. (EAL—*E. angustifolia* leaf, EPL—*E. purpurea* leaf, EPaL—*E. pallida* leaf). Reproduced in compliance with CC BY-NC-SA 3.0 license from Sandasi et al., 2014.

mycotoxins contributed by these plant pathogens. The potential of NIR-HSI technique in the role of the detection tool for mycotoxins in cereals was recently reviewed by Femenias et al., 2020. As summarized, the analytical performance of HSI approach is decisively superior to both the conventional methods of wet analytical chemistry and multispectral imaging methods. The capacity for high-throughput cereal sorting combined with quantitative analysis of mycotoxin content (e.g., deoxynivalenol, DON), in the scenario where visual assessment is far inferior, were emphasized as the most significant advantages of NIR-HSI technique used in this role. Nonetheless, Femenias et al. (2020) pointed out that certain limitations, e.g., prediction accuracy and limit of detection, need to be further improved in order to establish this technique as an industry standard.

Conventional methods, often human classification, are the standard routes in the industry for the selection of tobacco leaves according to their quality. Marcelo et al. (2019) demonstrated how application of NIR-HSI and chemometric tools could improve discriminant analysis of tobacco leaves. Imaging data collected for standard tobacco leaf bundles was subjected to rapid MVA classification by means of support vector machine-discriminant analysis (SVM-DA) within 5 s in real time. The developed classification models accurately predicted tobacco's chemical properties important for its quality characteristics in a more robust and reliable way, without subjectivity or bias of a human classifier.

Noteworthy, reliable and objective tools for detection and identification of *Cannabis sativa* are essential for uncovering illegal cannabis plantations. Pereira et al. (2020) developed a method based on NIR-HSI hyphenated to MVA algorithms (sparse PCA and soft independent modelling of class analogies, SIMCA) for this purpose. The method achieved sensitivity and specificity levels of 89.45% and 97.60%, in the case where evaluated samples included leaves of cannabis and similar plants and only 4 spectral variables (spectra points) were required for a sparse SIMCA model. That study evidenced a sufficient classification performance of NIR-HSI technique even when just four spectral bands are captured by the sensor, making it feasible to be implemented in low-cost airborne devices.

Optimization of the Cultivation Conditions

Point-spectroscopy accomplishes significant successes in monitoring the quality parameters of plants and finds use in practice, e.g., for the determination of the optimal harvest time of medicinal plants, which has strong correlation with the content of the bio-active compound in the plant and thus its therapeutic value. For example, Pezzei et al. (2017) in their trend-setting study applied novel miniaturized NIR spectrometer, which is laboratory independent and can operate directly in-the-field, to non-invasively examine common vervain (*Verbena officinalis*). Additionally, the analytical performance of the portable sensor towards quantitative analysis of the key constituents (verbenalin and verbascoside) was compared using a benchtop spectrometer as the spectroscopic reference method. The study revealed that NIR vibrational spectroscopy shows full feasibility for direct measurements of the pharmaceutical active ingredient (PAI) in fresh plant material and enables straightforward determination

of the ideal harvest time of a medicinal plant. It is expected, that with the emerging portable HSI systems, the current envelope will be pushed forward, with the new possibility to extend the information presently available from in-the-field spectroscopy (i.e., chemical composition, PAI content, quality parameters, growth conditions) by adding the ability to unveil spatial distribution of that information in a fresh plant.

Noteworthy, quality of wheat crops expressed by the nitrogen nutrition index NNI (Bowling et al., 1980) may be monitored over wide agricultural areas using novel NIR-HSI systems mounted on unmanned aerial vehicles (UAVs). This high-throughput data can be used with high efficiency for determining the optimal cultivation conditions, e.g., irrigation strategy. Remote sensing approaches advance rapidly and become capable of quantitative analysis. Liu et al. (2020) employed remote sensing setup with flight altitude of 50 m that operated over Vis/SW-NIR region (22,222–10,526 cm^{-1}; 450–950 nm) and spectral resolution of 4 nm to perform quantitative determination of NNI in winter wheat. For reference, the study additionally involved a ground-based NIR-HSI instrumentation with nominally better performance; wide spectral region of 28,571–4,000 cm^{-1} (350–2,500 nm) and maximum spectral resolution of 1 nm. The study showed that varying NNI in wheat yields a clear and distinct footprint in Vis/SW-NIR images, and the spectral resolution is a less critical factor here. The significant and highly characteristic response to NNI was observed in the green and red wavebands was observed. Further, quantitative models of NNI were successfully established and the accuracy of UAV-based imaging was fully satisfactory when compared with the performance achieved by ground-based benchtop instrument. The results collected from wide area monitoring by remote sensing HSI yielded highly useful practical information leading to improvements in the strategy for crop irrigation, to which the optimal approach changes depending on the growth stage of wheat. Rapid progress is made in the instrumentation and data analytical tools used for remote sensing HSI, with satellite-based HSI systems operating in NIR region (e.g., NASA's Surface Biology and Geology mission, Indian Hyperspectral Imaging Satellite) intended for global-wide monitoring of agricultural crops (Aneece and Thenkabail, 2018).

Phenotyping

Various imaging techniques attract current attention as the tools for phenotyping. This topic in a broader context was reviewed recently by Biswas et al. (2020). Special attention should be given here to HSI approaches based on SW-NIR and NIR spectroscopy, owing to particular practical advantages of the technique and instrumentation (*The Advantages and Limitations of Spectral Imaging for Examination of Plant Tissues, Products and Related Materials* and *Simultaneous Applications of Different Spectral Imaging Techniques*). Bodner et al. (2018) proposed a novel approach based on NIR-HSI for characterization of the root system architecture and its functional role in resource acquisition in soil grown plants. The imaging instrumentation operating in the spectral region of 10,000–5,882 cm^{-1} (1,000–1,700 nm) with a spatial resolution of 0.1 mm and 222 narrow detector channels was employed to scan root systems of durum wheat (*Triticum durum*) grown in soil-filled rhizoboxes. In contrast to the more

conventional root phenotyping by multispectral RGB imaging limited to assessing color contrasts between roots and growth media or artificial backgrounds, an imaging technique based on NIR spectral signatures largely improves the quality of information elucidated from the sample, including insights into chemical constituents and physico-chemical properties of roots and soil. To provide phenotyping capability, high-throughput data-analytical tools were optimized for a high degree of automation in image processing. Chemometric analysis with use of unsupervised clustering and thresholding approaches was critical to treat image segmentation and to enhance the effective spatial resolution. This unraveled distinctive radial composition of root axes and their decomposition dynamics. Compared to the routinely used multispectral RGB imaging, the developed high-throughput NIR-HSI approach improved significantly the amount of chemical information and phenotyping capability, although at the cost of more complex and less rapid measurements.

Briglia et al. (2019) examined drought stressed grapevines (*Vitis vinifera*) with aim to establish whether spectral imaging may be used for affordable, non-destructive phenotyping based on information about morphophysiological traits (leaf area, plant water consumption, leaf water potential). NIR-HSI and multispectral (RGB) techniques were evaluated for a possible implementation. That study aimed to identify water-stress combining both morphometric (leaf area) and physiological (water consumption) responses under various drought levels. RGB and NIR imaging techniques were confirmed to be feasible solutions for implementation as affordable phenotyping tools, and a suitable basis for development of new tools for a precision irrigation. As reported, the achieved results are meaningful for future standardization of phenotyping protocols meeting the current goals issued by the global phenotyping community.

Efficient food production in salinized lands is one of the major challenges of modern world. Search for rapid and reliable tools capable of selecting crops tolerant of salinity leads to focus is an active field of research. Non-destructive techniques capable of high throughput plant phenotyping based on morphological and physiological treats are deemed essential for accelerating plant breeding processes in sustainable agriculture. Feng et al. (2020) adopted Vis/SW-NIR-HSI (26,315–9,259 cm^{-1}; 380–1,030 nm) instrumentation to monitor the plant phenotypes of okra (*Abelmoschus esculentus*). Samples from the 13 okra genotypes were examined after 2 and 7 days of salt treatment. Novel approaches for image analysis based on deep learning enabled improved performance in segmentation of plant and leaf. Deleterious effects of salinity disturbing the physiological and biochemical processes in okra were manifested as substantial changes in the spectral information. Vis/SW-NIR-HSI combined with deep learning approach was reported to be highly capable tool for high-throughput phenotyping under salinity stress conditions.

Raman Imaging
Investigation of Plant Microstructure

Raman imaging offers high lateral resolution and confocal resolution, i.e., the ability to perform imaging at a controlled depth beneath the specimen surface. Raman imaging instrumentation based on the excitation laser emitting in NIR wavelength region has made possible investigations of green plant material by mitigating the autofluorescence in the sample. Structural studies on plant cell walls are established using 1064 nm laser-based Raman imaging (e.g., Agarwal, 2014). However, longer wavelength increases the diffraction-based limit of spatial resolution (*Basics of the Spectral Image Acquisition*). Hence, investigations into feasible application of Raman imaging that employs visible (Vis) laser for higher spatial resolution are important. Several studies showed that such systems could be successfully applied for examining micro-structure of plant specimens in their native state with no need for staining or complicated sample preparation. Therefore, this technique finds particular usefulness for exploring the structural complexity of plant organisms. The internal microstructure of plants is mostly defined at the molecular level by the cell walls. These consist of semi-crystalline cellulose fibrils of thickness in few µm range, which are embedded in an amorphous matrix constituted of biopolymers (pectins, hemicelluloses, and lignins). Structural arrangement of these constituents within the cell wall varies among different plant tissues; therefore, techniques capable of investigating the spatial inhomogeneity of this feature are essential. Recent review articles covering this topic should be noted (Durrant et al., 2019; Zhao et al., 2019; Prats Mateu et al., 2020).

