Multiscale vibrational spectroscopy of pollen

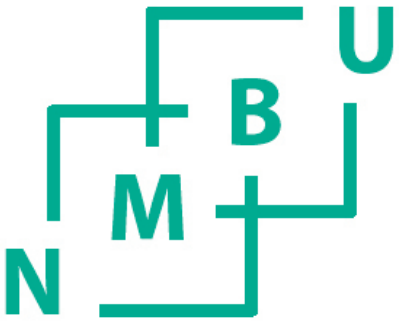
Flerskala analyser for vibrasjonsspektroskopi av pollen

Philosophiae Doctor (PhD) Thesis

Murat Bağcıoğlu

Department of Mathematical Sciences of Technology
Faculty of Environmental Sciences and Technology
Norwegian University of Life Sciences

Ås 2016



Thesis number 2016:10
ISSN 1894-6402
ISBN 978-82-575-1340-5

"If learning the truth is the scientist's goal... then he must make himself the enemy of all that he reads."

Al-Hassan Ibn al-Haytham (965-1040 AD)
a.k.a Alhazen, the "Father of Modern Optics"

Supervisors

Achim Kohler, Dr. (main supervisor)
Professor. Department of Mathematical Sciences and Technology
Norwegian University of Life Sciences (NMBU)
achim.kohler@nmbu.no

Arne Auen Grimenes, Dr. (co-supervisor)
Associate Professor. Department of Mathematical Sciences and Technology
Norwegian University of Life Sciences (NMBU)
arne.grimenes@nmbu.no

Mikael Ohlson, Dr. (co-supervisor)
Professor. Department of Ecology and Natural Resource Management
Norwegian University of Life Sciences (NMBU)
mikael.ohlson@nmbu.no

Siri Fjellheim, Dr. (co-supervisor)
Associate Professor. Department of Plant Sciences
Norwegian University of Life Sciences (NMBU)
siri.fjellheim@nmbu.no

Evaluation committee

Ganesh D. Sockalingum, Dr.
Professor. Biophotonics and Technologies for Health
Université de Reims Champagne-Ardenne
ganesh.sockalingum@univ-reims.fr

Nebojša Perišić, Dr.
Researcher, Vistin Pharma AS
nebojsa.perisic@vistin.com

Ingunn Burud, Dr.
Associate Professor. Department of Mathematical Sciences and Technology
Norwegian University of Life Sciences (NMBU)
ingunn.burud@nmbu.no

Preface

This thesis is submitted to be met the eligibility requirements of my PhD study. The research was conducted from November 2012 to May 2016 at Biospectroscopy and Data Modelling group, Department of Mathematical Sciences of Technology in Faculty of Environmental Sciences of Technology, Norwegian University of Life Sciences (NMBU, Norway). In addition, experimental works were performed at Norwegian Institute of Food, Fisheries and Aquaculture Research (NOFIMA, Ås, Norway), at French Synchrotron Radiation Facility (SMIS Beamline, SOLEIL, Paris, France) and Humboldt-Universität zu Berlin (Adlershof, Germany). The PhD study was performed under the supervision of Achim Kohler as main supervisor and Arne Auen Grimenes, Mikael Ohlson, Siri Fjellheim as co-supervisors as well as Boris Zimmermann as research collaborator. This thesis comprises introduction, methodological aspects, general discussions, conclusions and followed by attached scientific research articles produced during the thesis period.

May, 2016, Ås, Norway
Murat Bağcıoğlu

Acknowledgement

I would like to acknowledge the individuals listed below who contributed for accomplishing the thesis during my PhD trajectory. Without their time, effort and expertise, the excellence of the thesis could not be sustained.

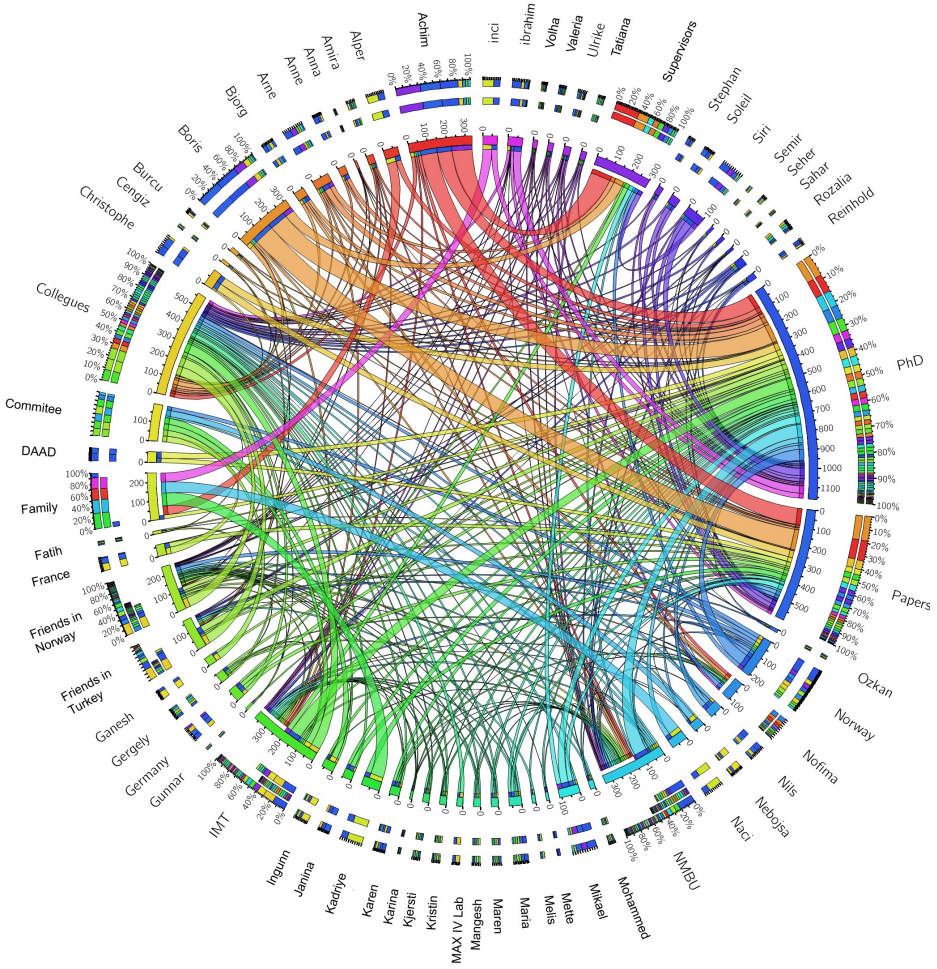


Figure i: Chord diagram for exploring relationships between a group of entities whose contribution segments are organized circularly and subdivided into individual contributions.

Abbreviations

- ATR: attenuated total reflectance
- EMSC: extended multiplicative signal correction
- CPCA: consensus principal component analysis
- HTS: high throughput screening
- MSC: multiplicative signal correction
- PCA: principal component analysis
- PLSR: partial least squares regression

Abstract

Pollen has a vital role in the reproduction of flowering plants by producing genetically diverse offspring via pollination. Environmental effects have strong influence on reproductive structures of plants, including on pollen. Phenotypic plasticity allows plants to adjust trait values to suit specific environmental conditions, and to persist in variable conditions. Plant phenotyping is the comprehensive assessment of complex plant traits related to their morphological, biochemical and physiological features. In order to determine the climate related plant adaptation and acclimation, it is imperative to improve plant phenotyping and monitoring of plant communities. Plant products are at the center of challenges posed by increasing requirements for food, feed and raw materials. Integrating new solutions across all scales, from molecular to field research, is necessary to develop sustainable plant production with higher yield by using limited resources. Therefore, plant phenotyping with a cost efficient and rapid methodology is highly desirable.

The main aim of this study was to develop a novel methodology for plant phenotyping by employing vibrational (infrared and Raman) spectroscopic techniques in combination with chemometrics. More specifically, the goal was to correlate pollen biochemical data, obtained by vibrational spectroscopy, with phylogenetic and environmental data by multivariate analyses.

In order to attain this, plant adaption and acclimation of a large sample set, including a large number of different plant species growing under a number of different environmental conditions, was studied. During the timeframe of the thesis, more than a thousand of outdoor samples have been collected from across Europe. The collected pollen sample set, consisting of Betulaceae and Fagaceae families in Fagales order, Pinaceae, Podocarpaceae and Cupressaceae families in Pinales order, as well as Poaceae and Cyperaceae in Poales order, offers phylogenetic variety and representation at the family and genus level. In addition, a group of selected grasses comprising more than five hundred individual plants were grown in controlled conditions in order to study plasticity and genetic effects. The degree and nature of variation in pollen chemical composition within and between species was studied, including phenotypic plasticity of pollen in response to differing temperatures and nutrient levels.

The main drawback of pollen analysis is that pollen grains are still mainly studied by measuring their morphology employing conventional microscopy. It is time-consuming and expensive and provides only identification based on morphological analysis, while chemical characterization of samples cannot be acquired. Vibrational (infrared and Raman) spectroscopies offer an alternative approach to pollen analysis. The vibrational spectroscopy enables biochemical characterization of pollen and detection of phylogenetic variation. Therefore, infrared and Raman spectroscopy are developing as an important tool in biology for studying phylogenetic differences between pollen grains of diverse plant species. Several diverse measurement techniques regarding single grains and bulk samples of pollen as well as microspectroscopic techniques were investigated subsequently.

Employment of the different spectroscopic techniques, such as Attenuated total reflection and microspectroscopy, which operate on different scales, can highlight different chemical aspects of pollen. Models based on chemometric methods have provided novel insight into acclimation and adaptation mechanisms of pollen. Multivariate methods allow to interrogate to what extent vibrational spectra can be used for classification and discrimination of pollen species. Namely, Consensus Principal Component Analysis (CPCA) and Partial Least Squares Regression (PLSR), were conducted in order to extract biochemical information on chemical building blocks of pollen grains. For the discrimination analysis of pollen species, chemical and physical information were separated by extended multiplicative signal correction (EMSC) and used together to establish classification models. With development of sparse PLSR classification models, the biochemical interpretation of the classification was obtained subsequently.

Overall, the value of vibrational spectroscopy as analysis method of pollen classification on taxonomic levels has been demonstrated in the peer review articles published as an outcome of the thesis. It has been shown that vibrational spectroscopy has a great potential for systematic collection of data on ecosystems in terms of environmental effects. The obtained phylogenetic variation can be well explained by the biochemical composition of pollen. Moreover, the novel findings regarding large variations in chemical composition of pollen prove that male functions and pollen quality deserve a wider research attention than they have received so far.

Norsk sammendrag

Pollen spiller en viktig rolle i reproduksjonen av blomstrende planter ved å produsere avkom via pollinering som er genetisk forskjellige. Det er kjent at miljøeffekter har en sterk innflytelse på reproduktive strukturer hos planter, inkludert pollen. Plantene er i stand til å tilpasse seg spesifikke miljøforhold og for å overleve i varierende forhold, ved å justere egenskapene. Evnen til planter til å tilpasse egenskapene til spesifikke miljøforhold kalles fenotypisk plastisitet. Komplekse planteegenskaper som morfologiske, biokjemiske og fysiologiske funksjoner vurderes i biologi gjennom plantefenotyping. For å bestemme den klimarelaterte tilpasningen og akklimatiseringen hos planter er det derfor viktig å forbedre fenotypingen og overvåkingen av ulike plantesamfunn.

Produkter fra planter er av hovedutfordringene knyttet til økt etterspørsel etter mat, fr og råvarer. Å integrere nye løsninger på tvers av flere felt, fra molekylre metoder til feltarbeid, er nødvendig for å utvikle en mer brekraftig planteproduksjon med høyere avkastning ved bruk av begrensede ressurser. Derfor er plantefenotyping med en kostnadseffektiv og hurtig metode meget ønskelig.

Hovedformålet med denne studien var å utvikle en ny metode for plantefenotyping ved å bruke vibrasjonsspektroskopiske teknikker (infrarød og Raman) i kombinasjon med kjemometri. Mer spesifikt var målet å korrelere biokjemiske data av pollen, innhentet ved vibrasjonsspektroskopi, med fylogenetiske data og miljødata ved hjelp av multivariabel analyse.

For å oppnå dette, ble store datasett med plantetilpasninger og -akklimatiseringer undersøkt, inkludert et stort antall forskjellige plante arter som vokser under en rekke forskjellige miljøforhold. Gjennom tidsrommet dekket av denne avhandlingen, ble mer enn tusen utendørs prøver samlet inn fra hele Europa. Det innsamlede prøvesettet av pollen, som består av Betulaceae og Fagaceae familier i Fagales orden, Pinaceae, Podocarpus og Cupressaceae familier i Pinales orden, samt Poaceae og Cyperaceae i Poales orden, garanterer fylogenetisk variasjon på familie- og slektsnivå. I tillegg ble en gruppe av utvalgte gress, som består av mer enn fem hundre individuelle planter, dyrket under kontrollerte betingelser for å studere plastisitet og genetiske effekter. Graden av og typen variasjon i den kjemiske sammensetningen av pollen innen og mellom arter ble studert, inkludert fenotypisk

plastisitet av pollen som svar på ulike temperaturer og nringstilgang.

Den største ulempen med tradisjonell pollenanalyse er at pollenkornene fortsatt hovedsakelig er studert ved å bestemme deres morfologi gjennom konvensjonell mikroskopi. Det er tidkrevende og kostbart, og gir bare identifikasjon basert på morfologisk analyse, mens kjemisk karakterisering av prøvene ikke blir registrert. Vibrasjonsspektroskopi (infrarød og Raman) tilbyr en alternativ tilnrming til pollenanalyse. Vibrasjonsspektroskopi muliggjør biokjemisk karakterisering av pollen og påvisning av fylogenetisk variasjon. Dette er grunnen til at infrarød- og Raman-spektroskopi i de siste årene har fåttstørre og større betydning som verktøy i biologi for å studere fylogenetiske forskjeller mellom pollenkorn av ulike plantearter. Flere ulike vibrasjonsspektroskopiske måleteknikker for å analysere enkeltkorn og sammensatte prøver av pollen, samt mikrospektroskopiske teknikker, har blitt utviklet i de siste årene.

I denne avhandlingen ble detvist at bruken av ulike spektroskopiske teknikker, for eksempel Attenuated Total Reflectance og mikrospektroskopi, som opererer på ulike skalaer, kan fremheve ulike kjemiske aspekter av pollenkornene. Modeller basert på kjemometriske metoder har gitt ny innsikt i akklimatiseringen og tilpasningsmekanismer hos pollen. Multivariate metoder kan brukes til å bestemme i hvilken grad vibrasjonsspektra kan brukes for klassifisering og diskriminering av pollenarter. Spesielt ble Consensus Principal Component Analysis (CPCA) og Partial Least Squares Regression (PLSR), utført for å trekke ut biokjemisk informasjon om de kjemiske byggesteinene i pollenkorn. For diskrimineringsanalyse av pollenarter, ble kjemisk og fysisk informasjon atskilt med Extended Multiplikative Signal Correction (EMSC) og brukt sammen for å etablere klassifiseringsmodeller. Med utviklingen av sparse PLSR klassifiseringsmodeller, ble den biokjemiske tolkningen av klassifiseringen forenklet.

Totalt sett har verdien av vibrasjonsspektroskopi som analysemetode for pollenklassifisering på taksonomiske nivåer blitt vist i peer review-artikler som er publisert som et resultat av doktorgradsarbeidet. Det er vist at vibrasjonsspektroskopi har et stort potensiale for systematisk innsamling av data om økosystemer i form av miljøeffekter. Dessuten kan de nye funnene om store variasjoner i kjemisk sammensetning av pollen bevise at mannlige funksjoner og pollenkvalitet fortjener en bredere forskningsoppmerksomhet enn de har fått så langt.

List of Papers

The following research articles are incorporated into this thesis. They are referred to by their roman numerals and are appended at the end of the thesis.

Paper I: Bağcıoğlu, Murat; Zimmermann, Boris; Kohler, Achim. A Multiscale Vibrational Spectroscopic Approach for Identification and Biochemical Characterization of Pollen. ***PLoS ONE*** 2015, 10:E0137899.

Paper II: Zimmermann, Boris; Bağcıoğlu, Murat; Sandt, Christophe; Kohler, Achim. Vibrational microspectroscopy enables chemical characterization of single pollen grains as well as comparative analysis of plant species based on pollen ultrastructure. ***Planta*** 2015, Volume 242. (5) p. 1237-1250.

Paper III: Zimmermann, Boris; Tafintseva, Valeria; Bağcıoğlu, Murat; Berdahl Høegh, Maria; Kohler, Achim. Analysis of allergenic pollen by FTIR microspectroscopy. ***Analytical Chemistry*** 2016, 88 (1), pp 803811.

Paper IV: Bağcıoğlu, Murat; Kohler, Achim, Stephan Seifert, Janina Kneipp, Zimmermann, Boris. Monitoring of plant-environment interactions by high throughput FTIR spectroscopy of pollen. ***Submitted*** 2016.

Paper V: Zimmermann, Boris; Bağcıoğlu, Murat; Tafinstseva, Valeria; Kohler, Achim; Ohlson, Mikael; Fjellheim, Siri. A greenhouse study for assessing environmental effects on pollen composition by FTIR spectroscopy. ***Submitted*** 2016.

Additional scientific contributions

Research articles

Lukacs, Rozalia; Blumel, Reinhold; Zimmermann, Boris; Bağcıoğlu, Murat; Kohler, Achim. Recovery of absorbance spectra of micrometer-sized biological and inanimate particles. ***The Analyst*** 2015; Volume 140. (9) p. 3273-3284.

Ingerslev, Anne Krog; Karaman, Ibrahim; Bağcıoğlu, Murat; Kohler, Achim; Theil, Peter Kappel; Bach Knudsen, Knud Erik; Hedemann, Mette Skou. Whole grain consumption increases gastrointestinal content of sulfate conjugated oxylipins in pigs a multicompartmental metabolomics study. ***Journal of Proteome Research*** 2015; Volume 14. (8) p. 3095-3110.

Presentations

Bağcıoğlu, Murat; Zimmermann, Boris; Sandt, Christophe; Kohler, Achim. Synchrotron Based Microspectroscopy of Pollen. ***MAX IV Lab Summer School in Synchrotron Radiation 2014-Back to Basics***; Lund, Sweden, May 19-27, 2014.

Bağcıoğlu, Murat; Zimmermann, Boris; Kohler, Achim. Vibrational Spectroscopy of Pollen. ***SPEC 2014, Shedding New Light on Disease***; Krakow, Poland, August 17-22, 2014.

Bağcıoğlu, Murat; Zimmermann, Boris; Kohler, Achim. Vibrational spectroscopy of pollen for plant phenotyping. ***Bilateral workshop on vibrational spectroscopy and data analysis***; Berlin, Germany, December 5, 2014.

Bağcıoğlu, Murat; Zimmermann, Boris; Seifert, Stephan; Kneipp, Janina; Kohler, Achim. High-throughput Fourier transform infrared spectroscopic approach for identification and biochemical characterization of pollen. ***16th European Conference on the Spectroscopy of Biological Molecules, ECSBM 2015***; Bochum, Germany, September 6-10, 2015.

Contents

If we knew what it was we were doing, it would not be called research, would it?

Albert Einstein

1 Scope

The main purpose of this thesis was to investigate vibrational spectroscopy of pollen for plant phenotyping in biology, ecology and palynology. In order to accomplish this, the following sub goals were set up.

1. To compare different vibrational spectroscopic techniques for identification and biochemical characterization of pollen.

2. To study vibrational microspectroscopy and hyperspectral imaging regarding chemical characterization of pollen grains substructures.

3. To study FTIR microspectroscopy of single pollen grain for aeroallergen monitoring.

4. To develop a high throughput FTIR approach for characterization of pollen and plant phenotyping.

5. To study the degree and nature of variation in pollen chemical composition within and between species, including plasticity of pollen in response to differing temperatures and nutrient levels.

The outline of the thesis is illustrated in Figure 1.1. Vibrational spectroscopy techniques including several sampling methods of infrared and Raman spectroscopy as well as microspectroscopy were applied to analyse pollen. Multivariate techniques were used to integrate the vibrational spectroscopy data that were obtained on different scales.

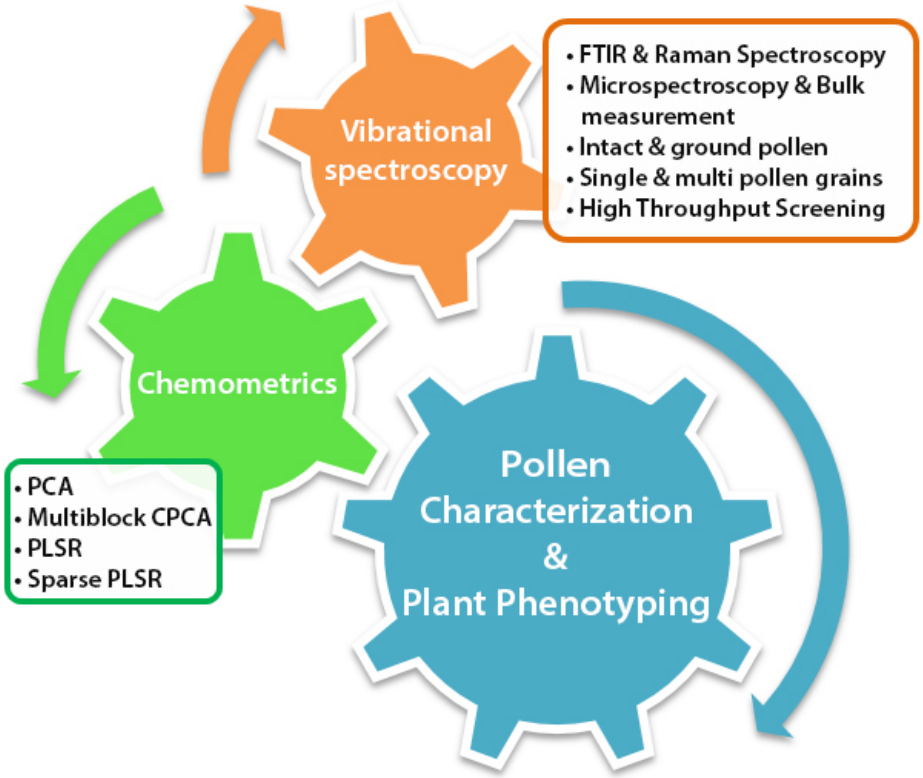


Figure 1.1: Framework of this PhD thesis

If you can't explain it simply, you don't understand it well enough.

Albert Einstein

2 Introduction

2.1 The need for efficient plant phenotyping

By the year 2050 the world population is expected to exceed 9 billion people. This will lead to a substantial increase in the demand for food and feed, and thus, yield per arable land has to be improved [1]. Progress in plant research is essential for dealing with the need for higher crop yields, and higher resistance to environmental stresses. This is even a bigger challenge due to expected impact of global climate change on ecosystems. Novel and improved methods for determination of plant traits, such as resistance to stress and the concentration of metabolites, are needed. Therefore, plant phenotyping is the key regarding sustainable food production, as well as use, conservation and protection of terrestrial ecosystems in a resource-constrained world.