Raman imaging operating at µm resolution has been established as a particularly potent tool for such purpose, as e.g., demonstrated by Gierlinger (2014) in her study of wood specimens. Two-dimensional Raman mapping with Vis (532 nm) excitation laser was performed for micro-cross-sections of spruce (softwood) and beech (hardwood). The imaging data was analyzed using both univariate (band integration, height ratios) and multivariate methods (vertex component analysis, VCA). It was concluded that MVA approach yielded superior information, as VCA algorithm successfully separated anatomical regions and cell wall layers according to the most different molecular structures. In comparison, univariate approach only visualized changes in selected band heights or areas with weaker correlation to the morphological features. With high spatial resolution (<1 µm), the investigation revealed subtle changes in lignin composition and content in the cell walls, which has been ascribed to a non-uniform lignification during growth of the specimens. Thus, detailed information on the inhomogeneous distribution of structural features of plant cell walls was derived. Alongside gains for basic plant science, these results had potential practical importance for optimizing the utilization of plant biomass (Gierlinger, 2014).

The application of Raman imaging technique for examination of the chemical and structural properties of cell walls remains a particularly active and fruitful area of research (Durrant et al., 2019; Zhao et al., 2019; Prats Mateu et al., 2020). Recent investigation by Zeise et al. (2018) demonstrated how high-resolution spatial distribution of chemical composition available from Raman microspectroscopy can be used to gain deeper insights into the properties of microscopic sub-structures of

the plant tissue, e.g., the building blocks of the cell walls. The study focused on cucumber (*Cucumis sativus*) with the experimental based on 55 Raman maps of root, stem, and leaf tissues. Through spectra analysis carried out with both univariate and multivariate algorithms, different spectral contributions from cellulose and lignin could be unraveled in Raman images. Contributions from the main cell wall components, lignin and cellulose, were suitable for assembling univariate (chemical) images of the sections and revealed substructures of the cell walls in the xylem tissue (**Figure 8**). Further, images constructed through MVA with hierarchical cluster analysis (HCA) and principal component analysis (PCA) algorithm identified different substructures in the xylem cell walls among different

tissues and visualized the cell wall substructures more clearly. Noteworthy, the laser excitation wavelength range (532 nm) enhanced the signal from carotenoid species (at 1,523 cm⁻¹, 1,156 cm⁻¹, and 1,005 cm⁻¹) through resonance effect. Zeise et al. (2018) presented different possible approaches of generating Raman images with high contrast and resolution suitable to analyze morphological information acquired from sections of native, unembedded root, stem, and leaf tissues of cucumber plants.

Raman imaging technique offers great potential for monitoring of structural and compositional changes at plant cellular level. With spatial resolution reaching down to ca. 300 nm, Raman imaging is capable of elucidating molecular structural features of plants at cellular level. For example,
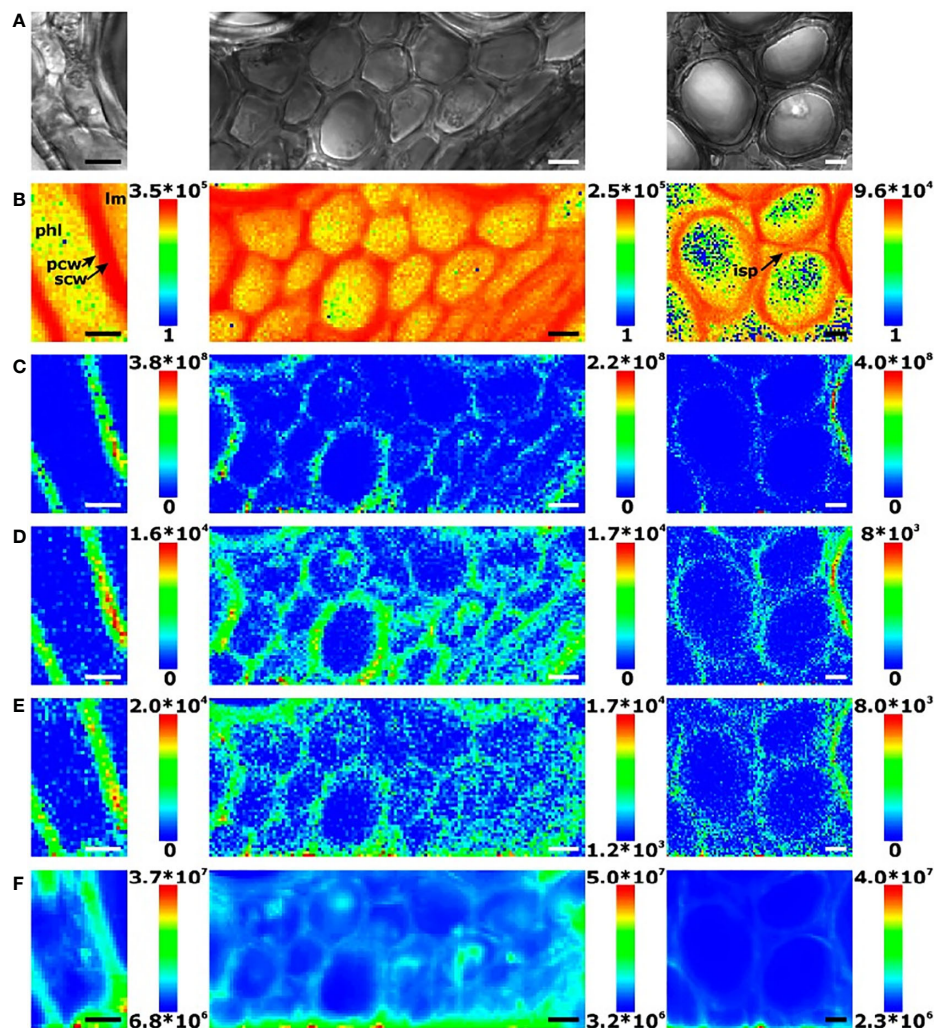

FIGURE 8 | Raman chemical images of three exemplary mapping data sets of cross sections of cucumber (*Cucumis sativus*) stem xylem. **(A)** Bright field images; **(B)** Integral intensity in the region 1,550–1,700 cm⁻¹, obtained after baseline correction; **(C)** Product of the intensities of the cellulose bands at 1,092 cm⁻¹ and 1,337 cm⁻¹, respectively, cf. **(D, E)**; **(D)** Intensity at 1,092 cm⁻¹ (baseline corrected in the range 1070–1108 cm⁻¹); **(E)** Intensity at 1337 cm⁻¹ (baseline corrected in the range 1,313–1,358 cm⁻¹); **(F)** Intensity integrated over the full spectral range 600–2,000 cm⁻¹. lm, lumen; phl, phloem; scw, secondary cell wall, pcw: primary cell wall; isp, intercellular space. Scale bars: 10 μm, mapping step size: 1 μm, excitation wavelength: 532 nm, excitation intensity: 1.7 × 106 W/cm², accumulation time: 1 s. Reproduced in compliance with CC-BY 4.0 license from Zeise et al., 2018.

Gierlinger et al. (2012) conducted a trend-setting study on chemical imaging of plant cell walls at sub 0.5 μm level by confocal Raman microscopy. Embedding and microcutting procedures of sample preparation were developed with aim to preserve plant tissues with intact cell walls. Alongside, data-analytical approaches were designed to present the images and to resolve molecular fingerprints among the multiple components appearing within the native cell walls. The study provided insights into polymer composition as well as the orientation of the cellulose microfibrils. The potential of this technique includes the ability to follow changes occurring in the structure and chemical composition of plants at single cell level; these temporal features can be monitored within different cells as well as between them. Research and results obtained at this direction were summarized in detail by Prats Mateu et al. (2014).

Li et al. (2020) employed confocal Raman microscopy to investigate the distribution of water content in apple tissues at the cellular level with valuable insights into the water molecular state with respect to hydrogen bonding (HB). The spectra of fresh apple tissues featured five peaks in the region of 3,000–3,800 cm^{-1}, that were assigned to the OH stretching mode of water molecules in different local environments resulting from the difference in the hydrogen bonding states. Interestingly, water molecules with the strongest and the weakest hydrogen bond (corresponding to the peaks at 3,050 and 3,630 cm^{-1}) were identified to be primarily located in the cell wall areas.

This observation demonstrated the potential for elucidating useful information on the water migration in plant tissue as cellular level, as well as practical value for molecular effects of food processing, e.g., freezing or drying.

Raman imaging shed light on the physiological changes underlying wood degradation in the study by Belt et al. (2019). The investigation focused on identifying the key steps in heartwood degradation in the incipient stages of brown rot decay in scots pine (Pinus sylvestris). The study unveiled that the degradation of heartwood begins in the innermost cell wall layers and then spread into the remaining cell walls and the middle lamella. One of the most prominent of the observed Raman spectral changes were identified as the decrease in the intensity of the bands due to pinosylvins. Further, this enabled monitoring an extensive degradation of pinosylvins in the cell walls, middle lamella and extractive deposits were observed. Other changes were observed as well, leading to the conclusion that the key driver of the incipient heartwood decay is the degradation of antifungal heartwood extractives, accompanied by the degradation of the inner cell wall and introduction of degradative agents into the cell walls and middle lamella.

Characterization of Spatial Distribution of Carotenoids

Carotenoids are common pigments abundant in numerous plants and algae as well as bacteria and fungi. They play an important physiological role in plant biology, as they form a protective measure against photodamage from overexcitations and are critical in absorbing energy of incident radiation and thus are essential for photosynthesis. Raman spectroscopy and imaging was extensively used for investigations of carotenoids in plant tissues (Schulz et al., 2014; Pećinar, 2019). Oleszkiewicz et al. (2020) provided a detailed example of the sample preparations aimed to preserve intact tissue for subsequent analysis using Raman spectroscopy. Characteristic bands of several important carotenoids (e.g., β-carotene, lutein, and lycopene) are located at the following wavenumbers; v(C=C) peak at ca. 1535-1500 cm^{-1}, v(C-C) at ca. 1,145–1,165 cm^{-1}, v(C–CH$_3$) at ca. 1,010–1,000 cm^{-1}. These bands are sensitive markers of the molecular structure of a carotenoid, band shift deliver information on the number of conjugated bonds and the side groups in the molecule, as well as its interaction with the chemical environment (i.e., matrix). Carotenoids are uniquely meaningful for Raman imaging studies of plants because of the resonance-based enhancement of their v(C=C) signal in Raman spectra. This gives the possibility to distinguish between the distributions of different carotenoids in Raman images. The study by Schulz et al. (2005) may serve here a classical example in which Raman imaging could be successfully used to extract individual distribution of carotenoids with 7, 8, and 9 conjugated bonds from the intact tissues of pot marigold (Calendula officinalis). As another example, β-carotene distribution in sweet potato, carrot and mango was examined by Brackmann et al. (2011) using imaging system based on coherent anti Stokes Raman scattering (CARS) principle. Sensitive probing of the β-carotene was achieved by monitoring its characteristic signal at 1,520 cm^{-1}. This enabled unraveling different macro-structural assemblies in each species. Sweet potato and carrot features rather densely accumulated β-carotene forming heterogeneous rod shaped structures, while homogeneous aggregates of carotenoid filled lipid droplets were identified in mango.

High-resolution Raman imaging can be used for highly detailed in situ characterization of carotenoid properties, as demonstrated by Roman et al. (2015). The conclusions drawn from that study shed light on the distribution and composition of crystalline and amorphous carotenoids in carrot cells. Crystalline carotenoid domains contain α-carotene, although the presence of lutein could not be excluded. In contrast, amorphous carotenoids are composed of β-carotene molecules. Additionally, through the band shifts observed in Raman spectra, it was concluded that amorphous carotenoids are involved in intermolecular interactions with other plant constituents, e.g., proteins or lipids. These accomplishments displayed the differences present in the carotenoid content in carrot with respect to crystalline and amorphous state. When combining Raman imaging with other techniques, further insights into the structure and properties of crystalline phase carotenoids in carrot root can be elucidated, as shown by Rygula and associates (2018). For the first time, that study unveiled the chemical and structural differences of carotenoid crystals, including the varying composition of the crystals. Evidence was presented on the uniform chemical composition of the crystals, regardless of their planar structure. The exception was observed for the helical crystals, where α-carotene was deemed absent.

Carotenoids play a role in the biochemical adaptive protection mechanisms of certain plants. Algae serve as highly suitable model plant organisms and are important part of the ecosystem. Therefore, they attract considerable attention in plant science (Coman and Leopold, 2017). Similar to various other algae, upon light irradiation *Chlorella* species (*C. protothecoides* and *C. vulgaris*) produce and accumulate large amount of carotenoids within their cells. This phenomenon was investigated by Grudzinski et al. (2016) using Raman mapping technique. It was found that light induced yellowing of *Chlorella* sp. results from the accumulation of xanthophylls, primarily zeaxanthin. It was accomplished at good sensitivity and selectivity levels, through arranging selective Raman resonance conditions using two spectral acquisitions with 488 and at 514 nm excitation lasers. At both of these wavelengths a resonance condition for xanthophyll pigments is achieved, while for zeaxanthin this occurs only for 514 nm line. It was revealed that under strong light exposure conditions carotenoids are formed at cell nucleus, with additional qualitative information about zeaxanthin being the major synthesized carotenoid. This study revealed an adaptive mechanism of plant against intense radiation in visible region, referred by the authors as 'the molecular sunglasses' presumably intended to shield the sensitive cell nuclei. These conclusions brought important insights into the adaptive mechanisms of algae against overexcitation. Further, practical gains from this discovery is the potential cultivation of *Chlorella* for use of as the alternative source of zeaxanthin, as it a macular pigment suiting protective role in human eyes against the age related macular degeneration (Grudzinski et al., 2016).

Noteworthy, alongside the major constituents, Raman imaging is capable of elucidating the distribution of other class of compounds within plant's body (Coman and Leopold, 2017). Polyacetylenes can be monitored at ca. 2100-2300 cm^{-1} through the ν(C≡C) vibration. For instance, Barańska et al. (2005) studied the distributions of falcarinol and falcarindiol in carrot and concluded that their accumulation occurs primarily in the outer sections of the roots.

Raman mapping was employed by Sharma et al. (2015) for unraveling the localization of major types of biomolecules within the cells of *Chlamydomonas reinhardtii* microalgae. The distribution of proteins, lipds and carotenoids was performed *via* their characteristic wavenumbers, at 1,003, 1,445, and 1,520 cm^{-1}, respectively. The generated Raman images highlighted the lipid rich, carotenoid rich and protein rich areas in the cells. Further, characterization of lipids enabled identifying among the mutated cells those with the increased lipid content.

Fluorescence Imaging

As outlined in Introduction, the aim of the present section is to briefly overview the most recent applications of fluorescence imaging, and on this background, to better present the potential of vibrational spectral imaging methods in plant science. CFI technique has found significant use in investigation of the influence of biotic stress delivered to living plants in various ways. Photosynthesis, as a metabolic process critical for plant physiology, is distinctly affected by external factors, e.g., exposition to pathogens and environmental aspects (Pérez-Bueno et al., 2016). This includes manipulations of the plant metabolism induced by the pests, but also plant's own defense mechanisms, in which photosynthesis rate is reduced in order to limit the availability of nutrients to the pathogens. Regardless of the cause, biotic stress typically causes spatially and temporarily heterogeneous effects, which can be elucidated by CFI through monitoring photosynthetic activity at cellular level.

Drought and salt stress are extremely meaningful sources of biotic stress in plants, and their primary significance for public stems from the impact they have on agriculture. These conditions perturb the functional properties of photosystem II (PSII) and lower the amount of energy available from the process, to which plants respond by reducing biochemical activities (Zhu, 2002). Further side-effects may be noted, including oxidative stress, osmotic stress, increased ion toxicity and disturbed homeostasis of Na+ and K+ cations (Zhu, 2002; Sun D. et al., 2019). Therefore, detailed knowledge on the metabolic effects these factors induce on agricultural plants is essential for the attempts to improve breeding programs towards better resilience to drought. CFI techniques can form extremely helpful tool in such research, as demonstrated, e.g., in a recent study by Sun D. et al. (2019). They examined the behavior of salt overly sensitive (SOS) mutants of thale cress (*Arabidopsis thaliana*) as a model plant using a time-series CFI analysis to dissect the chlorophyll fingerprints of salt overly sensitive (SOS) mutants under drought conditions. The investigation employed a potent set of data-analytical tools to unravel the chlorophyll fingerprint and its patterned change under drought and salt stress conditions. Authors employed PCA to resolve the shifting pattern of different genotypes in the examined specimen including SOS mutants and wild type plants. Subsequently, temporal features were elucidated using sparse auto encoders (SAEs) neural network, a time-series deep-learning algorithm. The resulting data were used in chemometric classifications based on linear discriminant analysis (LDA), k-nearest neighbor classifier (KNN), Gaussian naive Bayes (NB) and support vector machine (SVM), with very good accuracy of discrimination between the specimen. In addition, the authors employed sequential forward selection (SFS) algorithm was used to identify the most characteristic chlorophyll over-time responses to drought stress of each specimen. The complex workflow designed by for their study is presented in **Figure 9**. The accomplished results demonstrate the suitability of CFI approach supported by sophisticated data-analytical procedures, to monitor the gene function underlying plant's physiological response to biotic stress induced by drought and increased saltiness of the environment (Sun D. et al., 2019).

Noteworthy are selected earlier developments, e.g., those aimed at improving deep-tissue capability of plant imaging. Conventional approaches to this problem require problematic sample preparations, as explained in *The Advantages and Limitations of Spectral Imaging for Examination of Plant Tissues, Products and Related Materials*. Warner et al. (2014) developed an improved method with aim to simplify deep-tissue

imaging by fluorescence and make it suitable for examining intact plant organs and whole plants. The method was based on an alternative approach to chemical cleaning intended for greater tissue transparency and light transmission while preserving the ability to use common fluorescent stains and proteins. The procedure was based on clearing solutions with lowered intrusiveness, which lead to better preservation of subcellular features, enhance light transmission through the sample, and retain the ability to use common fluorescent stains and proteins. Warner et al. (2014) reported that their approach enabled successful imaging of fine cellular features in specimens, while maintaining the refractive indices of cleared plant tissues at the close level to that of untreated specimen.

Considerable attention should be given to the emerging applications of CFI as a part of integrated remote sensing techniques with potential to advance modern precision agriculture. In such applications, the pursued goals are to monitor large areas of crop fields in combination with the immense capabilities of spectral imaging methods. Unlike the typical applications of CFI, mostly intended to examine plant defense responses to stress factors, novel concepts focus on the ability to perform plant phenotyping, e.g., as reported by Pérez-Bueno et al. (2016). That study investigated the feasibility of applying multicolor CFI (MCFI) in combination with thermography for remote sensing and phenotyping purposes. Such abilities depend on processing large and complex data sets, and the study by Pérez-Bueno et al. (2016) included development of data-analytical methods valid for intended tasks. In the

process, a number of statistical models were evaluated, including logistic regression analysis (LRA) and artificial neural networks (ANN). These models have subsequently been validated in real life scenario and demonstrated the performance comparable to the established conventional techniques; however, the developed method of MCFI in combination with thermography is superior in scalability.