2.2 Basis of current plant phenotyping

The main sources of phenotyping variations are genetic differences within and among populations, effects of the environment on traits (phenotypic plasticity), and traits change over the course of growth (ontogenetic drift) [2]. The perception of plant phenotyping in agronomy refers to quantitative and qualitative assessments of traits in terms of breeds for new cultivars and improvement of processes in plant production [3]. Plant phenotyping deals with the morphometric traits such as leaf elongation, seed size and number, plant height and width, growth, development and yield, and the physiological traits such as photosynthesis, respiration, transpiration, quantum

yield, chlorophyll fluorescence, as well as phenological traits such as flowering time, germination time, and leaf and fruit abscission. It has become a widely used tool in the characterization of plant growth and productivity in response to biotic and abiotic stress [4].

Plants have haplodiplontic life cycles: during its life cycle a plant switches from multicellular sporophyte to multicellular gametophyte life stage and back. Diploid sporophytes produce haploid spores, which develop into haploid gametophytes that produce haploid gametes. Sexual reproduction in plants occurs with the fertilization of a female gamete by a male gamete, resulting with a seed (i.e. a new sporophyte generation). Traditional plant phenotyping of seed plants (spermatophytes) has been focused on *sporophytes* which comprises the whole multicellular body except pollen (male gametophyte) and *megagametophyte* (female gametophyte). Moreover, the impact of environment on plant reproductive fitness is predominantly based on studies of female functions while the contribution of male functions is mostly ignored. For instance, measurements regarding the female components of plant fitness, such as fruit or seed production, has been performed for investigation of impact of abiotic factors, such as nutrients and light, and biotic factors, such as herbivory and pollination [5–7]. However, contribution of pollen, and male fitness in general, needs to be addressed in order to better understand impact of environment on plant reproductive fitness [8–11].

2.3 A new candidate for plant phenotyping: pollen

Pollen grains, the male gametophytes, are known as the sperm producing reproductive microorganisms of seed plants such as conifers (Pinophyta), cycads (Cycadophyta), and flowering plants (Angiospermae). By pollination, pollen guarantee genetically diverse offspring of plants, which is basic for evolution and breeding. Pollen properties are characteristic for plant families, genera or even species, and therefore they are source of information in the diverse scientific fields, such as ecology, forensics, climate change, insect migration, food (including honey) and aero allergen [12–22].

2.3.1 Chemical properties of pollen

Since the life cycle of seed plants contains both multicellular haploid stage (gametophytes) and diploid stage (sporophyte), it is often referred to as *alternation of generations*. The sporophyte creates microspores via meiosis, while mitotic division of a microspore produces a gametophyte Figure 2.1. Further mitotic division of the gametophyte produces gametes. Fusion of male and female gametes during fertilization creates zygote, which divides mitotically to produce a sporophyte, thus restarting the life cycle.

Each pollen grain is comprised of both non-reproductive (vegetative) and reproductive (generative) cells Figure 2.1. The former reveal pollen tube while the latter holds male sperm nuclei, which is required for fertilization. Generative cell divides into two sperm cells, which will be transported via pollen tube to female gametophyte during fertilization. Primarily, pollen grains are composed of two core components: cytoplasm and grain wall. Cytoplasm, which is considered as the inner part of the pollen grain, is packed with structural and nutrient chemicals such as protein, carbohydrates, and lipids. The grain wall (i.e. outer shell of the grain) consists of two layers: intine and exine Figure 2.1. Intine is mainly made up by cellulose, which is the most common cell wall material in plants. Exine is composed of sporopollenin, a complex biopolymer based on phenylpropanoid building blocks. An additional extracellular lipidic matrix, the pollen coat, can cover the interstices of the exine. Pollen coat is mostly observed in insect-pollinated plants, and has significant purposes in pollen dispersal and pollen-stigma recognition.

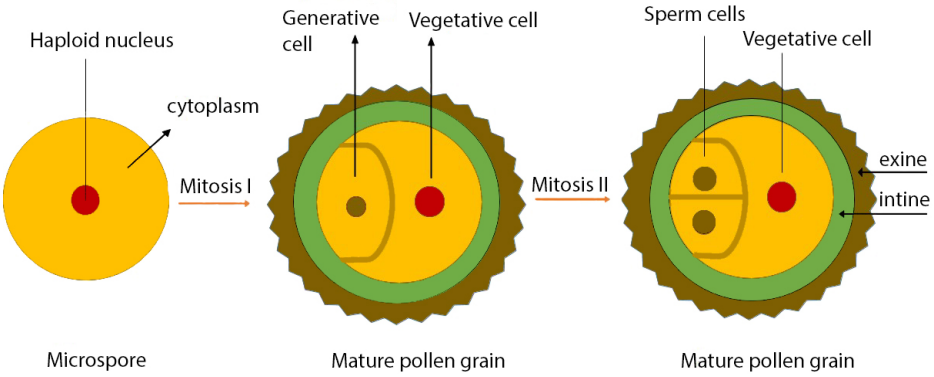


Figure 2.1: Parts of pollen grain

Pollen composition depends on sporophytic genome which controls development of exine (sporopollenin) and pollen coat, as well as gametophytic genome which controls development of intracellular materials and intine [23]. The sporopollenin and pollen coat precursors are both synthesized in the tapetum (tissue within the anther, which provides nutrition) under the control of the sporophytic genome, but at different stages of development. Pollen grains contain two major intracellular lipidic structures, namely storage oil bodies and an extensive membrane network. These intracellular lipids are synthesized in the vegetative cell of the pollen grain under the control of the gametophytic genome [24].

2.3.2 Pollination

Pollination is transfer of pollen grain (i.e. male genetic material) from anther, or male cone, to stigma, or female cone, preferably from one plant to another. Following germination of pollen grain, pollen tube will be developed from vegetative cell, and sperm cells will be transferred through pollen tube and fertilize female gamete. Pollination can be obtained by wind, water, or animals. The two most prevalent ways of pollination are wind pollination (anemophily) and insect pollination (entomophily) [25].

Plants sometimes utilise double strategy regarding pollination, such as in the case of oak (*Quercus*). Insects might collect pollen grains from anemophilous flowers when protein content of pollen from entomophilous flowers is limited, thus increasing pollination probability. Anemophilous plants are present in both gymnosperms i.e. conifers (including pine, fir and spruce) and angiosperms i.e. seed plants of which the latter do not develop scented flowers or produce nectar (including oak, beech, birch and hazel). In contrast, entomophiles and zoophilous species are almost exclusively angiosperms, while gymnosperms rarely rely on animals for pollination.

2.3.3 Pollen morphology and taxonomy

Pollen grains can have specific shape, size and texture that are unique for plant taxa. Figure 2.2 shows the variety of morphology of pollen grains. Pollen grains of anemophilous species tend to be smaller (diameter of 20 µm) and lighter than pollen from entomophiles species,

and thus it has ability to float in air easily. As can be seen in Figure 2.2A, conifer pollen of pine family (Pinaceae) has developed extra air-sacs in order to get larger surface area without increasing the weight. Anemophilous pollen grains have smoother and drier surface layers allowing to float in air easily without adhering at each other, whereas entomophilous pollen grains are ornamented, thus increasing the chance to attach to insects. Further, entomophilous pollen species often contain sticky pollen kit, which consists of lipids, carbohydrates, proteins, glycoproteins, carotenoids and flavonoids [26]

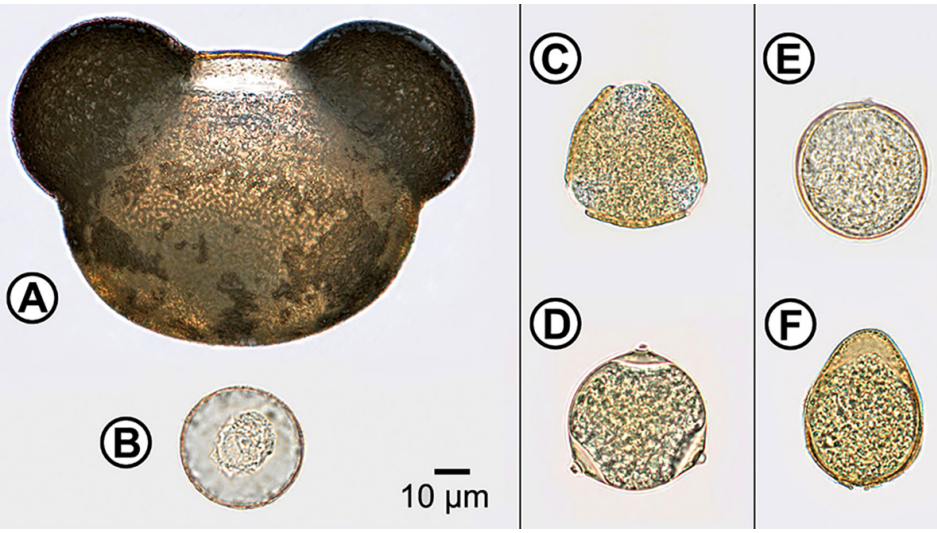


Figure 2.2: Microscope images of pollen grains. (A) *Abies cephalonica*, equatorial view; (B)*Cupressus sempervirens* (Cupressaceae), polar view; (C) *Quercus robur* (Fagaceae), polar view; (D) *Carpinus betulus* (Betulaceae), polar view; (E)*Bromus erectus* (Poaceae), equatorial view; (F) *Carex pendula* (Cyperaceae), equatorial view [PLoS ONE 10(4): e0124240]

2.4 The importance of pollen analysis

Pollen identification and characterization is important in diverse scientific fields such as in ecology for monitoring of life cycles of vegetation, in public health for allergy forecast, in botany for studying of plant-environment interactions, in paleoecology for providing insight into past plant-climate interactions, in melissopalynology for authentication of honey, and in forensics for crime scene assessment.

2.4.1 Environmental interactions studies of pollen

During the last three decades, there have been studies on the so-called pollen competition hypothesis and the relationship between pollen viability and offspring performance. Pollen viability has been considered as one of the critical factors for assessment of pollen mediated gene flow release to transgenic cultivars [27,28] . It designates the function of pollen grains in terms of compatible pollination strategies.

Environmental factors, such as herbivory, soil fertility, mycorrhizal association, temperature, relative humidity, and UV irradiation, can affect pollen viability and longevity substantially during the flowering period [29]. For instance, effect of pollen grain size and pollen production per flower as well as higher siring of seed can be observed when the environmental factors are changed [30]. Moreover, along with viability, chemical composition of pollen may be effected by change of environmental conditions during anther development [31–35]. For instance, pollen of *Petunia hybrida* (petunia) has shown that protein composition was dependent on the day temperature during the plant growth. The total amount of protein was higher for the plant grown at high temperature, whereas the amount of fats were the same [36]. Pollens grown at higher temperature compromised low molecular weight carbohydrates and phytic acid, revealing the information that pollen produced longer pollen tubes at higher temperatures and yielded more nutrients [37].

Moreover, herbivory and foliar leaf damage decreases the flower number, resulting in a decrease of pollen production per flower. Similar kinds of damage caused decreased pollen production in *Raphanus raphanistrum* (wild radish) [8] as well as in *Cucurbita texana* (Texas gourd) [38].

In addition, environmental factors have an impact on pollen viability. For instance, pollen viability can be reduced when the concentration of stratospheric ozone was diminished in geographic regions with elevated UV-B radiation because higher UV-B doses might reduce viability under direct sunlight [39].

Further, environmental factors, which diminish the pollen production (i.e. quantity), are also negatively related to the plant fitness [30, 33, 40, 41]. Pollen quality can be considered as the ability of creating new plant via pollen germination. For instance, the fertilization rates increase with the amount of pollen grains released onto stig-

mas of plants. Recently, heat stress tolerance of intact pea pollen grains was investigated via infrared spectroscopy [42, 43]. Genotypes which have been growing at higher temperature have had better performance on both pollen germination and on fertilization [43]. This property was correlated with the protein and lipid ratio of the pollen grains [42]. Thus, the pollen cell wall intine structure would change with the increasing temperature.

In the lights of aforementioned studies, knowledge about chemical composition of pollen, and thereby indirect assessment of pollen quality, is of outmost importance.

2.4.2 Pollen studies in health sciences for allergy forecast

Airborne pollen of wind pollinated plants, such as grasses and weeds, are a major cause of allergies [44, 45]. Approximately 30% of the population in developed countries are affected by pollen allergies. Pollen allergy can cause respiratory diseases in susceptible individuals, such as asthma, allergic rhinitis and hypersensitivity pneumonitis. The prevalence of pollen allergy is high and it is the most common form of respiratory disease in Europe [46]. Pollen identification is the cornerstone of aeroallergen monitoring networks and allergy forecast. Providing information on the occurrence of aeroallergens is essential for allergen avoidance and prevention, as well as for symptom treatments.

2.4.3 Pollen microfossils studies based on sporopollenin

It is known that pollen microfossils can have well preserved morphology for millions of years [47]. Therefore, in paleoecology, paleobotany, biostratigraphy and biogeography, identification of pollen microfossils is crucial for the reconstruction of past flora, population sizes and terrestrial communities. Once pollen samples are isolated from sediment layers and rocks, pollen grains can be counted and identified according to their unique shape and size using microscopy. These studies allow the reconstruction of past environments and provide an understanding of the causes of environmental changes.

Sporopollenin is an extremely robust and chemically stable substance, and is the most important component of pollen grain wall.

The ornamentation and morphology of the sporopollenin-rich exine layer is characteristic for plant taxa. Sporopollenin is one of the most resilient biological materials allowing pollen grains to survive under harsh conditions for long periods of time [48]. It is also important that concentration of phenylpropanoid components in sporopollenin can be used as UV-B proxy, thus allowing measurement of changes in the flux of UV-B radiation over geological time [49]. Therefore, the analysis of pollen substructures and their chemical components, such as grain wall biopolymers, are of foremost importance.

2.5 Conventional pollen analysis methods

Pollen research has remained basically unchanged in the last hundred years. Pollen grains are identified and characterised based on their morphological parameters by using optical microscopy. This traditional pollen analysis is based on conventional microscopic counting techniques, which are known to be labour intensive, time consuming, and limited in the reliability of identification of species [25, 50].

2.5.1 Pollen sampling

Pollen sampling depends on the analysis purpose. For instance, it can be sampled directly from the plant when assessment of plant development and plant-environment interactions is needed, particularly in greenhouse studies. In order to assess microfossil, sediment core samples are sampled by drilling or by sediment traps-gallon bottles suspended in wire cages [51]. For forensic studies, pollen samples are obtained from any kind of materials collected in crime scene such as dirt, dust and mud recovered from suspects clothing and bags [52–54]. Collection of airborne particles, including pollen and spores, from the atmosphere is obtained by air sampling devices, such as Burkard trap. It is based on a Hirst spore volumetric trap system (Burkard Scientific, UK), in which air is drawn in and the airborne particles are deposited onto adhesive tape mounted inside the trap [55]. Although sampling efficiency is high, it depends strongly on particle size, wind velocity and the types of adhesives used [56, 57].

2.5.2 Conventional microscopic counting technique: optical microscopy

The sampled pollens are usually identified on the basis of their distinct morphological features via optical microscopy. This step is strongly dependent on the expertise of the operator who is identifying and counting the pollen grains individually [58, 59]. Visual pollen identification by optical microscopy is challenging because pollen grains belonging to the same genus or the same family may have very similar morphological properties. This is demonstrated in Figure 2.3, where Poaceae (grass family) pollen grains of *Anthoxanthum*, *Deschampsia* and *Festuca* species share similar morphological features. This illustrates that even for a trained operator it is difficult to perform pollen analysis on species level by optical microscopy, and in case of grasses, even identification of genera is rarely obtained. Despite the drawbacks of conventional microscopy for pollen analysis, it is still considered as the standard method for pollen analysis [46, 59].

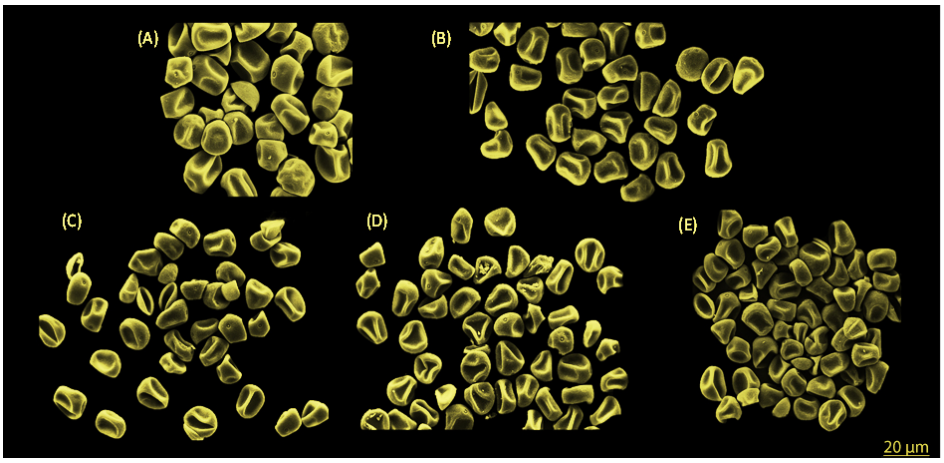


Figure 2.3: Scanning electron microscope images of pollen grains belonging the pollen species of: A) *Anthoxanthum nipponicum* , B) *Anthoxanthum odoratum*, C) *Avenella flexuosa*, D) *Deschampsia cespitosa*, E) *Festuca ovina* in top view

Different types of optical microscopic techniques have been employed in order to evaluate shape and surface texture of pollen grains in palynological studies. For instance, differential interference contrast microscopy [60] and phase contrast microscopy [61] have been performed for contrast enhancing in order to investigate fine morphological differences between pollen grains of different plant species. Flu-

orescence microscopy has been used to identify preserved specimens in samples of fossilized pollen grains [62]. Confocal laser scanning microscopy was performed to study ultrastructure of pollen grain walls [63,64]. Unfortunately, all these methods are time-consuming, expensive and provide only identification based on morphological analysis, while chemical characterization of pollen samples cannot be obtained.

2.5.3 Biomolecular methods for pollen analysis

In addition to optical microscopy, biomolecular and sequencing methods are applied for identification of pollen grains [65,66]. For instance, molecular biological methods such as quantitative real-time polymerase chain reaction (qPCR) have been employed to quantify the airborne plant materials of birch samples [66]. In qPCR instruments, the amplified DNA is labelled by fluorophores and optical components are employed for sensitive fluorescence detection. However, isolation of nucleic acids is difficult for pollen grains due to extremely resilient grain wall materials. Moreover, due to complex chemical composition of pollen grains, the extraction process may lead to traces of unwanted substances in the nucleic acid extract, which is challenging for biomolecular methods. Finally, due to costs of equipment, chemicals, and consumables, biomolecular methods have several unfavourable characteristics for standard pollen analysis [65, 67].

2.5.4 Emerging techniques for pollen analysis

During recent years, emerging spectroscopic techniques such as fluorescence spectroscopy [68], laser induced breakdown spectroscopy (LIBS) [69], matrix assisted laser desorption ionization mass spectrometry (MALDI-MS) [70, 71], Raman spectroscopy [69, 70, 72–79], and Fourier transform infrared spectroscopy (FTIR) [80–90] have been introduced for the chemical characterization of pollen grains. LIBS [69] is a versatile tool for studying relative element concentration distributions for different types of pollen. Further, it was demonstrated that MALDI-MS can be used for identification of allergenic pollen samples. MALDI-MS fingerprint spectra of whole pollen grains could be used for the classification of certain pollen species on the basis of chemical compounds such as sugars, proteins and glycoproteins [70, 71].

Vibrational spectroscopic techniques for pollen analysis

Infrared and Raman spectroscopy acquire a biochemical fingerprint of the main cell compounds such as lipids, carbohydrates and proteins. Different chemical bonds absorb infrared radiations at specific frequencies resulting in a molecular fingerprint of the irradiated sample. Both, the type of chemical bonds, as well as the molecular interactions, produce a specific biochemical and biophysical fingerprint of the sample [91]. A major asset of vibrational spectroscopies is that the samples can be measured without chemical pre-treatment [92–97]. In addition, vibrational microscopic techniques allow spatial measurement of different cell substructures [98,99]. Further, vibrational spectroscopy offers potential high-throughput approaches for measurements [98,99]. Due to the high sensitivity of biochemical variations in the spectroscopic fingerprint, and due to their high throughput characteristics, vibrational spectroscopic methods have a high potential for phenotyping [100–107]. This is the main reason why vibrational spectroscopy has attracted attention as a tool for pollen analysis and identification in recent years.

Each pollen species has characteristics biochemical composition, and thus it allows to perform the identification of pollen, including allergenic species [108]. Besides, sporopollenin has an important role to study plant adaptation to environmental changes, even information about pollen microfossils can be obtained via vibrational spectroscopy by examining sporopollenin [109–111]. Pollen microfossils are well preserved for millions of years and provide important information in the studies of paleoecology, paleobotany and biogeography. As a consequence, vibrational spectroscopy (infrared and Raman) can provide not only comprehensive chemical analysis of pollen, but can also enable simple, rapid and economic assessment of parent plants, their evolutionary history and living conditions [80, 83, 89, 90, 112].

Vibrational spectra of pollen grain comprise species-specific biochemical fingerprints of carbohydrates, lipids, proteins, biopolymers, such as cellulose and sporopollenins, and plant pigments, such as carotenoids and flavonoids [73, 81, 83, 89, 90] . These chemicals are the primary physical structures as well as nutritious constituents of pollen samples, and they are related to phenotypic features of pollen samples. Typical infrared and Raman spectra of pollen belonging to different conifer species are shown in Figure 2.4. Distinctive fea-

tures in a vibrational spectrum of pollen is a major reason why vibrational spectroscopy techniques have gained attractiveness during recent years.