CFI has been demonstrated to be a potent tool for monitoring the damage induced to plants by environmental pollution by heavy metals. The process of photosynthesis appears to be particularly sensitive to appearance of heavy elements, cadmium in particular. Bayçu et al. (2018) conducted a CFI-based study to provide new data on the mechanism of plant acclimation to exposure to Cd element (Parmar et al., 2013). They examined how the PSII of *Noccaea caerulescens* behaves during plant acclimation to the exposure conditions. CFI technique enabled a non-invasive visualization of the spatiotemporal variations in PSII efficiency and offered insights into the mechanism of PSII acclimation to Cd exposure. Further insights into cadmium effects could have been obtained by Moustakas et al. (2019) by combining CFI with a technique of laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS). With aim to unravel the impact of Cd accumulation on the plant's PSII photochemistry, they examined clary sage (*Salvia sclarea*) with focus on following the decrease of its photosynthetic efficiency as a result of Cd exposure. CFI and LA-ICP-MS could be successfully combined to monitor heavy metal effects and plant tolerance and the
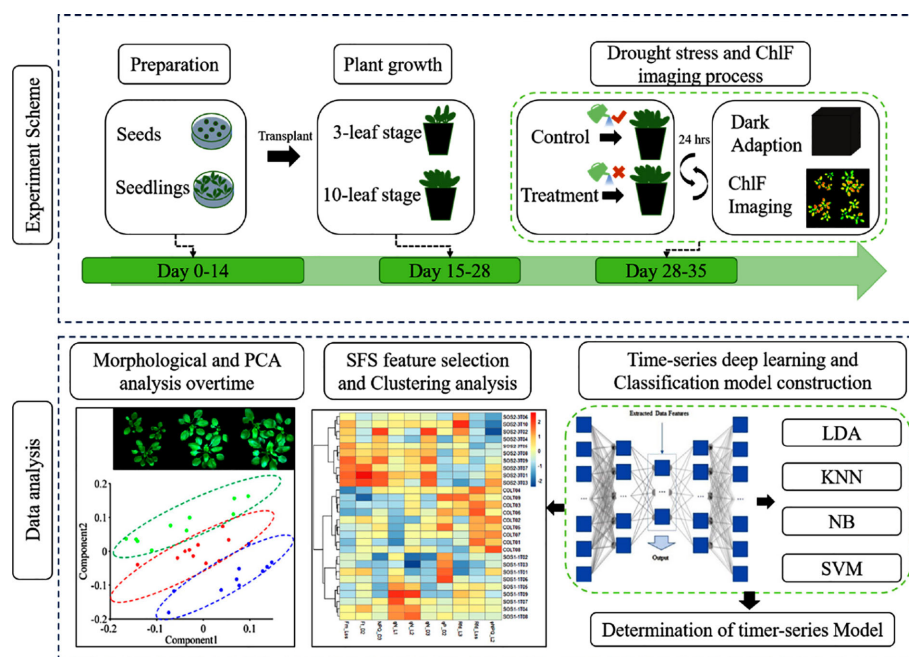

FIGURE 9 | Workflow of using chlorophyll fluorescence imaging (CFI) technique to dynamically monitor photosynthetic fingerprints caused by salt overly sensitive genes under drought condition. LDA, linear discriminant analysis; KNN, k-nearest neighbor classifier; NB, Gaussian naive Bayes; SVM, support vector machine. Reproduced in compliance with CC-BY 4.0 license from Sun D. et al. (2019).

mechanisms the plants develop for Cd detoxification. CFI enabled determining the spatial distribution of the heterogeneous changes occurring in the plant leaf. The study revealed that *Salvia sclarea* tends to accumulated Cd and exhibits good tolerance level to the presence of this element and may be used in the role of a phytoremediant plant (Moustakas et al., 2019).

Simultaneous Applications of Different Spectral Imaging Techniques

It may be noted, that the imaging instrumentation and spectra processing tools achieved a reasonable level of maturity nowadays. Further progress likely depends on improving the ability to interpret complex chemical signatures typically encountered in plant samples. In order to improve the chemical specificity and interpretability of spectral images, applications of complementary techniques is highly promising. For the reasons explained in *Basic Information Related to Spectra Origin and Interpretation*, a good combination is formed by IR and Raman imaging, as correlations between the spectral features can mutually aid interpretation of both types of imaging data. As demonstrated by Gierlinger (2018), interrogation of a plant specimen by FT-IR and Raman confocal microscopy leads to a deeper understanding of biological processes and structure–function relationships occurring therein. The combined techniques largely increase the chemical specificity and interpretability of the acquired images (**Figure 10**) offering better understanding of critical factors, e.g., plant cell wall chemistry or biochemical functions of microstructures. The complementary character of IR and Raman imaging techniques in the context of

plant science has been well summarized by Gierlinger (2018). It is our opinion, that NIR spectroscopy adds significantly to this combined potential, particularly owing to significant progress that is being currently achieved towards interpretability of NIR spectra. The synergy between the imaging techniques based on different modalities of vibrational spectroscopy is presented in **Table 4**. At the same time, MVA techniques such as unmixing methods (e.g., vertex components analysis) are helpful in addressing overlapping bands, which are common in the spectra of biochemically complex samples. Recent progress in developing unmixing algorithms with the purpose of improving the interpretability of Raman imaging of plant cell walls is noticeable (e.g., Prats Mateu et al., 2018). Over the current decade, one may anticipate increased attention paid to improving our understanding and ability to interpret the chemical information entangled in spectral images acquired from plant specimens.

Worth emphasizing are investigations that combine vibrational and fluorescence spectral imaging to elucidate even more complete information from the sample. Recently, a trend-setting investigation was reported by Vítek et al. (2020). In this study, the authors combined different imaging techniques (Raman, CFI and IR thermal) to resolve the spatial and temporal changes that occurs as a plant's response to PSII inhibition. A herbicide metribuzin was used as the inhibitor of the PSII in a model plant species *Chenopodium album*. High-resolution Raman imaging unveiled zones of local increase of carotenoid following the application of herbicide. The presence of carotenoids was highly correlated with the damaged tissue over time, as a result of the activation of defense mechanisms (**Figure 11**). Further, the Raman shift in the carotenoid band was a clear marker of the structural changes in carotenoids. CFI technique was employed to elucidate the spatial- and time-dependent variations in the quantum efficiency of PSII (**Figure 12**). In particular, it was possible to observe the spatial distribution of the key parameters of photosynthesis (F_v/F_m and NPQ) and to unveil that the movement of herbicide acropetally (in direction to the leaf tip) mostly through main veins occurs within hours from the application of metribuzin. CFI demonstrated that the herbicide affects sharply defined parts with no transition areas. IR thermal imaging enabled observing patterned changes of leaf temperature induced by the herbicide (**Figure 12**), in relation to control specimen untreated with metribuzin, with the temperature elevating from 0 to above 5°C during the period 96 h after the application of the agent. The observations based on IR thermal imaging remained in agreement with those made in the CFI experiment; the temperature increase was relatively greater towards the upper part of leaf, confirming the acropetal movement of the herbicide. Noteworthy, this technique was concluded to be inferior to chemical imaging, as the herbicide transport through veins could not be monitored, neither the areas affected by the agent were as well-defined.
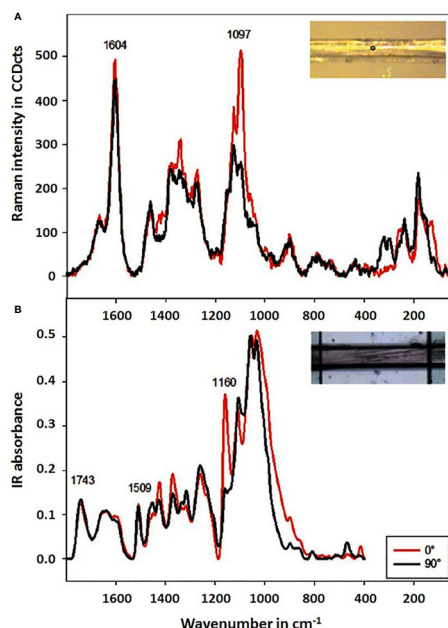

FIGURE 10 | Baseline corrected Raman **(A)** and Infrared spectra **(B)** acquired on mechanically isolated single spruce wood fibers (microfibril angle < 10°) with the incident polarization direction parallel to the fiber (0°, red spectra) and perpendicular to the fiber (90°, black spectra). Reproduced in compliance with CC BY-NC-ND 4.0 license from Gierlinger, 2018.
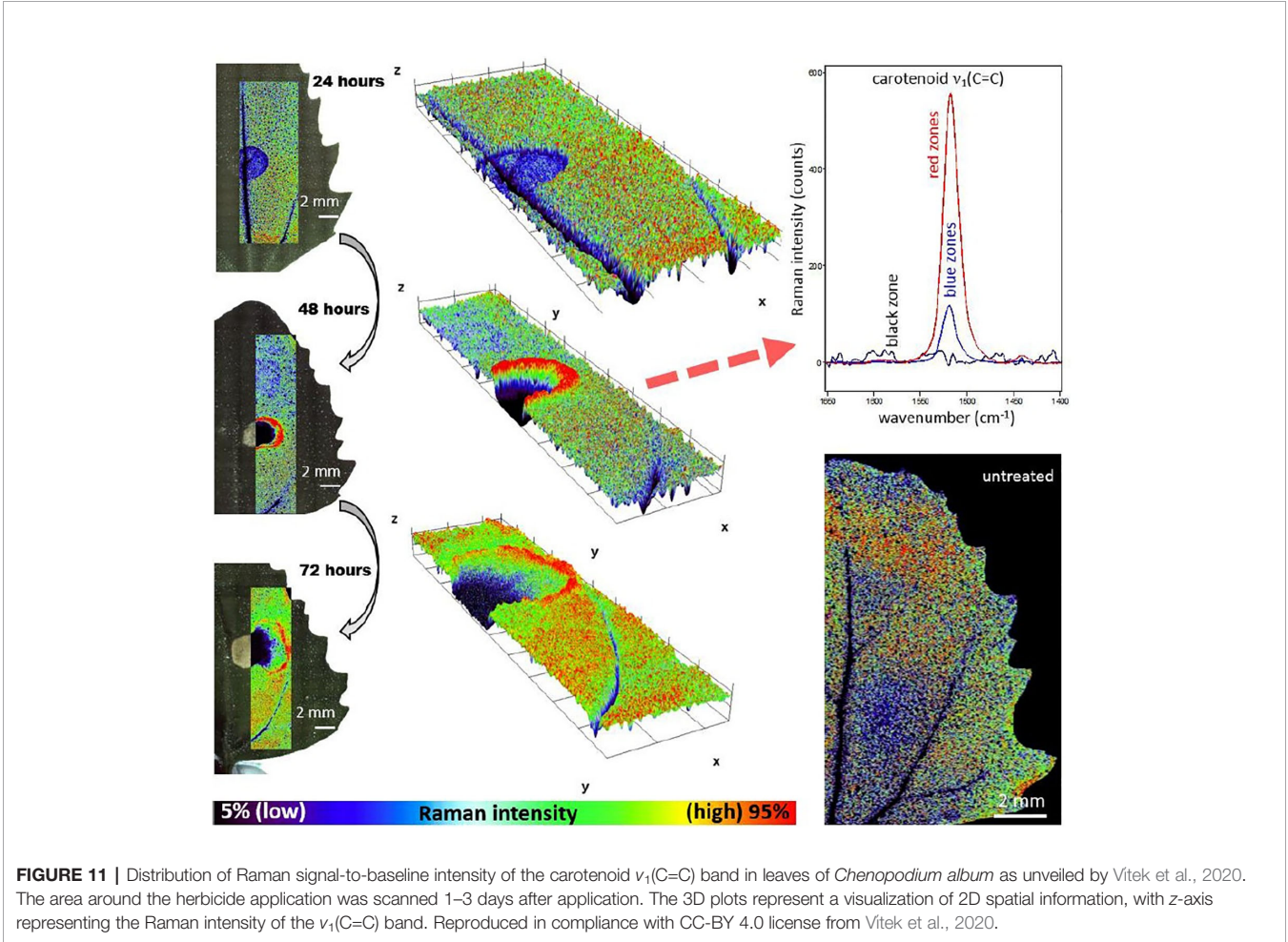
SUMMARY AND FUTURE PROSPECTS

The present state-of-the-art plant science takes advantage of spectral imaging technique across numerous research directions. IR

TABLE 4 | Comparison of the principal characteristics of near-infrared (NIR), infrared (IR), and Raman microspectroscopy relevant to imaging studies in plant science.

	NIR	IR	Raman
Excitation mechanism	Absorbance	Absorbance	Scattering
Complexity of instrumentation	Low	Medium	High
Selection rule (chemical sensitivity)	Change in dipole moment (polar moieties, enhanced signal of X-H groups, e.g., O-H, N-H, C-H)	Change in dipole moment (polar moieties)	change in polarizability (non-polar symmetrical bonds, e.g., C–C, skeletal vibrations)
Confocal resolution/ in-depth sampling	Not possible	Not possible	~0.6–4 μm
Sampling (i.e., spectra acquisition modes)	Transmission; diffuse reflection; transflection	Transmission; reflection; transflection; cassegrain or attenuated total reflectance (ATR)	Reflection (scattering); immersion microscope objectives with high numerical aperture possible
Remarks about sample preparation	No/minimal sample preparation needed; Moderate suitability of water as solvent or glass as container/optics	In transmission mode: optimal sample thickness; In ATR mode: optimal/stable sample-IRE contact surface;	Plane surface for mapping/imaging; no contact mode necessary; suitability of water as solvent or glass as container/optics
Chemical specificity	Low to moderate	High	High
Major issues and challenges	Low sensitivity; Overlapping contributions in the spectra; Difficult spectra interpretation;	Limited suitability of moist samples; Unsuitability of glass optics and materials (absorption of glass); Interfering signal from atmospheric H_2O and CO_2	Raman signal obscured by autofluorescence (stronger for excitation lasers with shorter emission wavelengths); Laser heating, danger of destruction of molecular structure, e.g., proteins; or sample thermal decomposition (particularly of dried material)

Information on the latter two techniques adapted in compliance with CC BY-NC-ND 4.0 license from Gierlinger, 2018.

FIGURE 11 | Distribution of Raman signal-to-baseline intensity of the carotenoid v_1(C=C) band in leaves of *Chenopodium album* as unveiled by Vítek et al., 2020. The area around the herbicide application was scanned 1–3 days after application. The 3D plots represent a visualization of 2D spatial information, with *z*-axis representing the Raman intensity of the v_1(C=C) band. Reproduced in compliance with CC-BY 4.0 license from Vítek et al., 2020.
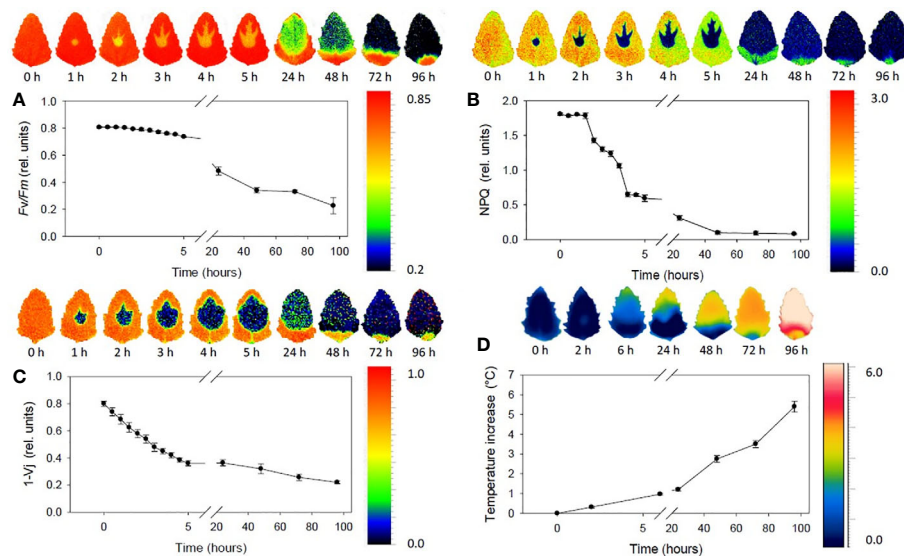
FIGURE 12 | Sequence of chlorophyll fluorescence and thermal imaging of leaves in the span 0–96 h after application as unraveled by Vítek et al., 2020. **(A)** The maximum quantum efficiency of PSII (F_V/F_m). **(B)** Non-photochemical quenching (NPQ). **(C)** Probability that a trapped exciton moves an electron into the electron transport chain beyond primary quinone electron acceptors of PSII (1-V_J). **(D)** Temperature increase in comparison to untreated control. Points represent means and error bars standard deviations (n=5). Reproduced in compliance with CC-BY 4.0 license from Vítek et al., 2020.

techniques benefit from chemical specificity and are well-suited to characterize molecular properties and spatial distribution of chemical contents, unveil microstructural features and follow biochemical processes ongoing in plants. NIR imaging does not require sample preparations, is more suitable for studies of moist samples and can be used to study water distribution in roots and soil. Because of deeper penetration of NIR light into the sample, this approach is more feasible to determine properties over larger volume in typically inhomogeneous plant material. This makes it preferable for applications where local variations of chemical compositions are not relevant or even detrimental, as it is in determinations of the total chemical contents, e.g., of pharmaceutically active ingredients in medicinal plants. For similar reasons, assessments of nutritional values, micro- and macro-nutrients concentrations and distributions, as well as other material quality parameters (e.g., relevant in agro/food applications), are easily obtained with NIR imaging. Further, deep-tissue sensing is possible in non-destructive manner. Direct interpretation of NIR spectra is difficult, although recent advances are promising. Raman imaging is a potent technique as it achieves superb spatial resolution, which makes it suitable for microstructure investigations and sub-cellular studies. Furthermore, it enables practical confocal resolution, enabling acquisition of information from beneath sample surface in a controlled manner. Care needs to be taken when examining specimens sensitive to thermal decomposition; recently a comprehensive dissection of this issue was published (Hauswald et al., 2019). This technique is fully suitable to study moist samples, although green plant material may pose a challenge because of interfering fluorescence signal. This feature may reduce the selectivity of Raman imaging, but still

selected compounds like carotenoids can be easily elucidated. The characteristics sketched above should not be considered in absolute categories, as the capabilities of the reviewed techniques overlap to some extent. Additionally, simultaneous application of several imaging techniques can mutually mitigate their limitations and further elevate the abundance of information elucidated from the sample. Combined approaches seem to be an increasingly commenced trend with promising outlook. Alongside, imaging instrumentation is continuously improving. Developments stimulated by research goals set up in other fields, such as biomedical spectral imaging, are adopted into plant investigations; for example, high-performing synchrotron IR imaging. This progress is accompanied by the development of data-analytical methods and image generation algorithms.

AUTHOR CONTRIBUTIONS

KB designed and prepared the manuscript. KB and JG performed literature search. KB and JG edited the manuscript. CH supervised the entire process. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Agarwal, U. P. (2014). 1064 nm FT-Raman spectroscopy for investigations of plant cell walls and other biomass materials. *Front. Plant Sci.* 5:490:490. doi: 10.3389/fpls.2014.00490

Albani, J. R. (2004). *Structure and dynamics of macromolecules: absorption and fluorescence studies* (Amsterdam, Netherlands: Elsevier Science).

Aneece, I., and Thenkabail, P. (2018). Accuracies achieved in classifying five leading world crop types and their growth stages using optimal Earth observing-1 hyperion hyperspectral narrowbands on Google Earth engine. *Remote Sens.* 10, 2027. doi: 10.3390/rs10122027

Arnold, T., Leitner, R., and Bodner, G. (2016). *Application of NIR hyperspectral imaging for water distribution measurements in plant roots and soil, 2016 IEEE SENSORS* (Orlando, FL: IEEE SENSORS), 1–3. doi: 10.1109/ICSENS.2016.7808534

Barańska, M., Schulz, H., Barański, R., Nothnagel, T., and Christensen, L. P. (2005). *In situ* simultaneous analysis of polyacetylenes, carotenoids and polysaccharides in carrot roots. *J. Agri. Food Chem.* 53, 6565–6571. doi: 10.1021/jf0510440

Barron, C., Parker, M. L., Mills, E. N. C., Rouau, X., and Wilson, R. H. (2005). FTIR imaging of wheat endosperm cell walls in situ reveals compositional and architectural heterogeneity related to grain hardness. *Planta* 220, 667–677. doi: 10.1007/s00425-004-1383-6

Bayçu, G., Moustaka, J., Gevrek, N., and Moustakas, M. (2018). Chlorophyll fluorescence imaging analysis for elucidating the mechanism of Photosystem II acclimation to cadmium exposure in the hyperaccumulating plant Noccaea caerulescens. *Materials* 11, 2580. doi: 10.3390/ma11122580

Beć, K. B., and Huck, C. W. (2019). Breakthrough potential in near-infrared spectroscopy: spectra simulation. A review of recent developments. *Front. Chem.* 7:48:48. doi: 10.3389/fchem.2019.00048

Beć, K. B., Grabska, J., and Huck, C. W. (2020). *Issues in Hyperspectral Traceability of Foods in Reference Module in Food Science* (Amsterdam, Netherlands: Elsevier).