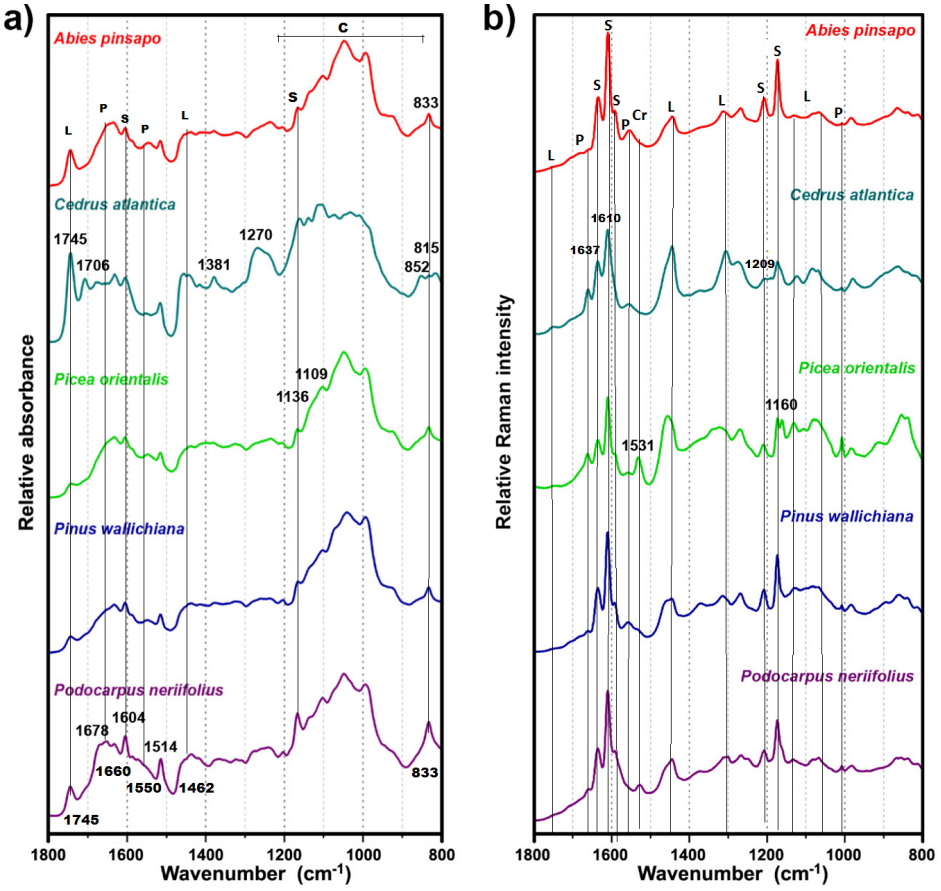


Figure 2.4: Spectra of representative samples of pollen measured as: (a) FTIR spectroscopy, (b) Raman spectroscopy. For better viewing the spectra are offset. The marked vibrational bands are linked with lipids (L), proteins (P), carbohydrates (C) sporopollenins (S) and carotenoids (Cr) (Adopted from paper I)

The spectral bands are summarized in Table 2.1.

Lipids are characterized by the strong vibrational band at 1745 cm^{-1} (C=O stretch) in FTIR, as well as by a weaker band at 1462 cm^{-1} (CH_2 deformation) in Raman. In addition to these bands, phospholipids are characterized by the bands at 1240 cm^{-1} (P=O asymmetrical stretch) and 1090 cm^{-1} (P=O symmetrical stretch). Proteins are characterized by two strong and broad bands at 1640 cm^{-1} (amide

Table 2.1: Characteristic infrared frequencies and Raman band shifts of pollen chemicals (spectral range: 1800-600 cm^{-1})

Compounds	***Frequency(cm^{-1})***	***Assignments***
Sporopollenin	1673 1605-1610 1590 1512 1209 1173 833	 vibrations of aromatic ring of phenylpropanoid building blocks
Lipids	1740-1750 1460-1444 1304 1240 1200-1000 1090 965 720	C=O str. C-H def. of CH_2 P=O str. (asym) C-O str P=O str (sym) P-O-C str. (asym) C-H def. of CH_2
Proteins	1700-1630 1550-1520 1310-1240	amide I amide II amide III
Carbohydrates	1202 1159 895 1200-1000	vibrations of pyranose units of the cellulose skeleton C-O str. C-C str. C-O-H def. C-O-C def.
Carotenoids	1160 1131	C-C str. C=C str.
Water	1640	OH_2 def.

I: C=O stretch) and 1535 cm^{-1} (amide II: NH deformation and C−N stretch) in FTIR. Carbohydrates have strong absorption bands in the 1200 - 900 cm^{-1} region (C−O−C and C−OH stretch) in FTIR, including some characteristic bands for certain types of carbohydrates, such as cellulose (at 1107, 1055 and 1028 cm^{-1}) and amylose (at 1076 and 995 cm^{-1}).

Carotenoids can be designated with the bands at 1531 cm^{-1} and 1160 cm^{-1}. Carotenoids have an allowed π-π^* electronic transition which occurs in the visible region and which gives rise to their strong colors. Even though pollen carotenoids are present at too low concentration to be observed by FTIR, they can be analyzed by Raman due to resonant effect. A resonance Raman spectrum of carotenoids is obtained when the wavelength of the incident laser coincides with electronic transition causing strong enhancement of vibrational bands, particularly those at 1531 and 1160 cm^{-1} that have strong electron-phonon coupling.

Sporopollenins, which consists of complex dehydrogenation-type biopolymers based on phenylpropanoid acids, have distinctive bands associated with the vibrations of aromatic rings at 1605, 1515, 1171, 853, 833 and 816 cm^{-1} in FTIR [80].

While in infrared spectroscopy, polar chemical bonds give rise to strong spectral bands, in Raman spectroscopy non-polar chemical bonds are predominant. In this sense, Raman spectroscopy is considered as a complementary method to infrared spectroscopy. It reveals information on the major chemical constituents of pollen [73, 89, 113]. The main spectral features, which are dominating on Raman spectra of pollen grains, are signals at 1637, 1610, 1590, 1209 and 1173 cm^{-1}, which can be associated with sporopollenin. These are bands referring to phenylpropanoid building blocks, such as p-coumaric acid found in cell walls. They are known to be species specific for pollen [114]. The lipids can be referred to the bands at 1750, 1444, 1304 and 1065 cm^{-1}. The proteins bands are related with 1660, 1455 and 1007 cm^{-1}.

Raman spectroscopy can be carried out with single pollen grains or fragments of pollen as in FTIR measurements. There are numerous studies on Raman spectroscopy for pollen analysis. In Raman spectroscopy, fluorescence and electronics absorption of the laser light can lead to heating and eventual decomposition of the sample. Most of the sample materials are colored and they absorb the light in a specific part of the visible spectrum. The absorbed light excites electrons of

the sample material to higher state of energy level and return rapidly (ps) to the ground state by emitting the same wavelength. However, for some sample materials the emission from excited state to ground state can take more time (ns) and can be accompanied by losing the part of its energy due to internal molecular process. Thus, light is emitted at longer wavelengths compared to the absorbed light which is called as fluorescence. Fluorescence is a problem in resonance Raman spectroscopy where the electrons are excited to higher electron state rather than in non-resonance Raman spectroscopy. It occurs when the virtual energy level overlaps an upper electronic level and as the energy of the laser gets higher which is shorter wavelength. Thus, it may results in low detection Raman signals due to the fact that the lifetime of electrons in excited states is longer [115]. In essence, fluorescence occurs due to real electronic transitions whereas Raman scattering occurs as a results of virtual electronic-vibrational transitions. Fluorescence can be seen when the excitation source is in the visible range. Near infrared (NIR) 785 nm laser can help to reduce the fluorescence contamination but there can be seen still some residual fluorescence effects on Raman spectra [116].

Fluorescence can be tackled with a variety of techniques such as performing confocal configuration, photobleaching or chemical bleaching, as well as by deployment of laser excitation at longer wavelengths [117–119]. Therefore, 1064 nm laser excitation in FT-Raman instruments can be used to avoid fluorescence background problems [80]. However, a longer wavelength excitation laser source results in lower spatial resolution and decrease of the sensitivity. Laucks et al [79] investigated single pollen grains by photo bleaching, i.e. photo depletion of the carotenoid molecules, in order to overcome fluorescence background problems in Raman. Photobleaching is usually performed when deployment of laser excitation in longer wavelength is not available. Nevertheless, photobleaching is not preferable since the sample is chemically altered, while the spectral acquisition time is significantly increased.

Schulte [77] performed experiments by High Performance Thin Layer Chromatography (HPTLC) combined with resonance Raman spectroscopy to investigate the pollen carotenoids. Carotenoids are extracted from pollen and separated with HPTLC before measurements by Raman spectroscopy. The spectra of resonance Raman spectroscopy of in situ carotenoids were obtained as similar to overall carotenoid

composition derived from whole pollen grains. Further, they have measured the dynamic process of pollen germination and pollen tube growth [73, 77]. Bernhard et al [120] studied chemistry and morphology of dried up pollen suspension residues, which can be extract from pollen interior. Guedes et al [74] established a Raman-based pollen spectral library for the applicability of database to detect and identify pollen in airborne samples. Recently, Wang et al [121] developed a new measurement technique, referred to as photophoretic trapping-Raman spectroscopy (PTRS), for measuring Raman spectra of bioaerosol particles, such as pollen and spores, by using photophoretic forces [122]. The combination of laser trapping with Raman spectroscopy (LTRS) allows that individual bioparticles in air or in liquid to be optically trapped and analyzed in an automated or manual manner via Raman spectroscopy [123, 124]. The photophoretic trap was combined with an aerosol delivery nozzle to allow efficient particle trapping for potential applications as an online aerosol characterization of biological molecules, fungal spores and pollen samples [125].

2.6 Vibrational spectroscopic measurement techniques for pollen analysis

Infrared spectroscopy of pollen can be applied via several diverse measurement techniques, including single- [42, 43, 80, 83, 89] and multi-reflection [126] attenuated total reflection (ATR) FTIR, diffuse reflectance FTIR (DRIFT) [90, 126], transmission FTIR measurements between windows [87, 127] and in pellets [80, 89, 90, 126], FTIR microspectroscopy measurements of single grain [83, 89, 90, 120, 128] as well as multi grains [79, 127, 128], and photoacoustic FTIR spectroscopy [82].

Raman spectroscopy of pollen can be performed as microspectroscopy measurements of single grains [77] and different parts of pollen grains [73, 75, 131] as well as bulk [80].

In this thesis, different measurement techniques of pollen analysis via vibrational spectroscopy were investigated focusing on infrared spectroscopic sampling techniques such as pellet, ATR and microspectroscopic measurements, as well as Raman microspectroscopy.

2.6.1 Pellet measurements

One of the oldest techniques for performing FTIR transmission measurements is the use of alkali halide pellets, such as potassium bromide (KBr), sodium chloride (NaCl) and potassium iodide (KI). Alkali halides are transparent in the majority of infrared region of the electromagnetic spectrum, and have plastic properties when subjected to pressure. KBr is the most commonly used alkali halide for preparing pellets. For KBr measurements, a background spectrum can be obtained by an empty pellet holder or using a blank (sample free) KBr pellet. However, blank pellets can create additional artefacts, since pellets are not identical and thus can have different moisture content and scattering effects. Therefore, it is often better not to use the blank pellets as a reference since it can be easier to remove scatter effects by spectral pre-processing [132].

KBr measurement of pollen samples has been considered as a benchmark method in this study due to its extensive use in FTIR spectroscopy of pollen [80, 89, 90, 126]. Pellet measurement is the method that obtains high quality pollen spectra, and thus any further development of methods needs to be compared with this method. It is important to notice that during the preparation of pellets, pollen sample is homogenised by grounding.

2.6.2 Attenuated Total Reflectance (ATR) measurements

Attenuated Total Reflectance (ATR) measurements are performed by guiding infrared light through a crystal with a high refractive index, such as diamond or germanium, while the infrared light undergoes total reflection at the surface of the crystal. The infrared beam generates an evanescing wave at the surface of the crystal, and will penetrate into the sample and be partially absorbed, provided that the sample is in direct contact with the crystal. The ATR measurement principle allows to measure solid or liquid samples without further preparation.

In ATR measurements, pollen grains can be measured both intact and ground. In contrast to transmission measurements, where infrared light interacts with the whole sample, the infrared light in ATR penetrates the pollen grains only up to 0.5-5 µm in depth, de-

pending on the wavelength [131]. It means that while in transmission measurements the whole pollen grain is measured, in ATR measurements predominantly the grain wall is measured if the grain is intact [42, 43, 80, 83, 89, 126]. So far, the pollen samples were measured as intact, therefore revealing the chemical spectral information from outer layer of pollen grains. However, when pollen samples are ground, it is possible to get information from the pollen grain interior.*In the thesis, the possibility of ground measurement of pollen was investigated for better differentiation of grain wall-related signals from cytoplasm-related ones.*

2.6.3 High throughput screening - Fourier transform infrared spectroscopy (HTS-FTIR)

A high throughput system (HTS), consisting a microplate extension unit coupled to a spectrometer, can be used for high-throughput screening of liquids and cell suspensions. Suspensions are transferred to 96 or 384 microplates and dried as thin films of a few micrometers. In order to achieve thin and homogeneous films, cell suspensions are often homogenized before transfer to the microplates [100]. The technique itself allow to measure high number of samples at once.

HTS has been already established in different areas, and successfully applied as automated high throughput technique for the investigation of cells and microorganisms [134–136]. Although HTS could be useful for rapid analysis of pollen, including for monitoring of terrestrial ecosystems, it has never been used in pollen analysis. Measurement with HTS should provide similar results as KBr pellet measurements since the pollen grains are in the form of ground homogenized samples. Having in mind that the KBr is standard and well-established method for getting high quality pollen spectra, HTS-FTIR should provide high quality spectra but in a high throughput way. However, due to thick resilient grain wall, pollen samples can be more difficult to homogenize compared to aforementioned samples commonly used for HTS measurements. *Therefore, in the thesis, we wanted to evaluate to what extent the HTS technique can be used for pollen analysis and to develop a new measurement protocol for pollen analysis via the HTS-FTIR approach.*

2.6.4 Infrared microspectroscopy

By using infrared microspectroscopy imaging one can obtain spatially resolved chemical information of pollen. Depending on the detector, spatially resolved measurements can be performed as point by point measurements by a single element detector or by the simultaneous measurement of a whole area by a focal plane array detector.

FTIR microspectroscopy has been applied broadly in plant science to study plant structures, especially cell walls, seeds, and leaves in the field of ecology, plant physiology, and developmental biology [137–145]. In FTIR microspectroscopy, different types of infrared radiation sources might be applied, such as traditional thermal globar sources, synchrotron radiation sources or quantum cascade lasers (QCLs) [142, 146, 147]. Globar sources consist of silicon carbide rods generating infrared radiation and providing a uniformly illuminated aperture of 10-200 µm in diameter at the sample [148].

The spatial resolution of FTIR and Raman microspectroscopies, which are defined by the diffraction limit, depends on the wavelength of the source light and the objective used. In Raman measurements, the lateral resolution is typically within 0.3-1.0 µm range, which is comparable to microscopes that are using light in the visible range. In FTIR the diffraction limit is within 2-20 µm range due to the long wavelength of mid-infrared light [149]. However, such spatial resolution is rarely obtained without an infrared light source with high brilliance, such as synchrotrons, and thus the typical lateral resolution is within 20-50 µm range.

Synchrotron radiation light sources are approximately 100-1000 fold times brighter than current benchtop thermal infrared light sources and allow smaller apertures and thus a higher resolution. They allow to illuminate smaller aperture areas with diameters of 10-20 µm with high signal to noise ratio, and even spectra down to the diffraction limit of the light can be obtained [150]. The main use of synchrotron sources has been primarily concerned with biological samples by coupling a standard FTIR microscope to a synchrotron source. So far, FTIR microspectroscopy measurements of single and multigrain pollen samples were conducted by globar sources [81, 83, 89, 90, 120, 128–130], although synchrotron sources or QCLs can be used as well [151].

When particle sizes are in the same order as the wavelength applied in spectroscopy, strong scattering effects occur. In FTIR microspectroscopy of single pollen samples, the infrared wavelengths are typically between 2.5 µm and 25 µm, while pollen grain sizes range from 5 to 200 µm with certain types of particle sizes. Broad oscillating baselines due to Mie scattering in pollen samples is one of the foremost scatter variabilities in infrared spectra. This effect is strong in transmission microspectroscopic measurements of spherical shaped single pollen grains [151,152]. Scattering effects in FTIR microspectroscopy have been considered as a major obstacle for the reliable interpretation and further use of the IR spectra in biological and biomedical science. Different approaches have been developed to deal with Mie-type scattering [153, 154]. Mie scattering consists of both broad oscillations in the spectrum and so-called dispersive artefacts. Dispersive artefacts occur due to refractive index fluctuations in absorption resonances. The refractive index fluctuates due to absorption and is linked to the pure absorbance spectrum through the Kramers-Kronig transform. Thus, Mie-distorted spectra can only be modelled if the pure absorbance spectrum is known. Different efforts have been made to estimate the pure absorbance spectrum from the Mie-scatter distorted spectrum by iterative algorithms, which iteratively estimate the pure absorbance spectrum [153]. However, the correction of Mie-type scattering remains difficult for biological samples, such as pollen, due to the complexity of the samples. Thus, it would be of great advantage to avoid Mie-type scattering experimentally. *Therefore, one of the goals of this thesis was to investigate to what extend strong scattering can be avoided experimentally in the infrared spectroscopy of single pollen grains.*

2.6.5 Raman microspectroscopy

The major conceptual difference between infrared and Raman spectroscopy is that Raman spectroscopy is a scattering technique whereas infrared spectroscopy is an absorption technique. The laser spot size is primarily defined by the laser wavelength and microscope objective being used. The minimum achievable spot size is diffraction limit: Laser spot diameter equals to $\frac{1.22\lambda}{NA}$ where λ is the wavelength of the laser, and NA is the numerical aperture of the microscope ob-

jective used. For instance, with a 632.8 nm laser, and a 0.90/100x objective, the laser spot diameter is 772 nm. The diffraction limit in infrared is much higher than in Raman, due to use of radiation source of longer (and non-coherent) wavelengths (up to 20 µm). The penetration depth of the laser used in Raman spectroscopy depends upon the wavelength of the laser as well. For instance, at 532 nm, the wavelength of the HeNe laser, light will penetrate 0.7 µm, at 633 nm it is approximately 3 µm, whereas at 785 nm it is 12 µm on the surface of silicon crystal [155, 156].

Due to the higher resolution, Raman spectroscopy can spatially resolve substructures of pollen grains. For instance, pollen grains of pine family Figure 2.2A have two large hollow projections (saccus) from the central body (corpus) of pollen grains. Therefore, measurement can be focused on either saccus or corpus regions of the grain. The saccus region, which consists of almost pure sporopollenin can be measured in order to obtain sporopollenin spectra, while measurement of corpus region can result with the information on pollen nutrients. *In the thesis, the possibility of measurement and chemical analysis of pollen substructures by Raman spectroscopy was investigated. Moreover, Raman microscopy of pollen was used to observe how the acquired spectra differ when the depth of the laser focus is altered from outer layers of pollen grain (i.e. pollen wall) to interior parts of the grain (i.e. pollen cytoplasm).*

In questions of science, the authority of a thousand is not worth the humble reasoning of a single individual.

Galileo Galilei

3 Material and Methods

3.1 Outdoor experiments: Fieldwork study

During the study period, several families of wind pollinated plant species (Table 3.1) were collected at 3 different locations (Croatia, Germany and Norway) during 3 different pollination seasons (2012, 2013 and 2014). Pollen samples from Croatia were obtained at the Botanical Garden of the Faculty of Science in the University of Zagreb by Boris Zimmermann (Ruder Bošković Institute, Zagreb, Croatia) in 2012. Pollen samples from Germany were obtained from the Botanical Garden in Berlin-Dahlem in Germany by Stefan Seifert (Humboldt-Universität zu Berlin, Adlershof, Germany) in 2013 and 2014. Pollen samples from Norway were collected in the campus area of Norwegian University of Life Sciences in Ås in 2013 and 2014.

All pollen samples were collected directly from plants at flowering time by shaking either male cones, i.e. strobili (Pinales) or male catkins (Fagales). Pollen samples of Poales were collected by cutting several flowering spikes from the same or different individuals. The samples were then kept at room temperature for 24 hours, cleaned, and subsequently stored in microcentrifuge tubes at -15 °C until measurements time. The chosen sample set, consisting of different families Betulaceae and Fagaceae in Fagales order, Pinaceae, Podocarpaceae and Cupressaceae families in Pinales order, as well as Poaceae and Cyperaceae in Poales order, offers phylogenetic variety and representation at the family and genus level (Table 3.1). Moreover, the samples show diversity in grain size, shape, and relative biochemical composition.

Table 3.1: List of pollen samples used in the thesis.

Order	Family	Genus	Species	Name
Pinales	**Pinacea**	***Abies***	*A. pinsapo*	Spanish Fir
			A. cephalonica	Greek Fir
		Cedrus	*C. atlantica*	Atlas Cedar
		Picea	*P. omorika*	Serbian Spruce
			P. orientalis	Caucasian Spruce
			P. abies	Norway Spruce
			P. pungens	Blue Spruce
		Pinus	*P. banksiana*	Jack Pine
			P. peuce	Macedonian Pine
			P. mugo	Mountain Pine
			P. nigra	European Black Pine
			P. resinosa	Red Pine
			P. sylvetris	Scots Pine
			P. tabuliformis	Chinese Red Pine
			P. wallichiana	Himalayan Pine
	Podocarpaceae	***Podocarpus***	*P. neriifolius*	Brown Pine
	Cupressaceae	***Cunnighamia***	*C. lanceolata*	China Fir
		Cupressus	*C. sempervirens*	Mediterranean Cypress
		Juniperus	*J. chinensis*	Chinese Juniper
			J. communis	Common Juniper
			J. excelsa	Greek Juniper
			J. phenicea	Phoenician Juniper
		Cryptomeria	*C. japonica*	Japanese Cedar
Fagales	**Fagacea**	***Quercus***	*Q. robur*	English Oak
			Q. rubra	Northern Red Oak
			Q. shumardii	Shumard Oak
			Q. cerris	Turkish Oak
			Q. cocciena	Scarlet Oak
			Q. libani	Lebanon Oak
			Q. palutris	Spanish Oak
			Q. petrea	Irish Oak
		Fagus	*F. sylvatica*	European Beech
	Betulaceae	***Betula***	*B. pendula*	Silver Birch
			B. alleghaniensis	Yellow Birch
			B. divaricata	Siberian Birch
			B. ermanii	Ermans Birch
			B. lenta	Black Birch
			B. papyrifera	Paper Birch
			B. pubescens	European White Birch
			B. raddeana	Raddes Birch
			B. utilis	Himalayan Birch
		Carpinus	*C. betulus*	European Hornbeam
		Alnus	*A. glutinosa*	Black Alder
			A. hirsuta	Manchurian Alder
			A. incana	Grey Alder
			A. viridis	Green Alder
		Corylus	*C. americana*	American Hazel
			C. avellana	Common Hazel
			C. chinensis	Chinese Hazel
			C. colurna	Turkish Hazel
			C. cornuta	Beaked Hazel
Poales	**Poaceae**	***Bromus***	*B. erectus*	Erect Brome
	Cyperaceae	***Carex***	*C. pendula*	Pendulous Sedge

Fagales taxa are considered to be one of the most abundant in the atmosphere. The most allergenic pollen species, responsible for a majority of allergy-related respiratory diseases, belong to Fagaceae (*Quercus*), and especially to Betulaceae (*Betula*, *Alnus*, *Corylus*, *Carpinus*). Betulaceae family has an early progressive flowering tendency in the Northern Europe. Moreover, birch (*Betula*) pollen grains produce cross allergic reaction with alder (*Alnus*) pollen as well as with hazel (*Corylus*) species, meaning that the potential cross reactivity can make diagnosing specific allergies more complicated and more severe allergic reactions. Furthermore, the concentration of grass pollen grains is highest typically 1-2 months after pollination season of Fagales order. Same as Fagales species, grasses can cause an extensive cross-reactivity for allergy sufferers. Lastly, Cupressaceae taxa, such as *Cryptomeria* and *Cupressus* can cause allergic reactions as well. *Cryptomeria* pollination season begins in January and lasts in April. Since the pollination period is quite elongated, cedar and cypress allergy are long lasting.

In addition, the reason for collecting pollen samples of Pinaceae (pine family) was typically related to their unique morphological characteristics. These are typically bisaccate pollen grains, which have two sacci and one corpus portion. (Figure 2.1). The saccus areas of pollen grains are principally made of grain wall materials, predominantly sporopollenins, whereas the corpus region of pollen grains comprises the nutrients chemicals. Due to the fact that saccate pollen has these two distinct morphological features, it makes pollen grains of Pinaceae a good candidate for vibrational spectroscopic analysis, since different pollen grain compartments can be addressed. Moreover, the obtained spectra on the different parts of the saccate pollen may differ depending on the focus depth of the Raman excitation laser.

Pollen samples were analyzed via several vibrational spectroscopic methods, followed by analyses of spectral data via different chemometric techniques.

3.2 Indoor experiments: Greenhouse study

In addition to fieldwork studies, where the pollen samples were collected as wild types in natural habitats, a greenhouse study was conducted as well. The aim was to investigate the climate effects and nutrients on the pollen composition in a controlled environment as well as phylogenetic relationships via FTIR spectroscopy. The greenhouse experiment has included species and populations originating from several diverse locations, spanning the North-South and the West-East axes. The following species of Poaceae were investigated; *Anthoxanthum odoratum* (Sweet vernal grass), *Festuca ovina* (Sheep fescue) and *Poa alpina* (Alpine meadow grass).

During the growth study, two different conditions for temperature (cold chamber at 14°C, and warm chamber at 20°C) and nutrition (nitrogen added watering and nitrogen deficient watering) were implemented. Three populations per species were used in the greenhouse study: *Poa alpina* from Sweden, Italy and Norway, *Anthoxanthum odoratum* from France, Greece and Finland, and *Festuca ovina* from Sweden, Finland and Italy. From each population, 15 genetically independent individuals were grown in separate pots, and kept outside at Norwegian University of Life Sciences, Ås, Norway over summer, before we divided each individual into four clones. During this period, plants were fertilized upon need with water containing 4% Yara Kristalon Indigo (Yara, Skøyen, Norway) and 3% YaraLiva Calcium Nitrate (Yara, Skøyen, Norway) adjusted to an electron conductivity of 1.5. In order to induce flowering, the plants were vernalized during 12 weeks at 4°C under short day length (8 h). After the vernalization period, the four clones were subjected to four different treatments: high (20°C) and low temperature (14°C) in combination with high and low levels of nutrients (+/- NU). The +NU plants were watered twice a week with nutrients containing water. The NU plants received only water, thus no extra fertilizer apart from what was in the soil at the start of the flowering period was added.

In total, 520 individuals were included in the flowering experiment, after removing misclassified individuals and some individuals with mould infection. 460 individuals produced inflorescence, and, out of those, 435 produced sufficient amount of pollen for spectroscopic analysis (at least 0.5 mg per individual).

Table 3.2: Sample set used in the greenhouse experiments

	Anthoxanthum odoratum			*Festuca ovina*			*Poa alpina*		
conditions	France	Greece	Finland	Sweden	Finland	Italy	Sweden	Italy	Norway
14 °C +NU	11	14	15	15	8	4	15	12	14
14 °C -NU	11	14	15	15	8	4	15	14	14
20 °C +NU	11	14	13	15	8	4	15	14	14
20 °C -NU	11	14	14	15	4	4	15	13	14

The data in Table 3.2 represents the number of pots (individual plants) which had produced sufficient amount of the pollen samples for measurements. For instance, all the plants of the Swedish population of *Festuca ovina* produced sufficient amount of pollen samples, whereas in case of Italian population only 20% percent of the plants created inflorescence and released pollen. As can be seen in Table 3.2, it is obvious that *Anthoxanthum odoratum* and *Poa alpina* created enough pollen samples for the analysis.

In addition to pollen collection, different morphological, phenological, and physiological parameters were measured. For morphological parameters, dry matter, number of stems with spikes and height of stems with spikes were recorded. For phenological parameters, time of pollination was recorded. For physiological parameters, chlorophyll a and b data were measured by UV spectroscopy. After pollen samples of grasses were harvested, they were analyzed applying the high-throughput FTIR spectroscopy protocol. Analysis of spectral data was performed by PCA and Sparse PLS classification methods.

3.3 FTIR spectroscopic measurements of pollen

3.3.1 Pellet measurements

The transmission infrared spectra of pollen samples were obtained by measuring potassium bromide (KBr) sample pellets, which is the most commonly used infrared transparent matrix. KBr pellets were prepared by mixing and grinding approx. 1 mg of pollen sample with approx. 100 mg of KBr using a pestle and mortar. KBr matrix was then cold-pressed without degassing into a transparent disk. Three KBr pellets were prepared and recorded for each pollen sample. Infrared transmission spectra of pollen pellets were recorded with 16 scans.

Sample free setup was used for background measurements. For the KBr method, the background measurements can be performed with an empty pellet holder or using a blank KBr pellet. However, the generated pellets (the blank one and the one with the sample) need to have uniform physical and optical properties in order to correct for scattering effects. If this requirement is not met, the resulting spectrum may contain two overlapping scatter effects, i.e. one from the sample pellet, and a different one from the blank pellet. Since obtaining perfect matched pellets of uniform thickness and crystal consistency is extremely difficult, it is usually better to use an empty pellet holder for background measurements. It may be reasoned that a blank KBr pellet could correct for moisture adsorbed on by the KBr pellet and by KBr powder during pellet preparation. However, the adsorbed water and respective spectral variations in infrared water bands was not a problem in our case since water signatures in our infrared spectra could be well separated from other chemical constituents in pollen spectra. By converting the spectral data into the second derivative form, the broad spectral bands of water vibrations were supressed [132].

3.3.2 Attenuated total reflectance (ATR) FTIR measurements

Total reflectance infrared spectra were obtained using an Equinox 55 FTIR spectrometer (Bruker Optik GmbH, Germany) with an ATR accessory. The ATR-FTIR spectra were recorded with a total of 32 scans and a spectral resolution of 4 cm^{-1}, using a 1.8 mm thick horizontal single-reflection ATR diamond prism with 45° angle of incidence on a MIRacle ATR accessory (PIKE Technologies, USA). Background (reference) spectra were recorded immediately before starting each measurement using the sample-free setup.

Reflectance FTIR spectra of pollen samples were obtained by using two different sampling methodologies: intact pollen samples as collected in nature, and pollen samples ground using an agate pestle and mortar. For ground measurements of pollen samples, 5 mg of pollen grains was grounded to a fine homogenous powder using an agate pestle and mortar. Then, approximately 1 mg of pollen samples was deposited onto the ATR crystal for FTIR measurement.

3.3.3 FTIR microspectroscopic measurements

Microscopic transmission FTIR measurements were performed using an Equinox 55 FTIR spectrometer with an IRScopeII infrared microscope (Bruker Optik GmbH, Germany). The system is equipped with a globar mid-infrared source and a liquid nitrogen-cooled mercury cadmium telluride (MCT) detector.

The pollen samples were deposited onto 3 mm thick ZnSe optical windows without any sample pretreatment. FTIR spectra of single pollen grains and multigrain were obtained in the spectral range from 4000 to 600 cm^{-1}, with a spectral resolution of 4 cm^{-1}, and with 128 scans. Samples were measured using a 15x objective, with different aperture sizes depending on the size of the sample. Single pollen grain measurements were performed by using different apertures sizes (ranging from 20 µm to 100 µm), which were adjusted to the respective pollen grain sizes. Multigrain measurements were performed with fully open aperture (200x200 µm). Background (reference) spectra were recorded immediately before starting each measurement by measuring empty areas of 3mm ZnSe slides.

3.3.4 Synchrotron FTIR microspectroscopic measurements

FTIR microspectroscopic measurements using a synchrotron radiation source were performed on intact and sectioned pollen grains at the synchrotron facility SOLEIL in Paris, France. The synchrotron spectra were measured on the SMIS infrared beamline, details of which can be found elsewhere [147]. Spectral maps were acquired as both linear maps and spectral images. For linear maps and images, step sizes of 5 or 10 100 µm were used in a linear direction or in x- and y-directions, respectively. In addition to intact pollen samples measurements, pollen samples were sectioned by embedding pollen samples into optical cutting temperature compound (OCT), which is one of the most common sectioning compound of tissues, followed by snap freezing in liquid nitrogen and sectioning in a cryostat at -20 °C. Sectioned samples were transferred onto ZnSe slides for transmission FTIR microspectroscopic measurements.

3.3.5 Scatter free FTIR microspectroscopic measurements of single pollen grains

In order to reduce scattering in FTIR microspectroscopy measurements of single pollen grains, we implemented a new approach for measuring single pollen grains. The pollen grains were embedded in a paraffin film between two polyethylene foils. The chosen polyethylene foils with paraffin are commonly used in aeroallergen sampling traps. The measurement setup is shown in Figure 3.1. A modified sample holder was used to stretch the foils to minimize the appearance of air bubbles in the sample as much as possible.

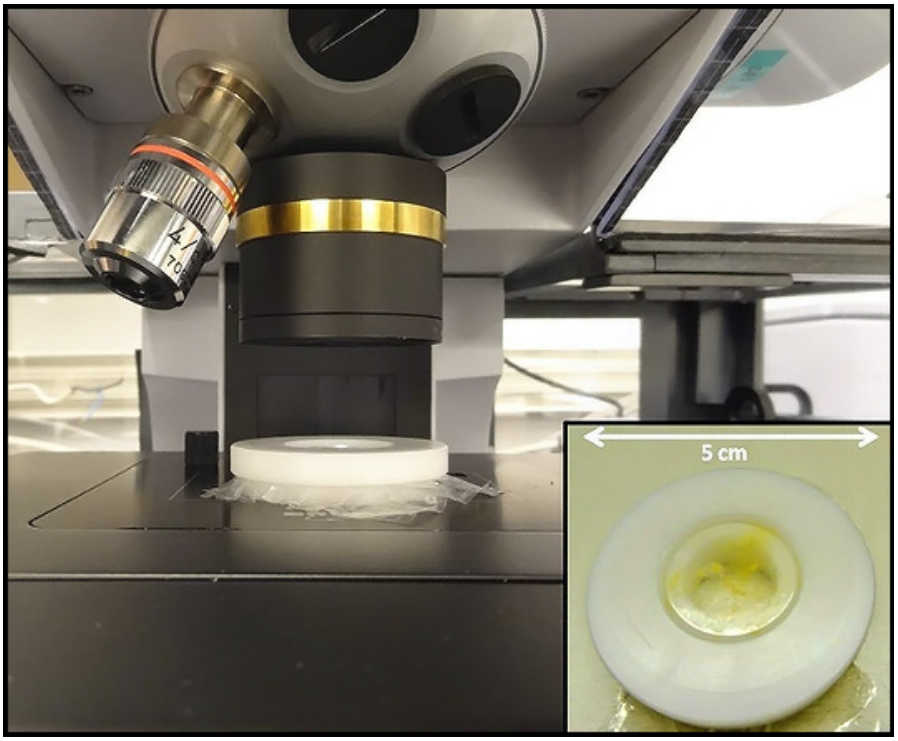


Figure 3.1: Illustration for the new experimental design to acquire scatter free single pollen grain spectra.

Mie type scattering problems are expected to be greatly reduced when the pollen grains are embedded into a surrounding medium with a comparable refractive index.

3.3.6 High throughput screening (HTS) FTIR measurements

For HTS-FTIR measurements of pollen, samples (ca. 1 mg of pollen in 500 µl water) were sonicated in ice bath by 2 mm probe coupled to ultrasonic processer under 100% power in order to get homogenous suspension of pollen. Sonication was lasted at two minutes in total (30 sec intermission after first minute to prevent overheating) due to extremely resilient grain wall structure of pollen grains. After sonication, samples were centrifuged with 13000 rpm for 10 mins and the suspension was concentrated by removing 400 µl of supernatant. From the remaining 100 µl, three aliquots each containing 8 µl were transferred on the infrared light transparent silicon 384 well microtiter plates. The microtiter plates were left for drying at room temperature for 1 hour in order to create a film that was suitable for FTIR measurements. The measurement protocol was tested by investigating the variability between different sonications and dry film preparations, different microtiter plates, different instruments as well as replicate measurements.

3.4 Raman spectroscopic measurements of pollen

Raman spectra were recorded by a LabRam HR 800 Raman microscope (Horiba Scientific, France). The excitation wavelength of 632.8 nm was generated by a HeNe laser. A 100x objective (Olympus, France) was used for focusing and collecting scattered Raman light. The confocal hole was set to 20 µm, with a laser spot diameter of approx. 0.8 µm, a spectral resolution of 2 cm^{-1}, and an exposure time of 4 x 10 s.

In order to test the spectral reproducibility and the chemical heterogeneity of single pollen grains, Raman spectra were collected from different positions on each single pollen grain. Spectra of different regions of pollen grains were investigated by manually focusing on either saccus or corpus region of the Pinales (pine order) pollen grains by Raman microspectroscopy.

3.5 Data analysis

3.5.1 Preprocessing of spectra

In order to remove unwanted variability in spectra, multivariate preprocessing methods are commonly used. For instance, baseline variations can be effectively suppressed by converting raw spectra into derivative form. Derivatives further allow separation of overlapping bands. The Savitzky-Golay algorithm is a commonly used technique to calculate derivatives of spectral data as well as supressing noise and smoothing of spectral data [157].

In addition to baseline corrections, normalization can be employed to correct so-called multiplicative effects. Multiplicative effects derive from variations in the sample thickness in infrared transmission spectroscopy. The correction of multiplicative effects may be done by dividing each variable in a spectrum by a normalization value. A normalization value may be obtained for example by one specific peak or the total area of a specific region, which is expected to be proportional to the total biomass.

Model-based pre-processing methods such as Multiplicative Signal Correction (MSC) and Extended Multiplicative Signal Correction (EMSC) are often used for correcting multiplicative effects, often in combination with spectral derivatives [158]. The advantage of MSC and EMSC is that several unwanted effects such as additive and multiplicative effects can be estimated simultaneously and removed from the spectra. The basic idea of MSC and EMSC is that in the first preprocessing step, every spectrum is modeled with respect to a reference spectrum $m(\tilde{\nu})$ according to

$$A_i(\tilde{\nu}) = a_i + b_i m(\tilde{\nu}) + E_i(\tilde{\nu}) \qquad (3.1)$$

where $m(\tilde{\nu})$ is the reference spectrum, and b_i the parameters corresponding to the reference spectrum, describing the multiplicative variation of the sample spectrum i with respect to the reference spectrum $m(\tilde{\nu})$. In transmission spectroscopy, the multiplicative effect is proportional to the effective optical path length. The parameter a_i in Equation 3.1 describes constant baseline variations, while the term $E_i(\tilde{\nu})$ accounts for unmodeled variations, which include both chemical variations and measurement noise. The MSC parameter estimation assumes that the un-modeled and informative chemical variation ends

up in the residual $E_i(\tilde{v})$ and has a spectral profile that is independent of the reference spectrum $m(\tilde{v})$ and the flat baseline. In order to apply Equation 3.1, a reference spectrum $m(\tilde{v})$, which is generally mean spectrum of spectra to be corrected, is used. In other words, the spectra $A_i(\tilde{v})$ are modified to be as similar to the reference spectrum $m(\tilde{v})$ as possible by estimating optimal constant vertical offset parameters a_i and scaling parameters b_i. For this estimation, an ordinary or weighed least squares estimation may be used. When the parameters a_i and b_i have been obtained, $A_i(\tilde{v})$ spectra are corrected as in the following equation:

$$A_{i,corr}(\tilde{v}) = \frac{(A_i(\tilde{v}) - a_i)}{b_i} \qquad (3.2)$$

If the largest differences between the spectra $A_i(\tilde{v})$ and reference spectrum $m(\tilde{v})$ are due to the baseline shifts and scaling effects, correction works well [151]. Due to the fact that overall shapes of mid infrared (MIR) spectra obtained from biological samples are very similar, model based preprocessing by simple MSC or EMSC tend to work very well in MIR spectroscopy of biological materials. The interesting differences between the spectra are very small and appear locally, which justifies the use of a reference spectrum by MSC to stabilize the model in Equation 3.1 and model all spectra that are to be corrected around this reference spectrum.

If non-constant baseline variations are present, the MSC model can be extended by a linear and a quadratic term:

$$A_i(\tilde{v}) = a_i + b_i m(\tilde{v}) + d_i(\tilde{v}) + e_i(\tilde{v})^2 + E_i(\tilde{v}) \qquad (3.3)$$

where $m(\tilde{v})$ is the reference spectrum and b_i the parameter corresponding to the reference spectrum, describing the multiplicative variation of the sample spectrum i with respect to the reference spectrum. The multiplicative effect is proportional to the effective optical path length. The parameter a_i describes the baseline variation. The terms $d_i(\tilde{v})$ and $e_i(\tilde{v})^2$ model linear and quadratic effects in the spectra, respectively. The term $E_i(\tilde{v})$ is the non-modeled residual containing chemical variation and random noise. The parameters a_i, b_i, d_i and e_i are estimated by least squares for each spectrum. Chemical variation may also, if possible, be estimated and build into the model in

Equation 3.3. However, in general chemical variation patterns are not known in advance.

3.5.2 Consensus Principal Component Analysis (CPCA)

Correlating different measurement techniques in vibrational spectroscopy of pollen

Since infrared spectra have high collinearity, methods based on latent variables may be used for analysis. The mother of all multivariate techniques based on latent variables is Principal Component Analysis (PCA). It allows the analysis of highly co-linear multivariate data. PCA has been extended for the analysis of multi-block data. Multi-block data is for example obtained, when the same sample set is measured by different vibrational spectroscopic techniques. In order to integrate different data blocks for multi-block analysis, a row-to-row correspondence between the data blocks are required [159–162]. It means that multiblock data set needs to be ordered in a way that measurements originating from the same biological replicate appears in the same row in each measurement block.

Consensus Principal Component analysis (CPCA) is a frequently used method for the analysis of multi-block data. It allows studying the co-variance patterns in multi-block data. By CPCA, global scores are calculated, which describe the consensus of all data blocks involved in the CPCA. In addition, block scores and block loadings are calculated, which show individual sample and variable variation patterns for each block. By inspecting block and global scores and block and global loadings, the clustering of samples and the molecular rationale as expressed by each vibrational spectroscopic technique can be examined.

In order to study clustering of samples, so-called block score plots and global score plots can be used. For studying variable variation patterns, correlation loading plots can be used. Correlation loading plots show correlations between variables and the global scores [159, 163]. Furthermore, information regarding experimental design variables can be added to the correlation loading plots by so-called indicator variables. This is done by defining a matrix of indicator variables with zeros and ones that identifies if a sample belongs to that

sample grouping or not [163]. Thus, the correlations between vibrational bands and the sample grouping information can be evaluated. An example of a correlation loading plot for a multi-block analysis

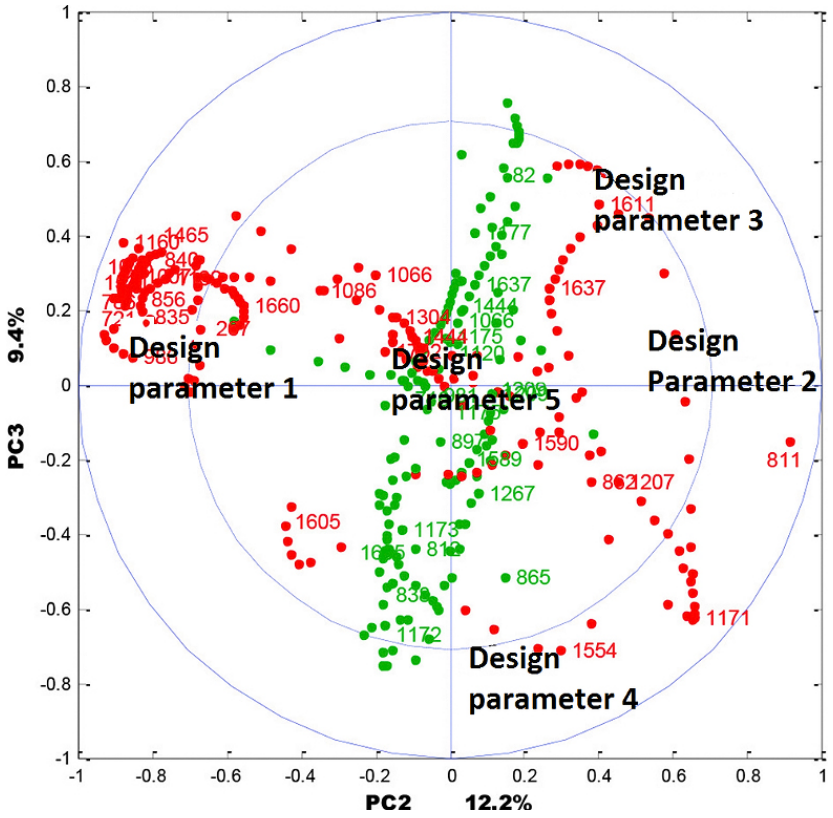


Figure 3.2: Representative correlation loading plot

is shown in Figure 3.2. The outer circle corresponds to 100% explained variance, while the inner circle corresponds to 50% explained variance. For illustration, in Figure 3.2 different design parameters and measurement variables are plotted. Differently coloured variables refer to different data blocks (for example different vibrational spectroscopic measurement techniques). Variables, i.e. absorbance values at given wavenumbers, explaining more than 50% variance of a respective combination of two components that are visualized in a correlation loading are located in the outer ring in Figure 3.2. In this window that is defined by two PC components (which account for PC2 12.2% and PC3 9.4% explained variance in the data set), we can see for example that the read variables at 9clock in Figure 3.2 are highly

positively correlated to the design parameter 1 (species/genus 1) and highly negatively correlated to design parameter 2 (species/genus 2) in PC2. Similar observations can be made for design parameter 3 and 4 and PC3. However, parameter 5 is not well explained by PC2 and PC3.