Belt, T., Altgen, M., Mäkelä, M., Hänninen, T., and Rautkari, R. (2019). Cellular level chemical changes in scots pine heartwood during incipient brown rot decay. *Sci. Rep.* 9, 5188. doi: 10.1038/s41598-019-41735-8

Biswas, D. K., Coulman, B., Biligetu, B., and Fu, Y.-B. (2020). Advancing Bromegrass Breeding Through Imaging Phenotyping and Genomic Selection: A Review. *Front. Plant Sci.* 10, 1673. doi: 10.3389/fpls.2019.01673

Bodner, G., Nakhforoosh, A., Arnold, T., and Leitner, D. (2018). Hyperspectral imaging: a novel approach for plant root phenotyping. *Plant Methods* 14, 84. doi: 10.1186/s13007-018-0352-1

Bonwell, E. S., Fisher, T. L., Fritz, A., and Wetzel, D. L. (2008). Determination of endosperm protein secondary structure in hard wheat breeding lines using synchrotron infrared microspectroscopy. *Vib. Spectrosc.* 48, 76–81. doi: 10.1016/j.vibspec.2008.04.002

Bowling, D. J. F., Mengel, K., and Kirkby, E. A. (1980). In principles of plant nutrition. *J. Ecol.* 68, 324. doi: 10.1097/00010694-198001000-00011

Brackmann, C., Bengtsson, A., Alminger, M. L., Svanberg, U., and Enejder, A. (2011). Visualization of β-carotene and starch granulesin plant cells using CARS and SHG microscopy. *J. Raman Spectrosc.* 42, 586–592. doi: 10.1002/jrs.2778

Briglia, N., Montanaro, G., Petrozza, A., Summerer, S., Cellini, F., and Nuzzo, V. (2019). Drought phenotyping in Vitis vinifera using RGB and NIR imaging. *Sci. Hortic.* 256:108555. doi: 10.1016/j.scienta.2019.108555

Burger, J., and Geladi, P. (2006). Hyperspectral NIR image regression part II: Dataset preprocessing diagnostics. *J. Chemom.* 20, 106–119. doi: 10.1002/cem.986

Burger, J., and Gowen, A. A. (2011). Data handling in hyperspectral image analysis. *Chemom. Intell. Lab. Syst.* 108, 13–22. doi: 10.1016/j.chemolab.2011.04.001

Butler, H. J., Adams, S., McAinsh, M. R., and Martin, F. L. (2017). Detecting nutrient deficiency in plant systems using synchrotron Fourier-transform infrared microspectroscopy. *Vib. Spectrosc.* 90, 46–55. doi: 10.1016/j.vibspec.2017.03.004

Carpita, N. C., Defernez, M., Findlay, K., Wells, B., Shoue, D. A., Catchpole, G., et al. (2001). Cell wall architecture of the elongating maize coleoptile. *Plant Physiol.* 127, 551–565. doi: 10.1104/pp.010146

Chan, K. L. A., and Kazarian, S. G. (2003). New opportunities in micro- and macro-attenuated total reflection infrared spectroscopic imaging: spatial resolution and sampling versatility. *Appl. Spectrosc.* 57, 381–389. doi: 10.1366/00037020360625907

Chen, J., Sun, S., and Zhou, Q. (2013). Direct observation of bulk and surface chemical morphologies of Ginkgo biloba leaves by Fourier transform mid- and near-infrared microspectroscopic imaging. *Anal. Bioanal. Chem.* 405, 9385–9400. doi: 10.1007/s00216-013-7366-3

Chen, J., Sun, S., and Zhou, Q. (2015). Data-driven signal-resolving approaches of infrared spectra to explore the macroscopic and microscopic spatial distribution of organic and inorganic compounds in plant. *Anal. Bioanal. Chem.* 407, 5695–5706. doi: 10.1007/s00216-015-8746-7

Chen, J., Sun, S., and Zhou, Q. (2016). Chemical morphology of Areca nut characterized directly by Fourier transform near-infrared and mid-infrared microspectroscopic imaging in reflection modes. *Food Chem.* 212, 469–475. doi: 10.1016/j.foodchem.2016.05.168

Colarusso, P., Kidder, L. H., Levin, I. W., and Lewis, E. N. (1999). "Raman and infrared microscopy," in *Encyclopedia of spectroscopy and spectrometry*. Eds. J. C. Lindon, G. E. Tranter and J. L. Holmes (New York, USA: Academic Press New York, USA).

Coman, C., and Leopold, L. F. (2017). *Raman Mapping: Emerging Applications, Raman Spectroscopy and Applications*. Ed. K. Maaz (Rijeka, Croatia: IntechOpen). doi: 10.5772/66097

Cooper, J. B., Abdelkader, M., and Wise, K. L. (2013). Sequentially shifted excitation Raman spectroscopy: novel algorithm and instrumentation for fluorescence-free Raman spectroscopy in spectral space. *Appl. Spec.* 67, 973–984. doi: 10.1366/12-06852

Cuello, C., Marchand, P., Laurans, F., Grand-Perret, C., Laine-Prade, V., Pilate, G., et al. (2020). ATR-FTIR microspectroscopy brings a novel insight into the study of cell wall chemistry at the cellular level. *Front. Plant Sci.* 11:105:105. doi: 10.3389/fpls.2020.00105

Dagdas, Y. F., Yoshino, K., Dagdas, G., Ryder, L. S., Bielska, E., Steinberg, G., et al. (2012). Septin-mediated plant cell invasion by the rice blast fungus, Magnaporthe oryzae. *Science* 336, 1590–1595. doi: 10.1126/science.1222934

de Juan, A., Gowen, A. A., Duponche, L., and Ruckebusch, C. (2019). "Image fusion," in *Data handling in science and technology*, vol. vol. 31 . Ed. M. Cocchi (Amsterdam, Netherlands: Elsevier), 311–344.

Djokam, M., Sandasi, M., Chen, W., Viljoen, A., and Vermaak, I. (2017). Hyperspectral imaging as a rapid quality control method for herbal tea blends. *Appl. Sci.* 7, 268. doi: 10.3390/app7030268

Dokken, K., and Davis, M. (2007). Infrared imaging of sunflower and maize root anatomy. *J. Agric. Food Chem.* 55, 10517–10530. doi: 10.1021/jf072052e

Durrant, B., Trappett, M., Shipp, D., and Notingher, I. (2019). Recent developments in spontaneous Raman imaging of living biological cells. *Curr. Opin. Chem. Biol.* 51, 138–145. doi: 10.1016/j.cbpa.2019.06.004

ElMasry, G. M., and Nakauchi, S. (2016). Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality – A comprehensive review. *Biosyst. Eng.* 142, 53–82. doi: 10.1016/j.biosystemseng.2015.11.009

Elsayad, K. (2019). Optical imaging spectroscopy for plant research: more than a colorful picture. *Curr. Opin. Plant Biol.* 52, 77–85. doi: 10.1016/j.pbi.2019.08.005

Feijó, J. A., and Moreno, N. (2004). Imaging plant cells by two-photon excitation. *Protoplasma* 223, 1–32. doi: 10.1007/s00709-003-0026-2

Femenias, A., Gatius, F., Ramos, A. J., Sanchis, V., and Marín, S. (2020). Use of hyperspectral imaging as a tool for Fusarium and deoxynivalenol risk management in cereals: A review. *Food Control* 108, 106819. doi: 10.1016/j.foodcont.2019.106819

Feng, X., Zhan, Y., Wang, Q., Yang, X., Yu, C., Wang, H., et al. (2020). Hyperspectral imaging combined with machine learning as a tool to obtain highthroughput plant salt-stress phenotyping. *Plant J.* 101, 1448–1461. doi: 10.1111/tpj.14597

Gardiner, D. J., and Graves, P. R. (1989). *Practical Raman spectroscopy* (Berlin, Heidelberg: Springer-Verlag).

Gente, R., Born, N., and Castro-Camus, E. (2013). Determination of leaf water content from terahertz time-domain spectroscopic data. *J. Infrared Millimeter Terahertz Waves* 34, 316–323. doi: 10.1007/s10762-013-9972-8

Gente, R., Rehn, A., Probst, T., Stübling, E. M., Castro-Camus, E., Covarrubias, A. A., et al. (2018). Outdoor Measurements of Leaf Water Content Using THz Quasi

Time-Domain Spectroscopy. *J. Infrared Milli. Terahz Waves* 39, 943–948. doi: 10.1007/s10762-018-0520-4

Gierlinger, N., Keplinger, T., and Harrington, M. (2012). Imaging of plant cell walls by confocal Raman microscopy. *Nat. Protoc.* 7, 1694–1708. doi: 10.1038/nprot.2012.092

Gierlinger, N. (2014). Revealing changes in molecular composition of plant cell walls on the micron – level by Raman mapping and vertex component analysis (VCA). *Front. Plant Sci.* 5, 306. doi: 10.3389/fpls.2014.00306

Gierlinger, N. (2018). New insights into plant cell walls by vibrational microspectroscopy. *Appl. Spectrosc. Rev.* 53, 517–551. doi: 10.1080/05704928.2017.1363052

Gonzalez-Torres, A., Rich, A. M., Marjo, C. E., and Henderson, R. K. (2017). Evaluation of biochemical algal floc properties using Reflectance Fourier-Transform Infrared Imaging. *Algal Res.* 27, 345–355. doi: 10.1016/j.algal.2017.09.017

Gorzsas, A., Stenlund, H., Persson, P., Trygg, J., and Sundberg, B. (2011). Cell-specific chemotyping and multivariate imaging by combined FT-IR microspectroscopy and orthogonal projections to latent structures analysis reveals the chemical landscape of secondary xylem. *Plant J.* 66, 903–914. doi: 10.1111/j.1365-313X.2011.04542.x

Gou, J. Y., Park, S., and Yu, X. H. (2008). Compositional characterization and imaging of "wall-bound" acylesters of Populus trichocarpa reveal differential accumulation of acyl molecules in normal and reactive woods. *Planta* 229, 15–21. doi: 10.1007/s00425-008-0799-9

Gowen, A. A., Feng, Y., Gaston, E., and Valdramidis, V. (2015). Recent applications of hyperspectral imaging in microbiology. *Talanta* 137, 43–54. doi: 10.1016/j.talanta.2015.01.012

Grudzinski, W., Krzeminska, I., Luchowski, R., Nosalewicz, A., and Gruszecki, W. I. (2016). Strong-light-induced yellowing of green microalgae Chlorella: a study on molecular mechanisms of the acclimation response. *Algal. Research* 16, 245–254. doi: 10.1016/j.algal.2016.03.021

Gu, Y., Zhang, Y., and Zhang, J. (2008). Integration of spatial–spectral information for resolution enhancement in hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* 46, 1347–1358. doi: 10.1109/TGRS.2008.917270

Hammes, G. G. (2005). *Spectroscopy for the biological sciences* (Hoboken, New Jersey: John Wiley & Sons, Inc).