3.5.3 Sparse Partial Least Square Regression (Sparse PLSR)

Partial Least Squares regression (PLSR) is a commonly used method for regression of highly collinear data on reference data [164,165]. The reference data can be either univariate or multivariate. In this thesis, sparse PLSR was used for classification, i.e. as discriminant sparse PLSR. The reference data was defined by indicator variables. Sparse PLSR is a further development of PLSR for variable selection [166]. It allows straightforward biochemical interpretation of spectral bands due the selection of a limited subset of variables in the modeling process [166]. For each PLS component, the Sparse PLSR algorithm selects variables in loading weights according to a parameter called degree of sparsity. This parameter defines how many variables will be neglected and needs to be predefined [167].

Everything must be made as simple as possible but not simpler.

Albert Einstein

4

Main results and discussions

4.1 Paper I: Multiblock analysis of FTIR and Raman spectroscopic data of pollen for studying plant adaptation to environmental changes

In paper I, different vibrational spectroscopic methodologies were compared regarding identification and biochemical characterization of pollen. In total, 15 different Pinales (pine order) pollen species were measured by seven infrared and Raman methodologies. For infrared measurements, spectra were collected by reflectance (ATR) and transmission (KBr pellet) measurements of ground and intact pollen grains, as well as via microspectroscopy of multigrain and single pollen grain. For Raman microspectroscopy measurements, spectra were collected from the same pollen grains by focusing onto two different substructures of a single pollen grain (i.e. saccus and corpus regions).

All spectral data from the seven methodologies was integrated into one data model via Consensus Principal Component Analysis (CPCA) in order to identify molecular fingerprints traced by different techniques. Infrared and Raman spectroscopy provided precise measurement of biochemical composition of pollen samples and allowed to detect the phylogenetic variation. Spectral differences, as manifested by the vibrational spectroscopic techniques, could be connected to specific chemical constituents in pollen grains, such as lipids, carbohydrates, carotenoids and sporopollenins.

Clear differences between pollen of *Cedrus* and the rest of Pinaceae family were explicitly associated with chemical composition of sporopol-

lenins in pollen grain walls. Moreover, pollen of *Picea* had significantly higher concentration of carotenoids than the rest of the measured genera of Pinaceae. It was shown that vibrational methodologies have great potential for systematic collection of data on ecosystems and that the obtained phylogenetic variation can be well explained by the biochemical composition of pollen.

The best taxonomic differentiation of pollen species was acquired via infrared measurements on macroscopic samples, as well as by Raman microspectroscopy measurements of the corpus region of single pollen grains. In addition, Raman measurements indicate that the depth of focus is of significant importance regarding spectral (i.e. chemical) information.

4.2 Paper II: Vibrational microspectroscopy and hyperspectral imaging of pollen

Pollen grains are highly scattering samples, since the sizes of pollen samples are at the same order as the infrared wavelength range. Singe pollen grains of *Juniperus* (diameter of 15 µm) and *Cunninghamia* (diameter of 35 µm) were measured by FTIR microspectroscopy employing a conventional infrared source (globar). The single pollen spectra showed that spectra of *Juniperus* has stronger scattering effects than *Cunninghamia* spectra.

In order to obtain high-quality images of pollen grain interior, we conducted synchrotron IR measurements on the assorted pollen samples belonging to Pinaceae (pine family). The high spatial resolution of the synchrotron IR measurements revealed that chemical composition and total absorption of the spectra were changing from the saccus region to corpus region of Pinaceae pollen grains. When moving to the denser corpus region, the optical path length increased until infrared light was absorbed totally. This demonstrated, that the spectra obtained from whole and single Pinaceae pollen grains contained mainly information from the saccus region rather than from the corpus region.

Furthermore, in order to obtain high-quality infrared spectra of Pinaceae single pollen grains of all cell compartments, grains were sectioned. Thus, approximately the same optical path-length was obtained for all cell compartments. The bisaccate pollen grains were

embedded into optimum cutting temperature (OCT) media, sectioned on a microtome-cryostat to 10 µm thick sections, transferred to BaF2 slides, and measured with a 10x10 µm microscope aperture. The FTIR spectra of the corpus region were dominated by lipids and proteins, while the FTIR spectra of the saccus region were dominated by cellulose and sporopollenin. This demonstrated a considerable chemical heterogeneity of the pollen interior. Raman spectroscopy of the pollen sections showed greatly reduced fluorescence backgrounds than those observed in Raman spectra of single pollen grains. Raman spectra of the sections showed complementary molecular information regarding the substructures of pollen grains.

The acquired FTIR and Raman microspectroscopic images demonstrate clearly the methods potential for determining spatially resolved biochemical fingerprints of pollen sub-structures. As a consequence, vibrational microspectroscopy enables chemical characterization of single pollen grains as well as comparative analysis of plant species based on pollen ultrastructure. High quality spectra can be obtained even when the size of the grains is as small as 10 µm by using synchrotron-based FTIR micro spectroscopy, although strong scatter effects remain a challenge.

4.3 Paper III: Analysis of allergenic pollen by FTIR microspectroscopy

Pollen monitoring is of importance for aerobiological studies, including aeroallergenic monitoring. Today, optical microscopic measurement is considered as the standard analysis method. For optical microscopic measurements, pollen grains are collected by pollen traps and examined under optical microscope by counting the individual grains. However, this method is expensive, time consuming and dependent on subjective expertise.

FTIR microspectroscopy is an alternative method for the classification based on single pollen grain spectra. However, the obtained FTIR spectra show strong scatter effects, which hamper pollen classification by multivariate analysis of spectral fingerprints. In this study, we developed a FTIR microspectroscopy-based methodology for analysis of pollen, which reduced scatter effects considerably. The methodology was tested on 11 pollen species. Pollen grains were embedded

in a paraffin film between polyethylene foils in order to reduce scattering. The experimental procedure was adapted to the conventional method used for pollen trapping, which uses polymer foil covered with a paraffin or silicone film.

In our study, 13 different plant samples were embedded in a paraffin matrix between two thin sheets of polyethylene foil and measured by FTIR microspectroscopy. Polyethylene foil and paraffin were chosen as pollen trapping media due to their favourable properties in infrared part of the spectrum. Thirty pollen grains per plant sample were measured. The embedding matrix comprised strong absorption signals in the spectral region from 1500 to 1350 cm^{-1}, but fortunately this region does not contain the most important spectral signals of pollen. Since pollen were imbedded in paraffin, which has a refractive index that is comparable to the refractive index of pollen, the obtained spectra were quasi scatter free and similar to benchmark spectra obtained by bulk measurements of KBr pellets.

In order to evaluate the classification and biochemical composition of pollen, PCA and Sparse PLSR were performed. For data analysis, second derivative and non-derivative spectra were used. Non derivative data was pre-processed by EMSC applying a weighting vector. This allows to avoid the influence of strong chemical absorbance variations on the pre-processing parameters.

PCA plots demonstrated the potential of the technique for pollen identification on species level. The hierarchical classification scheme was built following the phylogenetic tree, thus allowing classification at order, family, genus, species, and subspecies level. The hierarchical classification comprised nine classification sub-models, and each sub-model was established for 2 or 3 classes. Discriminant analysis in each sub-model was based on sparse PLSR. The model has shown excellent classification at all hierarchical levels, resulting in overall accuracy values of 95.4% and 98.0%, for the independent test set and full cross-validation respectively.

Discrimination between grass (*Bromus erectus*) and sedge (*Carex pendula*) was based on lipid-related and protein-related absorptions. Discrimination between birch (Betulaceae) and beech (Fagaceae) families was based on lipid and carbohydrate regions. The classification of Fagaceae pollen between oak (*Quercus*) and beech (*Fagus*) genera was based only on C=O stretch associated absorption at 1743 cm^{-1} thus demonstrating effectiveness of the sparse PLSR.

The presented study demonstrated that reproducible and scatter-free FTIR spectra of single pollen grains can be obtained by employing measurements in embedding matrix. The FTIR method is outperforming the previously published studies on development of an automated pollen analysis. Moreover, this approach is easily adaptable to routine aeroallergen analyses.

4.4 Paper IV: High throughput Fourier transform infrared spectroscopic method for biochemical characterization of pollen

High throughput screening (HTS) FTIR provides rapid characterization and routine analysis of microorganisms and fluids [134–136]. In this study, a new protocol for high-throughput screening of pollen was developed. The commercially available HTS system for FTIR spectroscopy [168] allows automated and robust high throughput transmission measurements of thin films on infrared light transparent silicon 384 well microtiter plates. Samples are deposited on microtiter plates as a thin and homogenous film in order to achieve good quality FTIR spectra. In order to obtain homogenous thin films of pollen grains, pollen grains were sonicated in water, transferred onto 384 microplates and subsequently dried.

The pollen species measured in this study were collected in Northern hemisphere from 3 different pollination seasons (2012, 2013 and 2014) and 3 different locations (Germany, Croatia and Norway). The sample set, consisting of 6 different genera of Fagales (beech order), compromises phylogenetic variety and representation at the family and genus level, thus showing a rich diversity in pollen grain size, shape and chemical composition.

The spectral data shows variability within Fagales species based on either location or pollination season. For instance, pollen samples of *Quercus robur*, which were collected during the two consecutive seasons at the same location in Norway, had big variations in triglyceride nutrients. Higher triglyceride content (compared to sporopollenin) was observed in the pollen grains from 2014 pollination season. Regarding the location, *Quercus robur* pollen samples from Germany

had 4 lower amounts of triglycerides, compared to Norwegian samples. In addition, German samples show slight differences in the sporopollenin composition, as indicated by changes in signals at 1605, 1515 and 833 cm^{-1}.

The relatively large intra-seasonal variations for *Alnus glutinosa*, *Betula pendula* and *Quercus robur* have been observed. Therefore, differences in pollen composition can be substantial, even for the same species living at the same location. This indicates that pollen chemical composition can be influenced by local climate conditions.

The sampling protocol was evaluated by studying variability between sonication pretreatment and dry film preparations, which are the critical steps in high throughput measurements of pollen. In addition, variability between two FTIR HTS systems was estimated. The variability within the different phylogenetic levels were approximately 1 to 2 orders of magnitude higher than on the technical replicates level, and even variability within species can be 20-30 times higher than the technical replicates.

It was shown that the high throughput FTIR spectroscopic protocol for pollen analysis was reproducible and robust. High quality FTIR spectra can be obtained after sonication of pollen grains revealing a fragmentation of pollen suspensions. Further, variations within measurements related to sample preparation, microplate holders and instrumentation were low. The new high-throughput screening method is thus an effective tool for studying plant environmental interactions.

4.5 Paper V: A greenhouse study for assessing enviromental effects on pollen composition by FTIR spectroscopy

A group of selected grasses (Poales), grown under the controlled environments in greenhouse, were investigated in order to study effects of phenotypic plasticity and genetics on chemical composition of pollen. The sample set included three different species of grasses (*Anthoxanthum odoratum*, *Festuca ovina*,*Poa alpina*), each represented with three different populations belonging to different country of origin. In total, 15 genetically independent individuals were grown per population, and each individual was divided into four clones.

The four clones were subjected to four different treatments: high (20°C) and low temperature (14°C) in combination with high and low levels of nutrients (+/- NU). Of the 520 individuals that were included in the flowering experiment, 460 individuals developed inflorescence, and out of those 435 created sizeable amount of pollen (0.5 mg) required for the FTIR measurement. In total, 761 samples of pollen were collected.

380 pollen samples were measured by high throughput FTIR approach: 140 samples of *Anthoxanthum odoratum* (44, 48 and 48 for populations France, Greece and Finland, respectively), 96 samples of *Festuca ovina* (48, 32 and 16 for populations Sweden, Finland and Italy, respectively) and 144 samples of *Poa alpina* (48 samples for each of the populations: Sweden, Italy and Norway). Thus, samples were uniformly distributed in order to cover each of the four growth conditions with a quarter of samples (95 samples in total per growth condition). The species had different flowering phenologies: *Poa alpina* individuals were pollinating within 3-4 days, while *Anthoxanthum odoratum* individuals were pollinating within several weeks and with almost continuous production of new inflorescence.

Classification and biochemical similarities between pollen samples were evaluated by using principal component analysis (PCA), partial least-squares regression (PLSR), and variability tests based on Pearson correlation coefficients and Fisher discriminant analysis. PLSR was used for building discrete and hierarchical classification schemes.

All three species could be easily separated based on distinctive biochemicals features. For instance, *Poa alpina* has higher protein-to carbohydrate ratio than the other two species. *Anthoxanthum odoratum* has higher concentration of amylose signals compared to other two species as well as higher lipid to carbohydrate ratio compared to *Festuca ovina*. Further, all three species can be distinguishable by composition of lipid bands.

Moreover, spectra of pollen samples show chemical variability of pollen related to different environmental conditions. For instance, *Poa alpina* grown in higher temperature had pollen with higher levels of proteins whereas pollen of the one grown in colder temperature showed higher levels of lipids and carbohydrates. Moreover, *Anthoxanthum odoratum* and *Poa alpina* grown in a nutrient-rich environment exhibited higher protein to carbohydrate ratio. Further,

amylose related FTIR signals decreased with temperature in *Anthoxanthum odoratum*. This may be interpreted as a manifestation of acclimation to warmer conditions. The decreased concentration of amylose could indicate that pollen starch is partly hydrolyzed into low molecular weight carbohydrates, such as sucrose, glucose and fructose.

In addition, FTIR spectra of pollen samples belonging to different populations differ significantly. For instance,*Poa alpina* could be distinguished from other populations by protein (amide I, amide II) bands. In addition to protein bands, Norwegian populations of *Poa alpina* could be differentiated from the other populations by higher ratio of carbohydrate (amylose bands) to lipids. Italian population of *Festuca ovina* was the only population that has not shown high correlation between pollen composition and growth conditions. This could be due to the lower spectral quality and the smaller number of samples that was encountered for this population.

Furthermore, in order to correlate growth conditions based on genotypes, cluster analysis was performed employing the Fisher clustering coefficient (FCC). It turned out that the Finish population of *Anthoxanthum odoratum* had the highest FCC value, strongly implying high phenotypic rigidity. *Poa alpina* populations had the lowest FCC value thus indicating the highest plasticity of pollen chemical composition. Regarding pollen plasticity, temperature conditions during plants growth had stronger effect on pollen chemical composition than nutrient conditions.

The study showed that pollen composition is both species-specific and related to life-history strategies. Moreover, the pollination season is species-specific and it reflects biological niches. Finally, environmental conditions influence pollen composition and reflect adaptation to local climate.

Research is creating new knowledge.

Neil Armstrong

5

General conclusions and future prospects

The importance of plant phenotyping for the world economy in the fields of agronomy and forestry may anticipate a huge demand for the interest in vibrational spectroscopy of pollen. Identification and characterization of pollen allows to obtain information in several fields, such as in public health for allergy forecasts, in ecology for monitoring of life cycles of vegetation, and in forensics for temporal and spatial determination of a criminal events.

However, current studies are still limited to visual investigation of pollen grains under optical microscope, and pollen research in general has been stagnant over the last hundred years. The identification based on optical microscopes requires laborious visual recognition of bioparticles by a skilled microscopist with specific knowledge of pollen and spore morphology.

Recent studies in infrared and Raman spectroscopy have demonstrated that vibrational spectra can serve as a fingerprint of pollens overall biochemical composition and thereby their quality and physiological status. The main purpose of this thesis was to attain correlation of pollen between biochemical data obtained by vibrational spectroscopy and phylogenetic and environmental data. As one of the sub goals, pollen analysis has been investigated by vibrational spectroscopy in a multiscale approach. In paper I, this was achieved by comparison of different infrared and Raman spectroscopic measurement techniques for identification and biochemical characterization of pollen. All spectral data from the different methodologies were integrated into one data model by Consensus Principal Component Analysis (CPCA). Out of all the methodologies performed, best tax-

onomy based discrimination was obtained by pellet and reflectance (ATR) FTIR, as well as Raman microspectroscopic measurements which are focused on the specific parts of the pollen grains. Thus, Raman measurements revealed that not only measurement area, but also depth of focus have significance on the spectral data. However, there might be still some room for further improvement. For instance, depth profiling measurements of the pollen grains from the outer surface could be performed in order to extract the information from the intine and exine layers of pollen grains more precisely. Another improvement might be regarding the data set. For instance, additional data blocks with different measurement techniques, such as SERS for Raman spectroscopy or synchrotron FTIR measurements of pollen substructures, could be added in the multiblock data set. Moreover, FTIR microspectroscopic measurements were underperformed compared to the other FTIR methodologies in that study because of the variations in spatial orientations of measured pollen grains and the scattering effects. Therefore, another sub-goal of the thesis was to study vibrational microspectroscopy and hyperspectral imaging regarding chemical characterization of pollen grain sub structures.

In paper II, an approach for chemical imaging of pollen grains by FTIR and Raman microspectroscopy was presented. Clear chemical images of pollen grains were achieved by cryosectioning of pollen grains. It was revealed that spatial chemical information can be obtained by using distinctive bands of proteins, lipids and sporopollenins. The chemical images indicate the chemical heterogeneity of pollen which allows to determine spatially resolved biochemical fingerprint of pollen ultrastructure.

In paper III, the aim was to investigate to what extend strong scattering effect can be avoided experimentally in the infrared spectroscopy of single pollen grains for aeroallergen monitoring. The presented study in Paper III showed that identification of single pollen grains can be based on transmittance FTIR microspectroscopy by using an embedding matrix to obtain scatter free spectra. The classification analysis shows promising results of the method. However, there might be still some further improvements required for the future measurements. For instance, the optimization of the FTIR instrumental setting needs to be focused on decreasing spectral acquisition time. In that study, instrumental settings were under control with a large safe margin to avoid any possible information shortfall regarding

spectral quality. However, this resulted in a relatively long spectral acquisition time of 1 min per grain. Further, embedding media can be optimized. For instance, it might be designated to allow measurements of both the infrared and the visible spectral region. Such type of system would acquire an image of pollen grains with infrared spectrum, thus obtaining both morphological and chemical information of pollen grains at the same time. Although the results have showed that identification accuracy is taxon-specific, it should be noted that the sample set contained relatively limited number of species. For example, only two Poales species were studied. Therefore, more measurements with a larger pollen set might be required to better evaluate performance of the methodology. For instance, the study has shown potential for differentiating pollen subspecies. Subspecies pollen identification can be useful for assessing the characterization the different allergenic variants present in the corresponding pollen. For instance, Japanese cedar (*Cryptomeria japonica*) has two varieties of trees that have morphologically similar pollen, but with different concentration of allergen. Thus, identification of subspecies would allow more accurate detection of allergens.

In addition, one of the other sub-goals of the thesis was to develop a high throughput FTIR approach for characterization of pollen and plant phenotyping. This was achieved and presented in Paper IV and Paper V. New measurement protocol has been established to obtain high quality pollen FTIR spectra. Sonication of pollen samples achieves sufficient fragmentation of pollen grains, thus allowing their high-throughput measurement on the microtiter plates. The results of the variability within the measurement protocol steps were considerably low. The high throughput approach allows measurement of large sample sets for studies on plant-climate interactions. This has been evaluated in the study presented in Paper IV via the pollen sample set comprising 31 species, collected during three different pollinations seasons and in three different locations. Moreover, the high throughput approach has been demonstrated in Paper V where a large set of pollen grass samples (380 samples), belonging to plants grown in controlled environments, was measured. Further, the role of pollen variability in terms of plant adaptation and success of male reproduction were investigated by high throughput FTIR approach. The study showed that pollen composition and pollination seasons are species-specific and related to life-history strategies as well as

biological functions. Further, pollen composition is effected by environmental changes due to adaptation to local climate.

In future, wider range of pollen samples could be collected in order to create broader pollen database library. The database with an online access would allow investigation of plant samples for the research in biology, agronomy, forestry, ecology and medicine. Moreover, more studies are necessary in order to estimate plasticity of pollen chemical composition, whether it is adaptive and how it effects fitness under different conditions.

Finally, the developed methods are considered as simple, rapid and inexpensive, therefore it is possible to test them straightforwardly which would be adapted and implemented to botanical and environmental studies. For instance, the Hirst-type volumetric pollen trap, which is the most common type of air sampler [169] can be implemented directly to our scatter free FTIR microscopic measurement technique in the future for monitoring of aeroallergens. Moreover, the presented high throughput screening technique is directly adjustable to automated and routine analysis for studying plant communities and plant environmental interactions.

Bibliography

[1] Freibauer, A., Sustainable food consumption and production in a resource-constrained world in The 3rd SCAR Foresight Exercise European Commission Standing Committee on Agricultural Research (SCAR) 2011: Brussels.

[2] Valladares, F., E. Gianoli, and J.M. Gmez, Ecological limits to plant phenotypic plasticity. New Phytologist, 2007. 176(4): p. 749-763.

[3] Fiorani, F. and U. Schurr, Future Scenarios for Plant Phenotyping. Annual Review of Plant Biology, 2013. 64(1): p. 267-291.

[4] Masuka, B., et al., Phenotyping for abiotic stress tolerance in maize. J Integr Plant Biol, 2012. 54(4): p. 238-49.

[5] Stanton, M.L., A.A. Snow, and S.N. Handel, Floral Evolution - Attractiveness to Pollinators Increases Male Fitness. Science, 1986. 232(4758): p. 1625-1627.

[6] Krupnick, G.A. and A.E. Weis, The effect of floral herbivory on male and female reproductive success in Isomeris arborea. Ecology, 1999. 80(1): p. 135-149.

[7] Lehtila, K. and S.Y. Strauss, Effects of foliar herbivory on male and female reproductive traits of wild radish, Raphanus raphanistrum. Ecology, 1999. 80(1): p. 116-124.

[8] Strauss, S.Y., J.K. Conner, and S.L. Rush, Foliar herbivory affects floral characters and plant attractiveness to pollinators: Implications for male and female plant fitness. American Naturalist, 1996. 147(6): p. 1098-1107.

[9] Reden, J., et al., The effect of a herbal combination of primrose, gentian root, vervain, elder flowers, and sorrel on olfactory function in patients with a sinonasal olfactory dysfunction. Rhinology, 2011. 49(3): p. 342-346.

[10] Bell, G., On the Function of Flowers. Proceedings of the Royal Society Series B-Biological Sciences, 1985. 224(1235): p. 223-265.

[11] Charnov, E.L., Simultaneous Hermaphroditism and Sexual Selection. Proceedings of the National Academy of Sciences of the United States of America, 1979. 76(5): p. 2480-2484.

[12] Xiang, C.L., et al., Pollen morphology of the East Asiatic genus Chelonopsis (Lamioideae: Lamiaceae) and allied genera, with reference to taxonomic implications and potential pollination ecology. Plant Biosystems, 2013. 147(3): p. 620-628.

[13] Lazarevic, M., et al., Pollen and seed morphology of resurrection plants from the genus Ramonda (Gesneriaceae): relationship with ploidy level and relevance to their ecology and identification. Turkish Journal of Botany, 2013. 37(5): p. 872-885.

[14] Fulkerson, J.R., J.B. Whittall, and M.L. Carlson, Reproductive Ecology and Severe Pollen Limitation in the Polychromic Tundra Plant, Parrya nudicaulis (Brassicaceae). Plos One, 2012. 7(3).

[15] Gunver-Dalkilic, G. and O. Dayi-Dogru, Determination of Pollen Grain Viability and Germination Levels for Pistachio and Terebinth in Aydin/Turkey Ecology. Pakistan Journal of Botany, 2011. 43(2): p. 841-848.

[16] Punekar, S.A. and K.P.N. Kumaran, Pollen morphology and pollination ecology of Amorphophallus species from North Western Ghats and Konkan region of India. Flora, 2010. 205(5): p. 326-336.

[17] Dieterich, H. and R. Hauff, Meaning of Pollen Analysis for Forest Ecology in Baden-Wurttemberg. Forstwissenschaftliches Centralblatt, 1980. 99(3): p. 120-128.

[18] Kaimakamis, E., et al., Counting airborne pollen at various elevations using an aircraft: how high can aeroallergens fly? Allergy, 2012. 67: p. 250-250.

[19] Skypala, I., et al., Pollen-food syndrome in United Kingdom subjects - which aeroallergens are commonly involved? Allergy, 2007. 62: p. 358-358.

[20] Atrouse, O.M., S.A. Oran, and S.Y. Al-Abbadi, Chemical analysis and identification of pollen grains from different Jordanian

honey samples. International Journal of Food Science and Technology, 2004. 39(4): p. 413-417.

[21] Pierre, J., et al., Efficiency of airborne pollen released by honeybee foraging on pollination in oilseed rape: a wind insect-assisted pollination. Apidologie, 2010. 41(1): p. 109-115.

[22] Zhang, Y., et al., Bayesian analysis of climate change effects on observed and projected airborne levels of birch pollen. Atmospheric Environment, 2013. 68: p. 64-73.

[23] Piffanelli, P., E.J.H. Ross, and J.D. Murphy, Biogenesis and function of the lipidic structures of pollen grains. Sexual Plant Reproduction. 11(2): p. 65-80.

[24] Jiang, J., Z. Zhang, and J. Cao, Pollen wall development: the associated enzymes and metabolic pathways. Plant Biology, 2013. 15(2): p. 249-263.

[25] Pollen Morphology, in Pollen Terminology. 2009, Springer Vienna. p. 15-25.

[26] R., S.K., Pollen Biology and Biotechnology 2003, Enfield, NH: Science Publisher.

[27] Shivanna, K.R., H.F. Linskens, and M. Cresti, pollen viability and pollen vigor. Theoretical and Applied Genetics, 1991. 81(1): p. 38-42.

[28] Williams, J.H. and S.J. Mazer, Pollen-Tiny and ephemeral but not forgotten: New ideas on their ecology and evolution. American Journal of Botany, 2016. 103(3): p. 365-374.

[29] Delph, L.F., M.H. Johannsson, and A.G. Stephenson, How environmental factors affect pollen performance: Ecological and evolutionary perspectives. Ecology, 1997. 78(6): p. 1632-1639.

[30] Lau, T.C. and A.G. Stephenson, Effects of Soil-Phosphorus on Pollen Production, Pollen Size, Pollen Phosphorus-Content, and the Ability to Sire Seeds in Cucurbita-Pepo (Cucurbitaceae). Sexual Plant Reproduction, 1994. 7(4): p. 215-220.

[31] Vesprini, J.L. and E. Pacini, Temperature-dependent floral longevity in two Helleborus species. Plant Systematics and Evolution, 2005. 252(1-2): p. 63-70.

[32] Vesprini, J.L., et al., Changes in cytoplasmic carbohydrate content during Helleborus pollen presentation. Grana, 2002. 41(1): p. 16-20.

[33] Van Herpen, M.M.A., Biochemical Alterations in the Sexual Partners Resulting from Environmental Conditions before Pollination Regulate Processes after Pollination, in Biotechnology and Ecology of Pollen, D. Mulcahy, G. Mulcahy, and E. Ottaviano, Editors. 1986, Springer New York. p. 131-133.

[34] Van Herpen, M.M.A., Effect of season, age and temperature on the protein pattern of pollen and styles in petunia hybrida. Acta Botanica Neerlandica, 1981. 30(4): p. 277-287.

[35] Sears, P. and E. Metcalf, The behaviour of pollen starch in a Geranium and its bud sport. Journal of Genetics, 1926. 17(1): p. 33-42.

[36] Herpen, M.M.A., Biochemical Alterations in the Sexual Partners Resulting from Environmental Conditions before Pollination Regulate Processes after Pollination, in Biotechnology and Ecology of Pollen: Proceedings of the International Conference on the Biotechnology and Ecology of Pollen, 911 July, 1985, University of Massachusetts, Amherst, MA, USA, D.L. Mulcahy, G.B. Mulcahy, and E. Ottaviano, Editors. 1986, Springer New York: New York, NY. p. 131-133.

[37] Van Herpen, M.M.A., Extracts from styles, developed at different temperatures, and their effect on compatibility of petunia hybrida in excised-style culture. Acta Botanica Neerlandica, 1984. 33(2): p. 195-203.

[38] Frazee, J.E. and R.J. Marquis, Environmental Contribution to Floral Trait Variation in Chamaecrista-Fasciculata (Fabaceae, Caesalpinoideae). American Journal of Botany, 1994. 81(2): p. 206-215.

[39] Wang, Z.Y., et al., Viability and longevity of pollen from transgenic and nontransgenic tall fescue (Festuca arundinacea) (Poaceae) plants. American Journal of Botany, 2004. 91(4): p. 523-530.

[40] Steinacher, G. and J. Wagner, Effect of temperature on the progamic phase in high-mountain plants. Plant Biology, 2012. 14(2): p. 295-305.

[41] Johannsson, M.H., J.A. Winsor, and A.G. Stephenson, Genetic and environmental effects on in vitro pollen tube growth in Cucurbita. Pollen-Pistil Interactions and Pollen Tube Growth, 1994. 12: p. 307-309.

[42] Lahlali, R., et al., ATR-FTIR spectroscopy reveals involvement of lipids and proteins of intact pea pollen grains to heat stress tolerance. Frontiers in Plant Science, 2014. 5.

[43] Jiang, Y., et al., Seed set, pollen morphology and pollen surface composition response to heat stress in field pea. Plant Cell Environ, 2015. 38(11): p. 2387-97.

[44] Johansson, S.G.O., Milestones in understanding allergy and its diagnosis. Clinical & Experimental Allergy Reviews, 2002. 2(1): p. 2-7.

[45] Weber, R.W., Pollen Identification. Annals of Allergy, Asthma & Immunology. 80(2): p. 141-148.

[46] Cecchi, L., Introduction, in Allergenic Pollen, M. Sofiev and K.-C. Bergmann, Editors. 2013, Springer Netherlands. p. 1-7.

[47] Descolas-Gros, C. and C. Scholzel, Stable isotope ratios of carbon and nitrogen in pollen grains in order to characterize plant functional groups and photosynthetic pathway types. New Phytol, 2007. 176(2): p. 390-401.

[48] Bedinger, P., The remarkable biology of pollen. The Plant Cell, 1992. 4(8): p. 879-887.

[49] Lomax, B.H., et al., A novel palaeoaltimetry proxy based on spore and pollen wall chemistry. Earth and Planetary Science Letters, 2012. 353: p. 22-28.

[50] Blackmore, S., Pollen and spores: Microscopic keys to understanding the earth's biodiversity. Plant Systematics and Evolution, 2007. 263(1-2): p. 3-12.

[51] Davis, M.B., Pollen grains in lake sediments: redeposition caused by seasonal water circulation. Science, 1968. 162(3855): p. 796-9.

[52] Hunter, P., All the evidence. EMBO Reports, 2006. 7(4): p. 352-354.

[53] Song, X.-Y., Y.-F. Yao, and W.-D. Yang, Pollen Analysis of Natural Honeys from the Central Region of Shanxi, North China. PLoS ONE, 2012. 7(11): p. e49545.

[54] Williams, S., et al., Establishing tobacco origin from pollen identification: an approach to resolving the debate. J Forensic Sci, 2014. 59(6): p. 1642-9.

[55] Mullins, J. and J. Emberlin, Sampling pollens. Journal of Aerosol Science, 1997. 28(3): p. 365-370.

[56] Bennett, K.D., Textbook of pollen analysis. K. Faegri, j. Iversen (4th edn by k. faegri, p. e. kaland, k. krzywinski), Publisher John Wiley and Sons, Chichester 1989 (328 pp) 51.00 ISBN 0 471 92178 5. Journal of Quaternary Science, 1990. 5(3): p. 254-255.

[57] Lacey, J. and J. Venette, Outdoor air sampling techniques- In: Bioaerosol Handbook, C.S. Cox and C.M. Wathes, Editors. 1995, CRC Press: Boca Raton FL.

[58] Longhi, S., et al., Biomolecular identification of allergenic pollen: a new perspective for aerobiological monitoring? Annals of Allergy, Asthma & Immunology, 2009. 103(6): p. 508-514.

[59] Hirst, J.M., An automatic volumetric spore trap. Annals of Applied Biology, 1952. 39(2): p. 257-265.

[60] Hochuli, P.A. and S. Feist-Burkhardt, A boreal early cradle of angiosperms? Angiosperm-like pollen from the Middle Triassic of the Barents Sea (Norway). Journal of Micropalaeontology, 2004. 23: p. 97-104.

[61] Holst, I., J.E. Moreno, and D.R. Piperno, Identification of teosinte, maize, and Tripsacum in Mesoamerica by using pollen, starch grains, and phytoliths. Proceedings of the National Academy of Sciences of the United States of America, 2007. 104(45): p. 17608-17613.

[62] Punyasena, S.W., et al., Classifying black and white spruce pollen using layered machine learning. New Phytologist, 2012. 196(3): p. 937-944.

[63] Salih, A., et al., Confocal imaging of exine as a tool for grass pollen analysis. Grana, 1997. 36(4): p. 215-224.

[64] Sivaguru, M., et al., Capturing the Surface Texture and Shape of Pollen: A Comparison of Microscopy Techniques. PLoS ONE, 2012. 7(6): p. e39129.

[65] Souza, C.D., et al., A Novel Fatty Acyl-CoA Synthetase Is Required for Pollen Development and Sporopollenin Biosynthesis in Arabidopsis. Plant Cell, 2009. 21(2): p. 507-525.

[66] Mller-Germann, I., et al., Quantitative DNA Analyses for Airborne Birch Pollen. PLoS ONE, 2015. 10(10): p. e0140949.

[67] Zhang, M., et al., DGAT1 and PDAT1 Acyltransferases Have Overlapping Functions in Arabidopsis Triacylglycerol Biosynthesis and Are Essential for Normal Pollen and Seed Development. Plant Cell, 2009. 21(12): p. 3885-3901.

[68] Wlodarski, M., et al. Fluorescence excitation-emission matrices of selected biological materials. 2006.

[69] Boyain-Goitia, A.R., et al., Single-pollen analysis by laser-induced breakdown spectroscopy and Raman microscopy. Applied Optics, 2003. 42(30): p. 6119-6132.

[70] Seifert, S., V. Merk, and J. Kneipp, Identification of aqueous pollen extracts using surface enhanced Raman scattering (SERS) and pattern recognition methods. Journal of Biophotonics, 2016. 9(1-2): p. 181-189.

[71] Seifert, S., et al., Taxonomic relationships of pollens from matrix-assisted laser desorption/ionization time-of-flight mass spectrometry data using multivariate statistics. Rapid Communications in Mass Spectrometry, 2015. 29(12): p. 1145-1154.

[72] Rsch, P., et al., On-Line Monitoring and Identification of Bioaerosols. Analytical Chemistry, 2006. 78(7): p. 2163-2170.

[73] Schulte, F., et al., Chemical Characterization and Classification of Pollen. Analytical Chemistry, 2008. 80(24): p. 9551-9556.

[74] Guedes, A., et al., Pollen Raman spectra database: Application to the identification of airborne pollen. Talanta, 2014. 119: p. 473-478.

[75] Schulte, F., U. Panne, and J. Kneipp, Molecular changes during pollen germination can be monitored by Raman microspectroscopy. Journal of Biophotonics, 2010. 3(8-9): p. 542-547.

[76] Willis, R.C., Classifying pollen with Raman. Analytical Chemistry, 2009. 81(1): p. 4-4.

[77] Schulte, F., et al., Characterization of pollen carotenoids with in situ and high-performance thin-layer chromatography supported resonant Raman spectroscopy. Anal Chem, 2009. 81(20): p. 8426-33.

[78] Manoharan, R., et al., Resonance Raman-Spectra of Aqueous Pollen Suspensions with 222.5-242.4-Nm Pulsed Laser Excitation. Applied Spectroscopy, 1991. 45(2): p. 307-311.

[79] Laucks, M.L., et al., Physical and chemical (Raman) characterization of bioaerosols - Pollen. Journal of Aerosol Science, 2000. 31(3): p. 307-319.

[80] Zimmermann, B., Characterization of pollen by vibrational spectroscopy. Appl Spectrosc, 2010. 64(12): p. 1364-73.

[81] Dell'Anna, R., et al., Pollen discrimination and classification by Fourier transform infrared (FT-IR) microspectroscopy and machine learning. Anal Bioanal Chem, 2009. 394(5): p. 1443-52.

[82] Parodi, G., P. Dickerson, and J. Cloud, Pollen Identification by Fourier Transform Infrared Photoacoustic Spectroscopy. Applied Spectroscopy, 2013. 67(3): p. 342-348.

[83] Gottardini, E., et al., Use of Fourier transform infrared (FT-IR) spectroscopy as a tool for pollen identification. Aerobiologia, 2007. 23(3): p. 211-219.

[84] Gergen, I., F. Radu, and M. Poiana, Bee's pollen moisture determination by halogen lamp infrared drying method. Revista De Chimie, 2005. 56(1): p. 54-56.

[85] Sowa, S. and K.F. Connor, Biochemical-Changes during Pollen Germination Measured in-Vivo by Infrared-Spectroscopy. Plant Science, 1995. 105(1): p. 23-30.

[86] Sowa, S., K.F. Connor, and E.E. Roos, Biochemical-Changes during Pollen Germination Measured Invivo by Infrared-Spectroscopy. Plant Physiology, 1993. 102(1): p. 136-136.

[87] Sowa, S., K.F. Connor, and L.E. Towill, Temperature-Changes in Lipid and Protein-Structure Measured by Fourier-Transform Infrared Spectrophotometry in Intact Pollen Grains. Plant Science, 1991. 78(1): p. 1-9.

[88] Bagcioglu, M., B. Zimmermann, and A. Kohler, A Multiscale Vibrational Spectroscopic Approach for Identification and Biochemical Characterization of Pollen. PLoS ONE, 2015. 10(9): p. e0137899.

[89] Zimmermann, B. and A. Kohler, Infrared spectroscopy of pollen identifies plant species and genus as well as environmental conditions. PLoS One, 2014. 9(4): p. e95417.

[90] Pappas, C.S., et al., New method for pollen identification by FT-IR spectroscopy. Applied Spectroscopy, 2003. 57(1): p. 23-27.

[91] Bhargava, R., B.G. Wall, and J.L. Koenig, Comparison of the FT-IR mapping and imaging techniques applied to polymeric systems. Applied Spectroscopy, 2000. 54(4): p. 470-479.

[92] Schmitt, J. and H.-C. Flemming, FTIR-spectroscopy in microbial and material analysis. International Biodeterioration & Biodegradation, 1998. 41(1): p. 1-11.

[93] Votteler, M., et al., Non-contact, Label-free Monitoring of Cells and Extracellular Matrix using Raman Spectroscopy. Journal of Visualized Experiments : JoVE, 2012(63): p. 3977.

[94] Fanti, G., et al., Non-destructive dating of ancient flax textiles by means of vibrational spectroscopy. Vibrational Spectroscopy, 2013. 67: p. 61-70.

[95] Culka, A., Non-destructive detection of biphosphammite and nickel-boussingaultite - two NH4 group containing minerals using vibrational spectroscopy. International Journal of Astrobiology, 2008. 7(1): p. 82-83.

[96] Schulz, H., Rapid analysis of medicinal and aromatic plants by non-destructive vibrational spectroscopy methods. WOCMAP III: Quality, Efficacy, Safety, Processing and Trade in MAPs, 2005(679): p. 181-187.

[97] Isaksson, T., et al., Non-destructive determination of fat, moisture and protein in salmon fillets by use of near-infrared diffuse spectroscopy. Journal of the Science of Food and Agriculture, 1995. 69(1): p. 95-100.

[98] Bellisola, G. and C. Sorio, Infrared spectroscopy and microscopy in cancer research and diagnosis. American Journal of Cancer Research, 2012. 2(1): p. 1-21.

[99] Diem, M., et al., A decade of vibrational micro-spectroscopy of human cells and tissue (1994-2004). Analyst, 2004. 129(10): p. 880-885.

[100] Kohler, A., et al., High-Throughput Biochemical Fingerprinting of Saccharomyces cerevisiae by Fourier Transform Infrared Spectroscopy. Plos One, 2015. 10(2).

[101] Shapaval, V., et al., A high-throughput microcultivation protocol for FTIR spectroscopic characterization and identification of fungi. Journal of Biophotonics, 2010. 3(8-9): p. 512-521.

[102] Rosa, F., et al., A comprehensive high-throughput FTIR spectroscopy-based method for evaluating the transfection event: estimating the transfection efficiency and extracting associated metabolic responses. Analytical and Bioanalytical Chemistry, 2015. 407(26): p. 8097-8108.

[103] Lacombe, C., et al., Rapid screening of classic galactosemia patients: a proof-of-concept study using high-throughput FTIR analysis of plasma. Analyst, 2015. 140(7): p. 2280-2286.

[104] Untereiner, V., et al., Bile analysis using high-throughput FTIR spectroscopy for the diagnosis of malignant biliary strictures: a pilot study in 57 patients. Journal of Biophotonics, 2014. 7(3-4): p. 241-253.

[105] Najbjerg, H., et al., High-throughput FTIR spectroscopy of intact HepG2 cells reveals additive and non-additive effects of individual fatty acids when given as mixtures. Journal of Biophotonics, 2013. 6(5): p. 446-456.

[106] Risse, B., et al., FIM, a Novel FTIR-Based Imaging Method for High Throughput Locomotion Analysis. Plos One, 2013. 8(1).

[107] Scholz, T., V.V. Lopes, and C.R.C. Calado, High-throughput analysis of the plasmid bioproduction process in Escherichia coli by FTIR spectroscopy. Biotechnology and Bioengineering, 2012. 109(9): p. 2279-2285.

[108] Zimmerman, B., et al., Analysis of Allergenic Pollen by FTIR Microspectroscopy. Anal Chem, 2016. 88(1): p. 803-11.

[109] Marshall, C.P., et al., Combined micro-Fourier transform infrared (FTIR) spectroscopy and micro-Raman spectroscopy of Proterozoic acritarchs: A new approach to Palaeobiology. Precambrian Research, 2005. 138(3-4): p. 208-224.

[110] Fraser, W.T., et al., UV-B absorbing pigments in spores: biochemical responses to shade in a high-latitude birch forest and implications for sporopollenin-based proxies of past environmental change. 2011, 2011.

[111] Lomax, B.H., et al., A novel palaeoaltimetry proxy based on spore and pollen wall chemistry. Earth and Planetary Science Letters, 2012. 353354: p. 22-28.

[112] Pepponi, G., et al., Total reflection X-ray fluorescence analysis of pollen as an indicator for atmospheric pollution. Spectrochimica Acta Part B: Atomic Spectroscopy, 2004. 59(8): p. 1205-1209.

[113] Ivleva, N.P., R. Niessner, and U. Panne, Characterization and discrimination of pollen by Raman microscopy. Anal Bioanal Chem, 2005. 381(1): p. 261-7.

[114] Blokker, P., et al., The Occurrence of p-coumaric Acid and Ferulic Acid in Fossil Plant Materials and their Use as UV-proxy. Plant Ecology. 182(1): p. 197-207.

[115] Krishnan, R.S. and R.K. Shankar, Raman effect: History of the discovery. Journal of Raman Spectroscopy, 1981. 10(1): p. 1-8.

[116] Smith, E. and G. Dent, Resonance Raman Scattering, in Modern Raman Spectroscopy A Practical Approach. 2005, John Wiley & Sons, Ltd. p. 93-112.

[117] Chen, T.C., D.A. Shea, and M.D. Morris, Effect of Hydrogen Peroxide Bleaching on Bone Mineral/Matrix Ratio. Applied Spectroscopy, 2002. 56(8): p. 1035-1037.

[118] Matousek, P., M. Towrie, and A.W. Parker, Fluorescence background suppression in Raman spectroscopy using combined Kerr gated and shifted excitation Raman difference techniques. Journal of Raman Spectroscopy, 2002. 33(4): p. 238-242.

[119] Golcuk, K., et al., Is photobleaching necessary for Raman imaging of bone tissue using a green laser? Biochimica et Biophysica Acta (BBA) - Biomembranes, 2006. 1758(7): p. 868-873.

[120] Pummer, B.G., et al., Chemistry and morphology of dried-up pollen suspension residues. Journal of Raman Spectroscopy, 2013. 44(12): p. 1654-1658.

[121] Wang, C., et al., Photophoretic trapping-Raman spectroscopy for single pollens and fungal spores trapped in air. Journal of

Quantitative Spectroscopy and Radiative Transfer, 2015. 153: p. 4-12.

[122] Redding, B., M.J. Schwab, and Y.L. Pan, Raman Spectroscopy of Optically Trapped Single Biological Micro-Particles. Sensors, 2015. 15(8): p. 19021-19046.

[123] Harvey, T.J., et al., Spectral discrimination of live prostate and bladder cancer cell lines using Raman optical tweezers. Journal of Biomedical Optics, 2008. 13(6).

[124] Harvey, T.J., et al., Classification of fixed urological cells using Raman tweezers. Journal of Biophotonics, 2009. 2(1-2): p. 47-69.

[125] Lin, J., A.G. Hart, and Y.Q. Li, Optical pulling of airborne absorbing particles and smut spores over a meter-scale distance with negative photophoretic force. Applied Physics Letters, 2015. 106(17).

[126] Mularczyk-Oliwa, M., et al., Comparison of fluorescence spectroscopy and FTIR in differentiation of plant pollens. Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 2012. 97: p. 246-254.

[127] Crowe, J.H., F.A. Hoekstra, and L.M. Crowe, Membrane Phase-Transitions Are Responsible for Imbibitional Damage in Dry Pollen. Proceedings of the National Academy of Sciences of the United States of America, 1989. 86(2): p. 520-523.

[128] Zimmermann, B., et al., Vibrational microspectroscopy enables chemical characterization of single pollen grains as well as comparative analysis of plant species based on pollen ultrastructure. Planta, 2015. 242(5): p. 1237-1250.