Hauswald, W., Förster, R., Popp, J., and Heintzmann, R. (2019). Thermal illumination limits in 3D Raman microscopy: A comparison of different sample illumination strategies to obtain maximum imaging speed. *PloS One* 14, e0220824. doi: 10.1371/journal.pone.0220824

Huck-Pezzei, V. A., Pallua, J. D., Pezzei, C., Bittner, L. K., Schönbichler, S. A., Abel, G., et al. (2012). Fourier transform infrared imaging analysis in discrimination studies of St. John's wort (Hypericum perforatum). *Anal. Bioanal. Chem.* 404, 1771–1778. doi: 10.1007/s00216-012-6296-9

Kazarian, S. G., and Chan, K. L. A. (2010). Micro- and macro-attenuated total reflection Fourier transform infrared spectroscopic imaging. *Appl. Spectrosc.* 64, 135A–152A. doi: 10.1366/000370210791211673

Kidder, L. H., Haka, A. S., and Lewis, E. N. (2006). "Instrumentation for FT-IR imaging," in *Handbook of vibrational spectroscopy*. Eds. P. Griffiths and J. M. Chalmers (Chichester, UK: John Wiley & Sons, Ltd).

Kokawa, M., Yokoya, N., Ashida, H., Sugiyama, J., Tsuta, M., Yoshimura, M., et al. (2015). Visualization of gluten, starch, and butter in pie pastry by fluorescence fingerprint imaging. *Food Bioprocess Technol.* 8, 409–419. doi: 10.1007/s11947-014-1410-y

Kumar, S., Lahlali, R., Liu, X., and Karunakaran, C. (2016). Infrared spectroscopy combined with imaging: A new developing analytical tool in health and plant science. *J. Appl. Spectrosc. Rev.* 51, 466–473. doi: 10.1080/05704928.2016.1157808

Largo-Gosens, A., Hernández-Altamirano, M., García-Calvo, L., Alonso-Simón, A., Álvarez, J., and Acebes, J. L. (2014). Fourier transform mid infrared spectroscopy applications for monitoring the structural plasticity of plant cell walls. *Front. Plant Sci.* 5, 303. doi: 10.3389/fpls.2014.00303

Larkin, P. J. (2018). "Instrumentation and sampling methods," in *Infrared and Raman spectroscopy. Principles and spectral interpretation, 2nd ed*. Ed. P. J. Larkin (Amsterdam, Netherlands: Elsevier Inc).

Li, D., Zhu, Z., and Sun, D.-W. (2020). Visualization of the in situ distribution of contents and hydrogen bonding states of cellular level water in apple tissues by confocal Raman microscopy. *Analyst* 145, 897–907. doi: 10.1039/C9AN01743G

Lichtman, J. W., and Conchello, J. A. (2005). Fluorescence microscopy. *Nat. Methods* 2, 910–919. doi: 10.1038/nmeth817

Lin-Vien, D., Colthup, N. B., Fateley, W. G., and Grasselli, J. G. (1991). *The handbook of infrared and Raman characteristic frequencies of organic molecules* (San Diego, California: Academic Press).

Liu, H., Zhu, H., Li, Z., and Yang, G. (2020). Quantitative analysis and hyperspectral remote sensing of the nitrogen nutrition index in winter wheat. *Int. J. Remote Sens.* 41, 858–881. doi: 10.1080/01431161.2019.1650984

Lu, G., and Fei, B. (2014). Medical hyperspectral imaging: a review. *J. Biomed. Opt.* 19:10901. doi: 10.1117/1.JBO.19.1.010901

Marcelo, M. C. A., Soares, F. L. F., Ardila, J. A., Dias, J. C., Pedó, R., Kaiser, S., et al. (2019). Fast inline tobacco classification by near-infrared hyperspectral imaging and support vector machine-discriminant analysis. *Anal. Methods* 11, 1966–1975. doi: 10.1039/C9AY00413K

Massart, D. L., Vandeginste, B. G. M., and Smeyers-Verbeke, J. (2014). *Handbook of chemometrics and qualimetrics: part A* (Amsterdam, The Netherlands: Elsevier).

McCann, M. C., Hammouri, M., Wilson, R., Belton, P., and Roberts, K. (1992). Fourier-transform infrared microspectroscopy is a new way to look at plant-cell walls. *Plant Physiol.* 100, 1940–1947. doi: 10.1104/pp.100.4.1940

Meijón, M., Satbhai, S. B., Tsuchimatsu, T., and Busch, W. (2014). Genome-wide association study using cellular traits identifies a new regulator of root development in Arabidopsis. *Nat. Genet.* 46, 77–81. doi: 10.1038/ng.2824

Monti, F., Dell'Anna, R., Sanson, A., Fasoli, M., Pezzotti, M., and Zenoni, S. (2013). A multivariate statistical analysis approach to highlight molecular processes in plant cell walls through ATR FT-IR microspectroscopy: The role of the a-expansin PhEXPA1 in Petunia hybrid. *Vib. Spectrosc.* 65, 36–43. doi: 10.1016/j.vibspec.2012.11.010

Moustakas, M., Hanč, A., Dobrikova, A., Sperdouli, I., Adamakis, I.-D. S., and Apostolova, E. (2019). Spatial heterogeneity of cadmium effects on Salvia sclarea leaves revealed by chlorophyll fluorescence imaging analysis and laser ablation inductively coupled plasma mass spectrometry. *Materials* 12, 2953. doi: 10.3390/ma12182953

Norris, K. H. (1983). "Extraction information from spectrophotometric curves. Predicting chemical composition from visible and near-infrared spectra," in *Food research and data analysis*. Eds. H. Materns and H. Russwurm Jr. (London, UK: Applied Science), 95–114.

Oleszkiewicz, T., Pacia, M. Z., Grzebelus, E., and Baranski, R. (2020). "Light microscopy and Raman imaging of carotenoids in plant cells in situ and in released carotene crystals," in *Plant and food carotenoids. Methods in molecular biology*, vol. vol 2083 . Eds. M. Rodríguez-Concepción and R. Welsch (New York, NY: Humana).

Op De Beeck, M., Troein, C., Siregar, S., Gentile, L., Abbondanza, G., Peterson, C., et al. (2020). Regulation of fungal decomposition at single-cell level. *ISME J.* 14, 896–905. doi: 10.1038/s41396-019-0583-9

Ozaki, Y., McClure, W. F., and Christy, A. A. (2006). *Near-infrared spectroscopy in food science and technology* (Hoboken: Wiley).

Ozaki, Y., Huck, C. W., and Beć,́ K. B. (2018). "Near-IR Spectroscopy and Its Applications," in *Molecular and laser spectroscopy. Advances and applications*. Ed. V. P. Gupta (San Diego, Calif: Elsevier), 11 – 38.

Paddock, S. W., and Eliceiri, K. W. (2014). "Laser scanning confocal microscopy: history, applications, and related optical sectioning techniques," in *Confocal microscopy. Methods in molecular biology (Methods and protocols)*, vol. 1075. Ed. S. Paddock (New York, NY: Humana Press), 9–47.

Pandey, P., Ge, Y., Stoerger, V., and Schnable, J. C. (2017). High throughput in vivo analysis of plant leaf chemical properties using hyperspectral imaging. *Front. Plant Sci.* 8, 1348. doi: 10.3389/fpls.2017.01348

Parmar, P., Kumari, N., and Sharma, V. (2013). Structural and functional alterations in photosynthetic apparatus of plants under cadmium stress. *Bot. Stud.* 54, 45. doi: 10.1186/1999-3110-54-45

Pećinar, I. (2019). "Raman Microscopy in Plant Science, Carotenoids Detection in Fruit Material," in *Application of Molecular Methods and Raman Microscopy/Spectroscopy in Agricultural Sciences and Food Technology*. Eds. B. Vucelić Radović, D. Lazić and M. Nikšić (London: Ubiquity Press), 177–186. doi: 10.5334/bbj.n

Pereira, J. F. Q., Pimentel, M. F., Amigo, J. M., and Honorato, R. S. (2020). Detection and identification of Cannabis sativa L. using near infrared hyperspectral imaging and machine learning methods. A feasibility study. *Spectrochim. Acta A* 237, 118385. doi: 10.1016/j.saa.2020.118385

Pérez-Bueno, M. L., Pineda, M., Cabeza, F. M., and Barón, M. (2016). Multicolor fluorescence imaging as a candidate for disease detection in plant phenotyping. *Front. Plant Sci.* 7, 1790. doi: 10.3389/fpls.2016.01790

Pesquet, E., Zhang, B., Gorzsás, A., Puhakainen, T., Serk, H., and Escamez, S. (2013). Non-cell-autonomous postmortem lignification of tracheary elements in Zinnia elegans. *Plant Cell.* 2, 1314–1328. doi: 10.1105/tpc.113.110593

Pezzei, C. K., Schönbichler, S. A., Kirchler, C. G., Schmelzer, J., Hussain, S., Huck-Pezzei, V. A., et al. (2017). Application of benchtop and portable near-infrared spectrometers for predicting the optimum harvest time of Verbena officinalis. *Talanta* 169, 70–76. doi: 10.1016/j.talanta.2017.03.067

Piqueras, S., Füchtner, S., Rocha de Oliveira, R., Gómez-Sánchez, A., Jelavić, S., Keplinger, T., et al. (2020). Understanding the formation of heartwood in larch using synchrotron infrared imaging combined with multivariate analysis and atomic force microscope infrared spectroscopy. *Front. Plant Sci.* 10, 1701. doi: 10.3389/fpls.2019.01701

Prats Mateu, B., Stefke, B., Hauser, M.-T., and Gierlinger, N. (2014). Elucidating structural and compositional changes in plant tissues and single cells by Raman spectroscopic imaging. *Spectrosc. Europe* 26, 11–14.