[129] Wolkers, W.F. and F.A. Hoekstra, Aging of Dry Desiccation-Tolerant Pollen Does Not Affect Protein Secondary Structure. Plant Physiology, 1995. 109(3): p. 907-915.

[130] Wolkers, W.F. and F.A. Hoekstra, Heat stability of proteins in desiccation-tolerant cattail (Typha latifolia L) pollen - A Fourier transform infrared spectroscopic study. Comparative Biochemistry and Physiology a-Physiology, 1997. 117(3): p. 349-355.

[131] Joseph, V., et al., Surface-enhanced Raman scattering with silver nanostructures generated in situ in a sporopollenin biopolymer matrix. Chemical Communications, 2011. 47(11): p. 3236-3238.

[132] Zimmermann, B. and A. Kohler, Optimizing Savitzky-Golay Parameters for Improving Spectral Resolution and Quantification in Infrared Spectroscopy. Applied Spectroscopy, 2013. 67(8): p. 892-902.

[133] Mirabella, F.M., Practical Spectroscopy Series: Internal ReectionSpectroscopy: Theory and Applications, M. Dekker, Editor. 1993, CRC Press: New York. p. 17-52.

[134] Giaever, G., et al., Functional profiling of the Saccharomyces cerevisiae genome. Nature, 2002. 418(6896): p. 387-391.

[135] Warringer, J., et al., High-resolution yeast phenomics resolves different physiological features in the saline response. Proceedings of the National Academy of Sciences, 2003. 100(26): p. 15724-15729.

[136] Parsons, A.B., et al., Exploring the Mode-of-Action of Bioactive Compounds by Chemical-Genetic Profiling in Yeast. Cell, 2006. 126(3): p. 611-625.

[137] Philippe, S., et al., Characterization using Raman microspectroscopy of arabinoxylans in the walls of different cell types during the development of wheat endosperm. Journal of Agricultural and Food Chemistry, 2006. 54(14): p. 5113-5119.

[138] Pietrzak, L.N. and S.S. Miller, Microchemical structure of soybean seeds revealed in situ by ultraspatially resolved synchrotron Fourier transformed infrared microspectroscopy. Journal of Agricultural and Food Chemistry, 2005. 53(24): p. 9304-9311.

[139] Rebuffo, C.A., et al., Reliable and rapid identification of Listeria monocytogenes and Listeria species by artificial neural network-based Fourier transform infrared spectroscopy. Applied and Environmental Microbiology, 2006. 72(2): p. 994-1000.

[140] Savolainen, O., et al., Genetic variation in cessation of growth and frost hardiness and consequences for adaptation of Pinus sylvestris to climatic changes. Forest Ecology and Management, 2004. 197(13): p. 79-89.

[141] Yu, P., R. Wang, and Y. Bai, Reveal protein molecular structural-chemical differences between two types of winterfat (forage) seeds with physiological differences in low temperature tolerance using synchrotron-based Fourier transform infrared microspectroscopy. Journal of Agricultural and Food Chemistry, 2005. 53(24): p. 9297-9303.

[142] Zhang, X.W. and P.Q. Yu, Using a Non-invasive Technique in Nutrition: Synchrotron Radiation Infrared Microspectroscopy Spectroscopic Characterization of Oil Seeds Treated with Different Processing Conditions on Molecular Spectral Factors Influencing Nutrient Delivery. Journal of Agricultural and Food Chemistry, 2014. 62(26): p. 6199-6205.

[143] Chylinska, M., M. Szymanska-Chargot, and A. Zdunek, Imaging of polysaccharides in the tomato cell wall with Raman microspectroscopy. Plant Methods, 2014. 10.

[144] Silge, A., et al., Shedding light on host niches: label-free in situ detection of Mycobacterium gordonae via carotenoids in macrophages by Raman microspectroscopy. Cellular Microbiology, 2015. 17(6): p. 832-842.

[145] Kutuzov, N.P., et al., Orientational Ordering of Carotenoids in Myelin Membranes Resolved by Polarized Raman Microspectroscopy. Biophysical Journal, 2014. 107(4): p. 891-900.

[146] Miller, L.M. and P. Dumas, Chemical imaging of biological tissue with synchrotron infrared light. Biochimica Et Biophysica Acta-Biomembranes, 2006. 1758(7): p. 846-857.

[147] Liu, J.N., et al., Optimally designed narrowband guided-mode resonance reflectance filters for mid-infrared spectroscopy. Optics Express, 2011. 19(24): p. 24182-24197.

[148] Miller, L.M. and R.J. Smith, Synchrotrons versus globars, point-detectors versus focal plane arrays: Selecting the best

source and detector for specific infrared microspectroscopy and imaging applications. Vibrational Spectroscopy, 2005. 38(1-2): p. 237-240.

[149] Dumas, P., et al., Synchrotron infrared microscopy at the French Synchrotron Facility SOLEIL. Infrared Physics & Technology, 2006. 49(1-2): p. 152-160.

[150] Duncan, W.D. and G.P. Williams, Infrared Synchrotron Radiation from Electron Storage-Rings. Applied Optics, 1983. 22(18): p. 2914-2923.

[151] Lukacs, R., et al., Recovery of absorbance spectra of micrometer-sized biological and inanimate particles. Analyst, 2015. 140(9): p. 3273-3284.

[152] Bohren, C.F. and D.R. Huffman, Introduction, in Absorption and Scattering of Light by Small Particles. 2007, Wiley-VCH Verlag GmbH. p. 1-11.

[153] Bassan, P., et al., RMieS-EMSC correction for infrared spectra of biological cells: Extension using full Mie theory and GPU computing. Journal of Biophotonics, 2010. 3(8-9): p. 609-620.

[154] Bassan, P., et al., Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples. Analyst, 2010. 135(2): p. 268-277.

[155] Everall, N., et al., Optimizing depth resolution in confocal Raman microscopy: a comparison of metallurgical, dry corrected, and oil immersion objectives. Appl Spectrosc, 2007. 61(3): p. 251-9.

[156] Whitley, A. and F. Adar, Confocal spectral imaging in tissue with contrast provided by raman vibrational signatures. Cytometry Part A, 2006. 69A(8): p. 880-887.

[157] Savitzky, A. and M.J.E. Golay, Smoothing + Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry, 1964. 36(8): p. 1627.

[158] Martens, H. and E. Stark, Extended Multiplicative Signal Correction and Spectral Interference Subtraction - New Preprocessing Methods for near-Infrared Spectroscopy. Journal of Pharmaceutical and Biomedical Analysis, 1991. 9(8): p. 625-635.

[159] Hassani, S., et al., Analysis of -omics data: Graphical interpretation- and validation tools in multi-block methods. Chemometrics and Intelligent Laboratory Systems, 2010. 104(1): p. 140-153.

[160] Ridgley, D.M., E.C. Claunch, and J.R. Barone, Characterization of Large Amyloid Fibers and Tapes with Fourier Transform Infrared (FT-IR) and Raman Spectroscopy. Applied Spectroscopy, 2013. 67(12): p. 1417-1426.

[161] Perisic, N., et al., Characterising protein, salt and water interactions with combined vibrational spectroscopic techniques. Food Chemistry, 2013. 138(1): p. 679-686.

[162] Westerhuis, J.A., T. Kourti, and J.F. MacGregor, Analysis of multiblock and hierarchical PCA and PLS models. Journal of Chemometrics, 1998. 12(5): p. 301-321.

[163] Kohler, A., et al., Interpreting Several Types of Measurements in Bioscience, in Biomedical Vibrational Spectroscopy. 2007, John Wiley & Sons, Inc. p. 333-356.

[164] Martens, H. and T. Ns, Multivariate Calibration. 1992, Chichester,UK: Wiley.

[165] Geladi, P. and B.R. Kowalski, Partial Least-Squares Regression - a Tutorial. Analytica Chimica Acta, 1986. 185: p. 1-17.

[166] L Cao, K.-A., et al., A Sparse PLS for Variable Selection when Integrating Omics Data, in Statistical Applications in Genetics and Molecular Biology. 2008.

[167] Karaman, ., et al., Comparison of Sparse and Jack-knife partial least squares regression methods for variable selection. Chemometrics and Intelligent Laboratory Systems, 2013. 122: p. 65-77.

[168] BOPT-4000018-03, B.O., Bruker Optics HTS-XT FT-IR High Throughput Screening Extension. 2013.

[169] Galan, C., et al., Pollen monitoring: minimum requirements and reproducibility of analysis. Aerobiologia, 2014. 30(4): p. 385-395.

Publish or Perish
Logan Wilson (1942)

6

Appendix: Peer review articles

This chapter consists of the manuscripts produced as an outcome of the PhD study.

(i) A Multiscale Vibrational Spectroscopic Approach for Identification and Biochemical Characterization of Pollen. ***PLoS ONE*** 2015, 10:E0137899.

(ii) Vibrational microspectroscopy enables chemical characterization of single pollen grains as well as comparative analysis of plant species based on pollen ultrastructure. ***Planta*** 2015, Volume 242. (5) p. 1237-1250.

(iii) Analysis of allergenic pollen by FTIR microspectroscopy. ***Analytical Chemistry*** 2016, 88 (1), pp 803811.

(iv) Monitoring of plant-environment interactions by high throughput FTIR spectroscopy of pollen. ***Submitted*** 2016.

(v) A greenhouse study for assessing environmental effects on pollen composition by FTIR spectroscopy. ***Submitted*** 2016.

6.1 Paper I

PLOS ONE

RESEARCH ARTICLE

A Multiscale Vibrational Spectroscopic Approach for Identification and Biochemical Characterization of Pollen

Murat Bağcıoğlu[1]*, Boris Zimmermann[1], Achim Kohler[1,2]

1 Department of Mathematical Sciences and Technology, Faculty of Environmental Science and Technology, Norwegian University of Life Sciences, Ås, Norway, **2** Nofima AS, Ås, Norway

* murat.bagcioglu@nmbu.no

Citation: Bağcıoğlu M, Zimmermann B, Kohler A (2015) A Multiscale Vibrational Spectroscopic Approach for Identification and Biochemical Characterization of Pollen. PLoS ONE 10(9): e0137899. doi:10.1371/journal.pone.0137899

Editor: Jonathan A Coles, Glasgow University, UNITED KINGDOM

Received: March 9, 2015

Accepted: August 22, 2015

Published: September 16, 2015

Data Availability Statement: All spectral files are available from the Dryad database (DOI: doi:10.5061/dryad.b7g8p).

Funding: This work was supported by the European Commission through the Seventh Framework Programme (FP7-PEOPLE-2012-IEF project N°. 328289) (BZ,AK); The Ministry of Education, Sciences and Sports of the Republic of Croatia (project N°. 098-0982904-2927) (BZ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Nofima AS provided support in the form of a salary for author AK, but did not have any

Abstract

Background

Analysis of pollen grains reveals valuable information on biology, ecology, forensics, climate change, insect migration, food sources and aeroallergens. Vibrational (infrared and Raman) spectroscopies offer chemical characterization of pollen via identifiable spectral features without any sample pretreatment. We have compared the level of chemical information that can be obtained by different multiscale vibrational spectroscopic techniques.

Methodology

Pollen from 15 different species of Pinales (conifers) were measured by seven infrared and Raman methodologies. In order to obtain infrared spectra, both reflectance and transmission measurements were performed on ground and intact pollen grains (bulk measurements), in addition, infrared spectra were obtained by microspectroscopy of multigrain and single pollen grain measurements. For Raman microspectroscopy measurements, spectra were obtained from the same pollen grains by focusing two different substructures of pollen grain. The spectral data from the seven methodologies were integrated into one data model by the Consensus Principal Component Analysis, in order to obtain the relations between the molecular signatures traced by different techniques.

Results

The vibrational spectroscopy enabled biochemical characterization of pollen and detection of phylogenetic variation. The spectral differences were clearly connected to specific chemical constituents, such as lipids, carbohydrates, carotenoids and sporopollenins. The extensive differences between pollen of *Cedrus* and the rest of Pinaceae family were unambiguously connected with molecular composition of sporopollenins in pollen grain wall, while pollen of *Picea* has apparently higher concentration of carotenoids than the rest of the family. It is shown that vibrational methodologies have great potential for systematic collection of data on ecosystems and that the obtained phylogenetic variation can be well explained by the biochemical composition of pollen. Out of the seven tested methodologies,

additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific role of this author is articulated in the "Author Contributions" section.

Competing Interests: The authors have declared the following interests: Achim Kohler is employed by Nofima AS. There are no patents, products in development or marketed products to declare. This did not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

the best taxonomical differentiation of pollen was obtained by infrared measurements on bulk samples, as well as by Raman microspectroscopy measurements of the corpus region of the pollen grain. Raman microspectroscopy measurements indicate that measurement area, as well as the depth of focus, can have crucial influence on the obtained data.

Introduction

Pollen grains are characteristic to plant families, genera or even species, thus revealing valuable information on ecology, forensics, climate change, insect migration, food sources and aeroallergens. They are extremely resilient to decay, and pollen microfossils can have well-preserved morphology for millions of years. Thus in palaeoecology, palaeobotany, biostratigraphy and biogeography, identification of pollen microfossils is important for the reconstruction of past flora, population sizes and terrestrial communities. Moreover, microfossil data is vital for the reconstruction of past environments and for the understanding of the causes of environmental changes.

The main drawback of pollen analysis is that pollen grains are still studied by measuring their morphology employing conventional microscopy. Contemporary pollen identification is predominantly performed by optical microscopy. Thus, routine analysis is labour-intensive and uneconomical, where thousands of discrete pollen grains are examined by an experienced scientist after complex chemical pre-treatment of samples. In addition to optical microscopy, different microscopic techniques have been attempted for measurement of pollen, such as differential interference contrast microscopy, phase contrast microscopy, fluorescence microscopy, scanning electron microscopy, confocal laser scanning microscopy and other alternative methods for obtaining detailed morphology of pollen [1]. Unfortunately, these methods are time-consuming and expensive and provide only identification based on morphological analysis, while chemical characterization of samples cannot be obtained. Alternative analyses, such as biochemical and sequencing methods, are even more complex and expensive [2].

Vibrational (infrared and Raman) spectroscopies offer an alternative approach to pollen analysis. Instead of morphological characterization of grains, in vibrational spectroscopy pollen analysis is based on chemical characterization via identifiable spectral features without any sample pretreatment [3–11]. Fourier transform infrared (FTIR) spectroscopy provides a direct assessment of biochemical composition of pollen which has proven to be species specific [7, 8]. Therefore, infrared spectroscopy is developing as an important tool in biology for studying phylogenetic differences between pollen grains of diverse plant species.

While most of the studies conducted with infrared spectroscopy were performed by bulk measurements of pollen samples, single grain measurements by FTIR microspectroscopy on single pollen grains are rare [4]. In addition to FTIR spectroscopy, characterization and classification of pollen grains have been performed by Raman spectroscopy, including detailed microspectroscopy measurements [9, 10, 12]. Although Raman measurements of pollen are often obstructed by the strong fluorescence background and laser-induced degradation, the studies have nevertheless shown a potential of Raman spectroscopy in ecology and aerobiology.

Although several comparative studies employing vibrational spectroscopy have been performed on pollen [5, 8], none of the studies compared the level of chemical information that can be obtained by the different spectroscopic techniques that operate on different scales and can highlight different chemical aspects of pollen. Pollen studies are almost exclusively focused on identification of pollen grains, while the chemical characterization has remained

underutilized. However, chemical characterization of pollen could present important information for palaeosciences regarding that some pollen chemicals have long-term stability spanning millions of years [13]. For example, sporopollenin, the main constituent of pollen grain wall, is extremely resilient to a variety of abiotic stresses and well preserved in the fossil record, while its chemical composition depends on environment experienced during the growth of the parent plant [14]. Since phenolic 98i blocks of sporopollenins have characteristic spectral features, vibrational spectroscopy offers a novel approach for chemical analysis of pollen grain wall and improvement of palaeoecological studies [15].

In this study, pollen samples of Pinales (conifer) species were investigated. The measured species are growing in Northern hemisphere and they belong to the most numerous extant group of conifers. The chosen sample set, consisting of 15 different species of Pinaceae and Podocarpaceae families, offers phylogenetic assortment and representation at the family and genus level due to diversity in grain size, shape, and relative biochemical composition. Different vibrational spectroscopic methodologies have been used in a multiscale setting to obtain spectra from the same pollen samples by bulk measurements as well as microscopic measurements at sub-grain level. Application of pollen measurements often requires measurement of an individual pollen grain, such as in monitoring of aeroallergens. Therefore, it is important to assess if the advancement in bulk FTIR measurements of pollen can be repeated at the single grain level by FTIR microspectroscopy, despite the anticipated technical difficulties [3, 16]. Five methodologies were used in FTIR measurements: attenuated total reflectance (ATR) measurements of ground and intact pollen grains, microscopic transmission measurements of multi-grain and single pollen grains, and potassium bromide (KBr) pellet measurements of ground pollen. For Raman measurements, two different regions from the same pollen grains were measured, namely saccus and corpus area of bisaccate pollen samples. In total, seven different FTIR and Raman methodologies have been covered, and subsequently combined with multi-block data analysis. The different vibrational methodologies obtain complementary information on pollen substructures, such as grain wall and grain interior, as well on pollen chemical components, such as sporopollenins, proteins, lipids and carotenoids. Multi-block multivariate methods have been used for integrating the different spectroscopic data sets and for extracting and visualizing common underlying patterns of mutual information of the different spectroscopic data blocks, thus enabling interpretation of spectroscopic measurements.

Materials and Methods

Samples

15 pollen samples, belonging to 15 different Pinales species, were collected at the Botanical Garden of the Faculty of Science, the University of Zagreb, during pollination season in the spring of 2012 (Table 1). Each pollen sample was collected from one tree, and it includes pollen grains from at least 10 male cones (pollen strobili). The pollen samples were kept at room temperature for 24 hours, and subsequently stored at -15°C in microcentrifuge tubes until measurements time. The study was a part of government-funded research, and has been conducted with the full cooperation of the administrations of the University of Zagreb. The field studies did not involve endangered or protected species.

For chemical characterization of pollen a set of model compounds was measured by FTIR (ATR) and Raman spectroscopies to correlate with high positive or negative values in the principal component analyses correlation plots. Tristearin (1,3-di(octadecanoyloxy)propan-2-yl octadecanoate), triolein (2,3-bis[[(Z)-octadec-9-enoyl]oxy]propyl (Z)-octadec-9-enoate), phosphatidistearoylcholine (1,2-distearoyl-rac-glycero-3-phosphocholine), phosphatidioleylcholine (1,2-dioleoyl-sn-glycero-3-phosphocholine), stearic acid (octadecanoic acid), oleic acid

Table 1. List of the pollen samples.

Family	Genus	Species	Name
Pinaceae	*Abies*	*A. pinsapo*	Spanish Fir
		A. cephalonica	Greek Fir
	Cedrus	*C. atlantica*	Atlas Cedar
	Picea	*P. omorika*	Serbian Spruce
		P. orientalis	Caucasian Spruce
		P. pungens	Blue Spruce
	Pinus	*P. banksiana*	Jack Pine
		P. peuce	Macedonian Pine
		P. mugo	Mountain Pine
		P. nigra	European Black Pine
		P. resinosa	Red Pine
		P. sylvetris	Scots Pine
		P. tabuliformis	Chinese Red Pine
		P. wallichiana	Himalayan Pine
Podocarpaceae	***Podocarpus***	*P. neriifolius*	Brown Pine

doi:10.1371/journal.pone.0137899.t001

((9Z)-octadec-9-enoic acid), β-carotene, *p*-coumaric acid, ferulic acid, caffeic acid, sinapic acid, hydro-*p*-coumaric acid, hydroferulic acid, hydrocaffeic acid, cellulose, amylose, amylopectin, arabinoxylan, pectin, β-D-glucan, sucrose, trehalose, fructose, glucose, gluten, α-casein, β-casein, κ-casein were purchased from Merck (Darmstadt, Germany) and Sigma-Aldrich (St. Louis, United States), and used without further purification.

Spectroscopic analyses

Seven different FTIR and Raman methodologies have been used for the measurement of the same pollen samples. For infrared measurements: reflectance FTIR spectroscopy by Attenuated Total Reflectance (ATR) of ground (1) and intact pollen grains (2); transmission FTIR spectroscopy of KBr pellets (3); transmission FTIR microspectroscopy of multigrain (4) and single pollen grain (5). For Raman microspectroscopy measurements, spectra were obtained from the same pollen grains by focusing on the different substructures: saccus (6) and corpus part (7) of pollen grain.

Infrared measurements. Microscopic transmission FTIR measurements (μFTIR) were performed using an Equinox 55 FTIR spectrometer with a IRScopeII IR microscope (Bruker Optik GmbH, Germany). The system is equipped with a globar mid-IR source and a liquid nitrogen-cooled mercury cadmium telluride (MCT) detector. The pollen samples were deposited onto 3 mm thickness of ZnSe optical windows without any sample pretreatment. The spectra were obtained in the 4000–600 cm^{-1} spectral range, with a spectral resolution of 4 cm^{-1}, and with 128 scans. Samples were measured using a 15× objective, with different aperture sizes, depending on the size of the sample. The aperture sizes for single pollen grain measurements were ranging from 30xm (for smallest) to 100μm (for biggest) whereas for multigrain pollen samples, fully open aperture (200x200μm) was used. Background (reference) spectra were recorded immediately before starting each measurement by measuring empty areas of 3mm ZnSe slides. Visible images of the measured pollen grains were obtained by a CCD-camera coupled to the microscope. The microscope was equipped with a computer-controlled x, y stage. The Bruker system was controlled with OPUS 6.0 software (Bruker Optik GmbH, Germany).

The reflectance infrared spectra were recorded using an Equinox 55 FTIR spectrometer (Bruker Optik GmbH, Germany) with an ATR accessory. The ATR IR spectra were recorded with a total of 32 scans and spectral resolution of 4 cm^{-1}, using the 1.8 mm thickness of horizontal single-reflection ATR diamond prism with 45° angle of incidence on a MIRacle ATR accessory (PIKE Technologies, USA). Background (reference) spectra were recorded immediately before starting each measurement using the sample-free setup. Pollen samples were measured using two different sampling methodologies: intact samples as collected in nature, and samples ground using an agate pestle and mortar. For ground measurement, batch of 5 mg pollen sample was grounded to a fine homogenous powder using an agate pestle and mortar, and roughly 1 mg of sample per measurement was deposited onto ATR crystal (3 replicate measurements were obtained). Data acquisition and instrument control was carried out using the OPUS 6.0 software (Bruker Optik GmbH, Germany).

The transmission spectra of KBr sample pellets were collected with 4 cm^{-1} resolution acquiring a total of 30 scans on an MB102 single beam FTIR spectrometer (ABB Bomem, Canada), equipped with CsI optics and a DTGS detector (Deuterated Triglycine Sulfate). Background (reference) spectra were recorded immediately before starting each measurement using the sample-free setup. KBr pellets were prepared by mixing approx. 1 mg of a sample with approx. 100 mg of KBr and grounded the mixture to a fine homogenous powder mixture using an agate pestle and mortar. Ground powder samples were transferred to the 10 mm pellet die assembly and the KBr matrix was then cold-pressed without degassing into a transparent disk. Three KBr pellets were prepared and recorded for each pollen sample. Data acquisition and instrument control was carried out using the GRAMS/32 software (Galactic Industries Corp., USA)

Raman measurements. By Raman microspectroscopy (μRaman), spectra of different regions of pollen grains were obtained by manually focusing on either saccus or corpus region of the grain. Raman spectra were recorded by a LabRam HR 800 Raman microscope (Horiba Scientific, France). The excitation wavelength of 632.8 nm was generated by a He–Ne laser. A 100× objective (Olympus, France) was used for focusing and collecting scattered Raman light. The laser power was approximately 15 mW on the sample surface. The confocal hole was set to 20 μm, with a laser spot diameter of approx. 0.8 μm, a spectral resolution of 2 cm^{-1}, and an exposure time of 4 x 10 s. The Raman scattering was dispersed with a 300 lines/mm grating, which resulted in spectra in range 409–2611 cm^{-1}. Data acquisition and instrument control was carried out using the LabSpec 5.45 software (Horiba Scientific, France).

Scanning electron microscopy. The scanning electron microscope (SEM) images were taken with Zeiss EVO 50 Extended Pressure scanning electron microscope (Carl Zeiss AG, Jena, Germany). The desiccated pollen samples were attached to SEM stubs (covered with double-stick tape) without prior pre-treatment with chemicals. The samples were coated with gold-palladium and measured with the SEM.

Data analysis

Spectral pre-processing. For the analysis of the spectral sets, the spectral region of 1900 to 700 cm^{-1} was selected. This spectral region contains bands that are distinctive for pollen grains [3, 4, 7, 8]. Spectra were smoothed and transformed to second derivative form by the Savitzky-Golay (SG) algorithm using a polynomial of power 2 with different window sizes depending on the spectral data sets (15 for bulk FTIR measurements, 17 for FTIR microspectroscopy, and 25 for Raman measurements). The window sizes were chosen according to the number of channel readings and the noise level in the respective data set [17]. After derivation by the SG algorithm, spectra were processed using extended multiplicative signal correction (EMSC). The SG

algorithm was used to enhance spectral features, while the EMSC pre-processing was used for normalization and for separation of chemical and physical variations in vibrational spectra [17, 18].

Multiblock principal component analysis. Principal component analysis (PCA) of pre-processed spectra was used to evaluate biochemical differences between different pollen taxa by displaying main variation patterns in score plots for each spectroscopic method. The multiblock method Consensus Principal Component Analysis (CPCA) was used to combine the seven measurement methods by integrating multivariate signals from each technique (i.e. data block) in one data model [19–22]. Following this approach, the relations between the molecular signatures traced by different techniques were obtained, as well as the variation pattern in the whole (multiblock) data set. In order to integrate different data blocks for CPCA, a row-to-row correspondence between the data blocks was obtained. Therefore, the spectra were averaged over all replicates for each spectroscopic method and integrated as seven blocks: Block 1) KBR (transmission FTIR of KBr pellets), block 2) ATI (ATR-FTIR of intact pollen), block 3) ATG (ATR-FTIR of ground pollen), block 4) MGR (transmission μFTIR of multigrain), block 5) SGR (transmission μFTIR of single grain), block 6) RMC (Raman of corpus region) and block 7) RMS (Raman of saccus region).

504 spectra were analysed in total, including at least 3 replicates per sample for each methodology. For bulk measurements (KBR, ATI and ATG), replicates refer to measurements of distinct aliquots of the pollen sample (approx. 1 mg per measurement). For multigrain μFTIR measurements (MGR), replicates refer to measurements of distinct regions on the microscope slide. For single grain measurements (SGR, RMC and RMS), replicates refer to measurements of distinct pollen grains. For MGR and SGR measurements, the number of replicate measurements was between 5 and 10 spectra, and it was not uniform for all species. CPCA was performed as described in literature [19, 23]. In order to illustrate the relations and variation patterns in sample and variable spaces, score and correlation loading plots were used [19, 24, 25]. In order to visualize variables in correlation loading plots and in order to avoid overcrowded plots, spectral band positions were identified by determining minima in second derivative spectra. When band shifts were occurring in a data set, all shifted band positions were used for each spectral band. It is important to note though, that CPCA models were performed on all variables, not only the selected band positions. Thus, all score plots are based on the whole spectral region from 1900 to 700 cm^{-1}. All pre-processing methods and data analysis were performed using in-house developed routines written in MATLAB V. 8.3.0.532 (The MathWorks, Natick, USA).

Results and Discussion

The sample set covers conifer taxa with characteristic morphological and biochemical properties (Table 1). In general, pollen grains have double layered grain walls, with an inner layer (intine) that is predominantly composed of pectin and cellulose, and an outer layer (exine) that is a complex biopolymer sporopollenin [26]. The grains of Pinaceae and Podocarpaceae families have a distinctive hollow projection called the saccus from the central body of pollen grain called the corpus. While the grain interior has a complex composition of proteins, lipids, carbohydrates and nucleic acids, the saccus region is composed mainly of the exine grain wall. All the pollen grains we measured were bisaccate pollen since they have two sacci. In Fig 1 it can be seen that these two sacci are only visible when the orientation of the pollen grain allows it. The grain sizes of the measured taxa range from 50 μm for the species of *Podocarpus*, 55–75 μm for *Pinus*, 75 μm for *Cedrus*, 85 μm for *Picea*, to 120 μm for *Abies*. The sample set has a well defined phylogenetic relationship of the species, with one clear outlier: The species

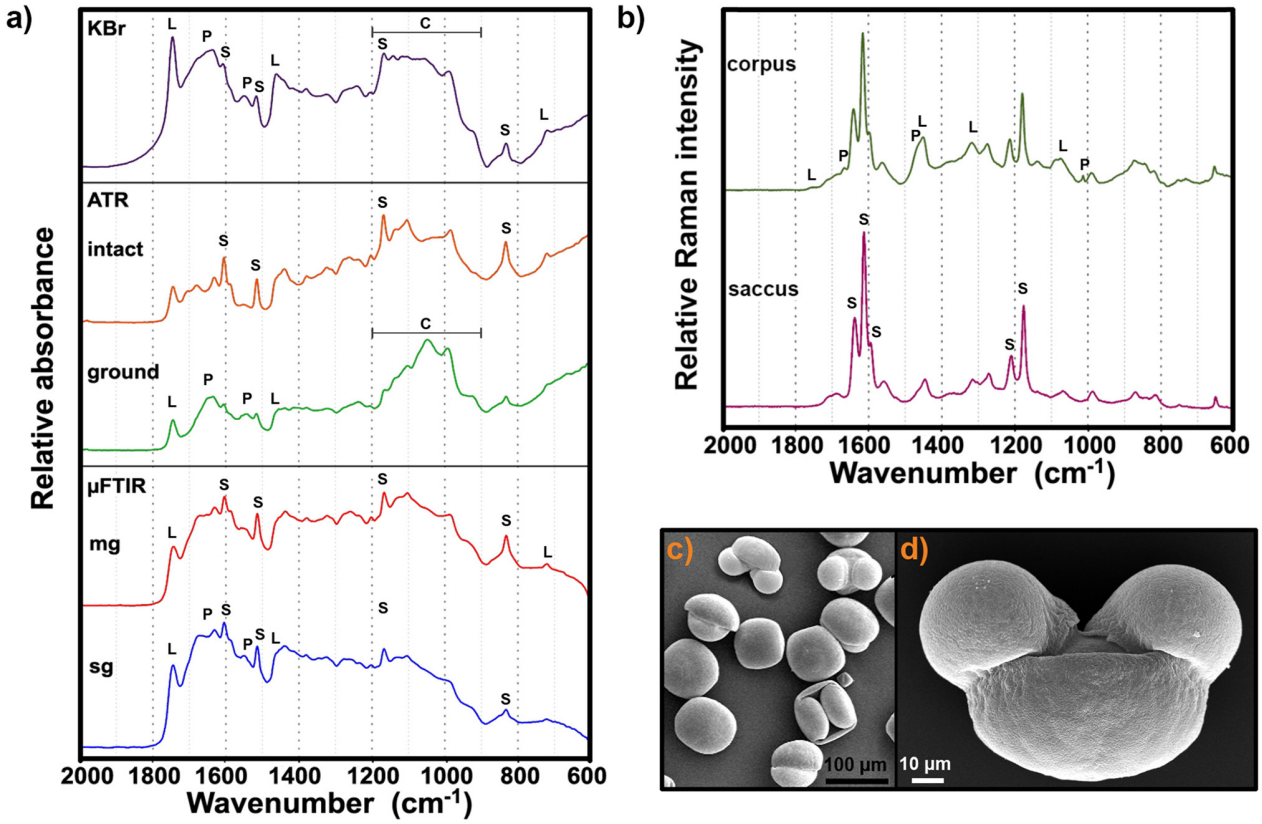


Fig 1. Measurement of *Abies cephalonica* pollen. (a) FTIR spectra obtained from different sampling techniques (from top downwards): transmission IR of KBr pellet, ATR-IR of intact and ground grains, IR microspectroscopy of multigrain (mg) and single grain (sg). **(b)** Raman spectra obtained from two different regions of a single grain: corpus and saccus regions. The marked signals are associated with the vibrational bands of (P) proteins, (L) lipids, (C) carbohydrates and (S) sporopollenins. **(c)** SEM image of pollen grains in various orientations: equatorial view (up left), distal polar view (up right) and proximal polar view (down left). **(d)** SEM image of pollen grain in equatorial view, with saccus (up, two hemispherical substructures) and corpus regions (down, large hemispherical substructure).

doi:10.1371/journal.pone.0137899.g001

Podocarpus neriifolius belongs to the family Podocarpaceae while the rest of the samples belong to the family Pinaceae. Within the Pinaceae, the set contains one large subset belonging to the clade of sister genera *Pinus* comprising 8 species and *Picea* comprising 3 species [27]. In addition, it contains one smaller subset, with a clade that is far related to the genera *Pinus* and *Picea*, belonging to the clade of the sister genera *Abies* comprising 2 species and *Cedrus* comprising 1 species.

Infrared measurements

A set of representative FTIR spectra, belonging to the same *Abies cephalonica* pollen sample and obtained by different methodologies, are shown in Fig 1. The vibrational spectra of five pollen species are shown in Fig 2. Regarding the amount of the measured pollen samples, the different vibrational techniques comprise bulk (KBr and ATR) and microscopic measurements (μFTIR and μRaman). Regarding the preparation of the pollen samples, the measurements include both ground (KBr and ATR ground) and intact samples (ATR intact, μFTIR and μRaman). Moreover, transmission μFTIR of single grain and multigrain measurements were performed. Finally, different substructures of pollen grains, namely saccus and corpus, were measured by μRaman.

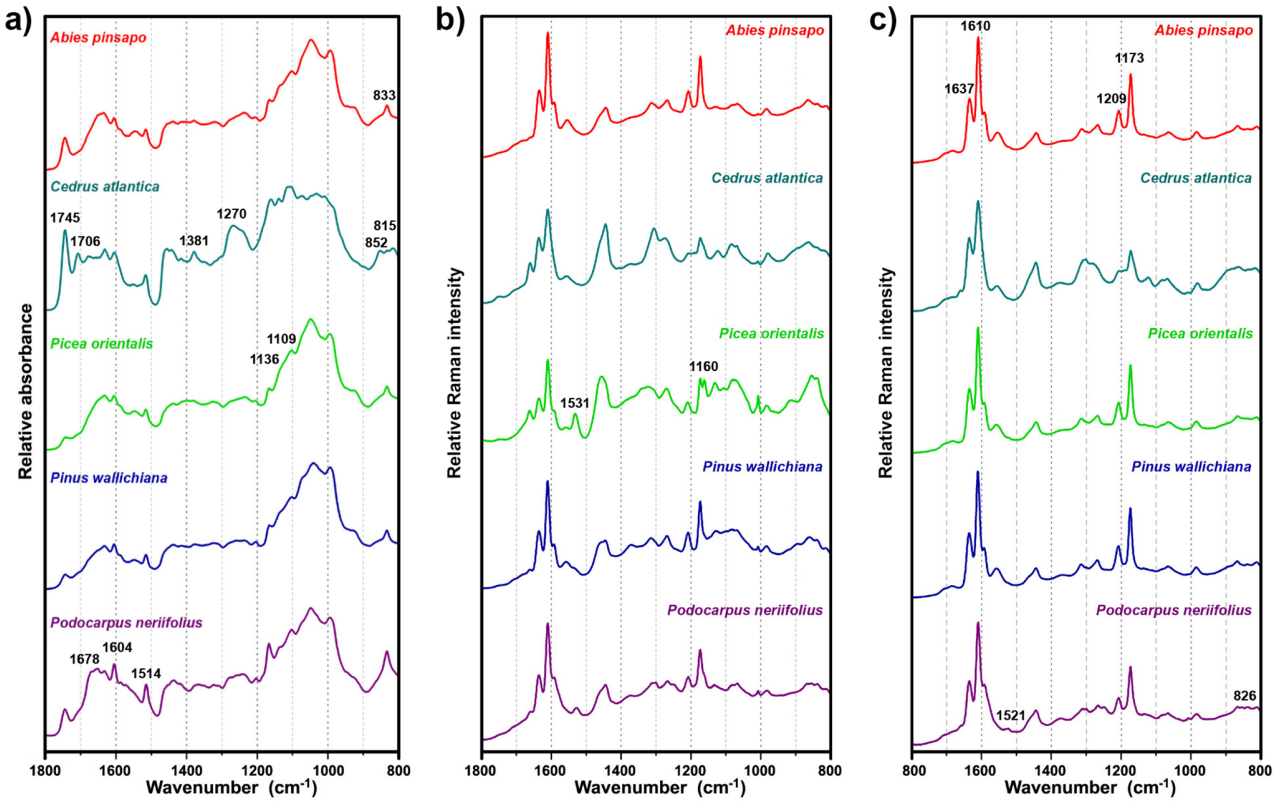


Fig 2. Spectra of representative samples of pollen measured as: (a) ATR of ground pollen, (b) Raman of corpus region, and (c) Raman of saccus region. The spectral set consists of EMSC normalized average spectra; for better viewing the spectra are offset.

doi:10.1371/journal.pone.0137899.g002

The FTIR spectra of pollen grains contain information on the major chemical constituents of pollen, such as lipids, proteins, carbohydrates, and grain wall biopolymers sporopollenins and cellulose [4, 5, 8]. The reference infrared spectra of pure compounds that can be typically found in pollen are shown in Fig 3. Lipids are characterized by the strong vibrational band at 1745 cm^{-1} (C = O stretch), as well as by a weaker band at 1462 cm^{-1} (CH_2 deformation), while proteins are characterized by two strong and broad bands at 1640 cm^{-1} (amide I: C = O stretch) and 1535 cm^{-1} (amide II: NH deformation and C–N stretch). Carbohydrates have strong absorption in the 1200–900 cm^{-1} region (C–O–C and C–OH stretch) including some characteristic bands for certain types of carbohydrates, such as cellulose (at 1107, 1055 and 1028 cm^{-1}) and amylose (at 1076 and 995 cm^{-1}). Sporopollenins are an extremely resistant group of biopolymers present in wall of pollen and spores. They are complex dehydrogenation-type biopolymers based on phenylpropanoid acids [26]. Therefore, their vibrational spectra show distinctive bands associated with the vibrations of aromatic rings at 1605, 1515, 1171, 853, 833 and 816 cm^{-1} [8]. These bands are present in the spectra of phenylpropanoid acids (Fig 3).

Although all FTIR spectra shown in Fig 1 contain information on the same chemical constituents, it can be seen that the ratio of chemical signals varies substantially from spectrum to spectrum. For instance, there is a considerable difference between the ATR spectra of ground and intact pollen. While the spectrum of intact pollen has predominant signals of sporopollenins, the spectrum of ground grains has predominant signals of lipids and proteins. In contrast, the spectra of KBr pellets and ATR ground samples contain similar information irrespective of the measurement technique. Both spectra are associated with the chemical components from

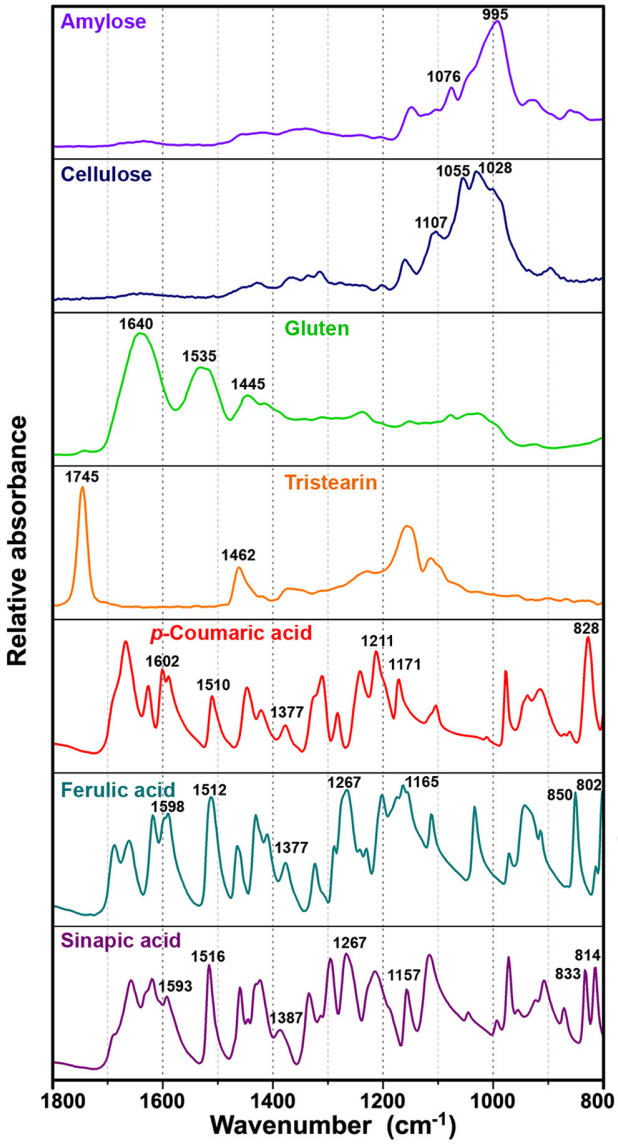


Fig 3. FTIR spectra of biochemicals. carbohydrates (amylose and cellulose), protein (gluten), lipid (tristearin) and phenylpropanoids (*p*-coumaric, ferulic and sinapic acids); for better viewing the spectra are offset.

doi:10.1371/journal.pone.0137899.g003

the grain interior (lipids and proteins), rather than grain wall materials (sporopollenins). As opposed to transmission measurement where infrared light interacts with the whole sample, the infrared light measured in reflection measurements penetrates the pollen grains only up to 0.5–5 μm in depth, depending on the wavelength [28]. It means that while in transmission measurements the whole pollen grain is measured, predominantly the grain wall is measured in reflection measurements [7, 8]. Therefore, for large pollen grains, such as Pinaceae and Podocarpaceae covered in this study, a more complete biochemical fingerprint is obtained by grinding the sample.

Regarding FTIR microspectroscopy measurements, it can be seen that both settings (multi-grain and single grain) obtain high quality spectra. The measured pollens have quite large sizes,

which enables measurement of single grain spectra with high signal-to-noise ratio. For pollen grain that are smaller than 50 μm, the quality of the obtained spectra is often poor due to strong Mie-type scattering effects [16, 29].

The main difference between the obtained multigrain and single grain spectra is their reproducibility. While the reproducibility of multigrain spectra is generally high, the reproducibility of single grain spectra is rather poor. Single grain spectra show high variation in ratio of principal bands belonging to grain interior and grain wall [29]. The reason for this is spatial orientation of a measured grain, which can either favour signals of saccus compounds such as sporopollenins or signals from corpus compounds such as nutrients (see grain orientations in Fig 1c). For instance, a measurement of the grain with equatorial profile orientation results in relatively stronger signals of proteins than sporopollenins, while the reverse is true for the grain with distal polar orientation [29].

It should be noted that spectral differences between pollen samples are related not only to phylogenetic differences between pollen species but to environmental conditions affecting parent plants as well. Pollen samples obtained from the same trees in different years can show noticeable inter-annual variations of biochemical composition. For instance, pollen samples of the specific *Abies cephalonica* tree, sampled in this study, were collected in consecutive pollination seasons, in 2011, 2012 and 2013. The infrared spectra of samples collected in different seasons show noticeable spectral differences associated with ratios of principal chemical constituents. The ATR (intact) FTIR spectra for *Abies cephalonica* in Fig 1a (pollination season 2012) can be compared with the ATR spectra in [30] (pollination season 2011), while the results of inter-annual variations of pollen data for this and other conifer trees was presented in [7]. Therefore, identification of pollen species based on their vibrational spectra is a complex issue that needs to take into consideration variability within species (i.e. variability within population and between different populations), as well as temporal variability within individual mother plant (i.e. variability between different pollination seasons).

Raman measurements

A set of representative Raman spectra obtained from samples belonging to the same species *Abies cephalonica* are shown in Fig 1. In order to investigate the spectral reproducibility and the chemical heterogeneity of single pollen grains, Raman spectra were collected from different positions on each single pollen grain. As shown in Fig 1, there are distinct differences between spectra obtained from the saccus and corpus structures. The presence of additional signals in the spectrum of the corpus structure is obvious. Spectra obtained from Raman measurements of saccus and corpus region of pollen species are shown in the Supplementary information (Fig J in S1 File). Similar to FTIR spectra of pollen, Raman spectra contain information on the major chemical constituents [6, 8, 31]. The main spectral features, that are dominating both saccus and corpus Raman spectra, are signals at 1637, 1610, 1590, 1209 and 1173 cm^{-1}, which can be associated with sporopollenins, that is with phenylpropanoid building blocks. The additional signals, present in the corpus spectrum, can be associated with lipids (1750, 1444, 1304 and 1065 cm^{-1}) and proteins (1660, 1455, 1007 cm^{-1}). Reference Raman spectra of typical pure constituents of pollen samples are shown in the Fig 4. The results are in agreement with pollen grain anatomy, since the saccus regions are predominantly made of sporopollenins while the corpus region contains nutrients and coding chemicals.

Since pollen grains have a layered and chemically inhomogeneous structure, the acquired spectra differ when the depth of the laser focus is changed. For saccate pollen, the inhomogeneity is not only due to a layered structure, but also due to the fact that the grain has a saccus and a corpus region which represent two completely different substructures. Since the saccus region