Prats Mateu, B., Felhofer, M., de Juan, A., and Gierlinger, N. (2018). Multivariate unmixing approaches on Raman images of plant cell walls: new insights or overinterpretation of results? *Plant Methods* 14, 52. doi: 10.1186/s13007-018-0320-9

Prats Mateu, B., Bock, P., and Gierlinger, N. (2020). "Raman Imaging of Plant Cell Walls In The Plant Cell Wall: Methods and Protocols," in *Methods in Molecular Biology, 2nd ed.*, vol. 2149. Ed. Z. A. Popper (New York: Springer Science+Business Media, LLC, part of Springer Nature), 2020. doi: 10.1007/978-1-0716-0621-6_15

Rinnan, A., van den Berg, F., and Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trend. Anal. Chem.* 28, 1201–1222. doi: 10.1016/j.trac.2009.07.007

Roman, M., Marzec, K. M., Grzebelus, E., Simon, P. W., Baranska, M., and Baranski, R. (2015). Composition and (in)homogeneity of carotenoid crystals in carrot cells revealed by high resolution Raman imaging. *Spectrochim. Acta A* 136, 1395–1400. doi: 10.1016/j.saa.2014.10.026

Rygula, A., Oleszkiewicz, T., Grzebelus, E., Pacia, M. Z., Baranska, M., and Baranski, R. (2018). Raman, AFM and SNOM high resolution imaging of carotene crystals in a model carrot cell system. *Spectrochim. Acta A* 197, 47–55. doi: 10.1016/j.saa.2018.01.054

Salzer, R., and Siesler, H. W. (2014). *Infrared and Raman spectroscopic imaging. 2nd edition* (Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA).

Sandasi, M., Vermaak, I., Chen, W., and Viljoen, A. M. (2014). Hyperspectral imaging and chemometric modeling of Echinacea — A novel approach in the quality control of herbalm. *Molecules* 19, 13104–13121. doi: 10.3390/molecules190913104

Sandasi, M., Chen, W., Vermaak, I., and Viljoen, A. M. (2018). Non-destructive quality assessment of herbal tea blends using hyperspectral imaging. *Phytochem. Lett.* 24, 94–101. doi: 10.1016/j.phytol.2018.01.016

Saulnier, L., Robert, P., Grintchenko, M., Jamme, F., Bouchet, B., and Guillon, F. (2009). Wheat endosperm cell walls: Spatial heterogeneity of polysaccharide structure and composition using micro-scale enzymatic fingerprinting and FT-IR microspectroscopy. *J. Cereal Sci.* 50, 312–317. doi: 10.1016/j.jcs.2009.05.003

Schulz, H., and Baranska, M. (2007). Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vib. Spectrosc.* 43, 13–25. doi: 10.1016/j.vibspec.2006.06.001

Schulz, H., Baranska, M., and Baranski, R. (2005). Potential of NIR-FT-Raman spectroscopy in natural carotenoid analysis. *Biopolymers* 77, 212–221. doi: 10.1002/bip.20215

Schulz, H., Krähmer, A., Naumann, A., and Gudi, G. (2014). *Infrared and Raman spectroscopic mapping and imaging of plant materials in infrared and Raman spectroscopic imaging. 2nd ed.* Eds. R. Salzer and H. W. Siesler (Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA), 227–293.

Segone, R. T., Tankeu, S. Y., Chea, W., Combrinck, S., Schmidt, M., and Viljoen, A. (2020). Rapid differentiation of Piper methysticum (kava) plant parts using single point and imaging vibrational spectroscopy. *J. Appl. Res. Med. Aromatic. Plants* 16, 100235. doi: 10.1016/j.jarmap.2019.100235

Sharma, S. K., Nelson, D. R., Abdrabu, R., Khraiwesh, B., Jijakli, K., Arnoux, M., et al. (2015). An integrative Raman microscopy-based workflow for rapid in situ analysis of microalgal lipid bodies. *Biotechnol. Biofuels* 8, 164. doi: 10.1186/s13068-015-0349-1

Singh, A. K., Pérez-López, A. V., Simpson, J., and Castro-Camus, E. (2020). Three-dimensional water mapping of succulent Agave victoriae-reginae leaves by terahertz imaging. *Sci. Rep.* 10, 1404. doi: 10.1038/s41598-020-58277-z

Sun, D., Zhu, Y., Xu, H., He, Y., and Cen, H. (2019). Time-series chlorophyll fluorescence imaging reveals dynamic photosynthetic fingerprints of sos Mutants to drought stress. *Sensors* 19:2649. doi: 10.3390/s19122649

Sun, J., Zhou, X., Hu, Y., Wu, X., Zhang, X., and Wang, P. (2019). Visualizing distribution of moisture content in tea leaves using optimization algorithms and NIR hyperspectral imaging. *Comput. Electron. Agric.* 160, 153–159. doi: 10.1016/j.compag.2019.03.004

Suutarinen, J., Anakainen, L., and Autio, K. (1998). Comparison of light microscopy and spatially resolved Fourier transform infrared (FT-IR) microscopy in the examination of cell wall components of strawberries. *Lebensmittel-Wissenschaft Technol.* 31, 595–601. doi: 10.1006/fstl.1997.0331

Truernit, E., Bauby, H., Dubreucq, B., Grandjean, O., Runions, J., Barthélémy, J., et al. (2008). High-resolution whole-mount imaging of three dimensional tissue organization and gene expression enables the study of phloem development and structure in Arabidopsis. *Plant Cell* 20, 1494–1503. doi: 10.1105/tpc.107.056069

Türker-Kaya, S., and Huck, C. W. (2017). A review of mid-infrared and near-infrared imaging: principles, concepts and applications in plant tissue analysis. *Molecules* 22:168. doi: 10.3390/molecules22010168

Vidal, M., and Amigo, J. M. (2012). Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemometr. Intell. Lab. Syst.* 11, 138–148. doi: 10.1016/j.chemolab.2012.05.009

Vítek, P., Veselá, B., and Klem, K. (2020). Spatial and temporal variability of plant leaf responses cascade after PSII inhibition: Raman, chlorophyll fluorescence and infrared thermal imaging. *Sensors* 20, 1015. doi: 10.3390/s20041015

Vongsvivut, J., Pérez-Guaita, D., Wood, B. R., Heraud, P., Khambatta, K., Hartnell, D., et al. (2019). Synchrotron macro ATR-FTIR microspectroscopy for high-resolution chemical mapping of single cells. *Analyst* 144, 3226–3238. doi: 10.1039/C8AN01543K

Wang, X.-J., Jin, G., Li, L.-Q., Liu, Y., Kianpoor Kalkhajeh, Y., Ning, J.-M., et al. (2020). NIR hyperspectral imaging coupled with chemometrics for nondestructive assessment of phosphorus and potassium contents in tea leaves. *Infrared Phys. Technol.* 108:103365. doi: 10.1016/j.infrared.2020.103365

Warner, C. A., Biedrzycki, M. L., Jacobs, S. S., Wisser, R. J., Caplan, J. L., and Sherrier, J. (2014). An optical clearing technique for plant tissues allowing deep imaging and compatible with fluorescence microscopy. *Plant Physiol.* 166, 1684–1687. doi: 10.1104/pp.114.244673

Warren, F. J., Perston, B. P., Galindez-Najera, S. P., Edwards, C. H., Powell, P. O., Mandalari, G., et al. (2015). Infrared microspectroscopic imaging of plant tissues: spectral visualization of Triticum aestivum kernel and Arabidopsis leaf microstructure. *Plant J.* 84, 634–646. doi: 10.1111/tpj.13031

Williams, P., and Norris, K. (2001). *Near-infrared technology: in the agricultural and food industries. 2nd edn* (St. Paul: American Association of Cereal Chemists).

Xin, H., Zhang, X., and Yu, P. (2013). Using synchrotron radiation-based infrared microspectroscopy to reveal microchemical structure characterization: Frost damaged wheat vs. normal wheat. *Int. J. Mol. Sci.* 14, 16706–16718. doi: 10.3390/ijms140816706

Yang, H.-Y., Inagaki, T., Ma, T., and Tsuchikawa, S. (2017). High-Resolution and Non-destructive Evaluation of the Spatial Distribution of Nitrate and Its Dynamics in Spinach (Spinacia oleracea L.) Leaves by Near-Infrared Hyperspectral Imaging. *Front. Plant Sci.* 8, 1937. doi: 10.3389/fpls.2017.01937

Yu, P., Jonker, A., and Gruber, M. (2009). Molecular basis of protein structure in proanthocyanidin and anthocyanin-enhanced Lc-transgenic alfalfa in relation to nutritive value using synchrotron-radiation FTIR microspectroscopy: A novel approach. *Spectrochim. Acta A* 73, 846–853. doi: 10.1016/j.saa.2009.04.006

Yu, K.-Q., Zhao, Y.-R., Li, X.-L., Shao, Y.-N., Liu, F., and He, Y. (2014). Hyperspectral imaging for mapping of total nitrogen spatial distribution in pepper plant. *PloS One* 9 (12), e116205. doi: 10.1371/journal.pone.0116205

Yu, P. (2006). Molecular chemical structure of barley proteins revealed by ultra-spatially resolved synchrotron light sourced FTIR microspectroscopy: Comparison of barley varieties. *Biopolymers* 85, 308–317. doi: 10.1002/bip.20661

Zeise, I., Heiner, Z., Holz, S., Joester, M., Büttner, C., and Kneipp, J. (2018). Raman imaging of plant cell walls in sections of Cucumis sativus. *Plants* 7:7. doi: 10.3390/plants7010007

Zhao, Y., Man, Y., Wen, J., Guo, Y., and Lin, J. (2019). Advances in Imaging Plant Cell Walls. *Trends Plant Sci.* 24, 867–878. doi: 10.1016/j.tplants.2019.05.009

Zhong, R., Lee, C., McCarthy, R. L., Reeves, C. K., Jones, E. G., and Ye, Z. H. (2011). Transcriptional activation of secondary wall biosynthesis by rice and maize NAC and MYB transcription factors. *Plant Cell Physiol.* 52, 1856–1871. doi: 10.1093/pcp/pcr123

Zhu, J.-K. (2002). Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53, 247–273. doi: 10.1146/annurev.arplant.53.091401.143329

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership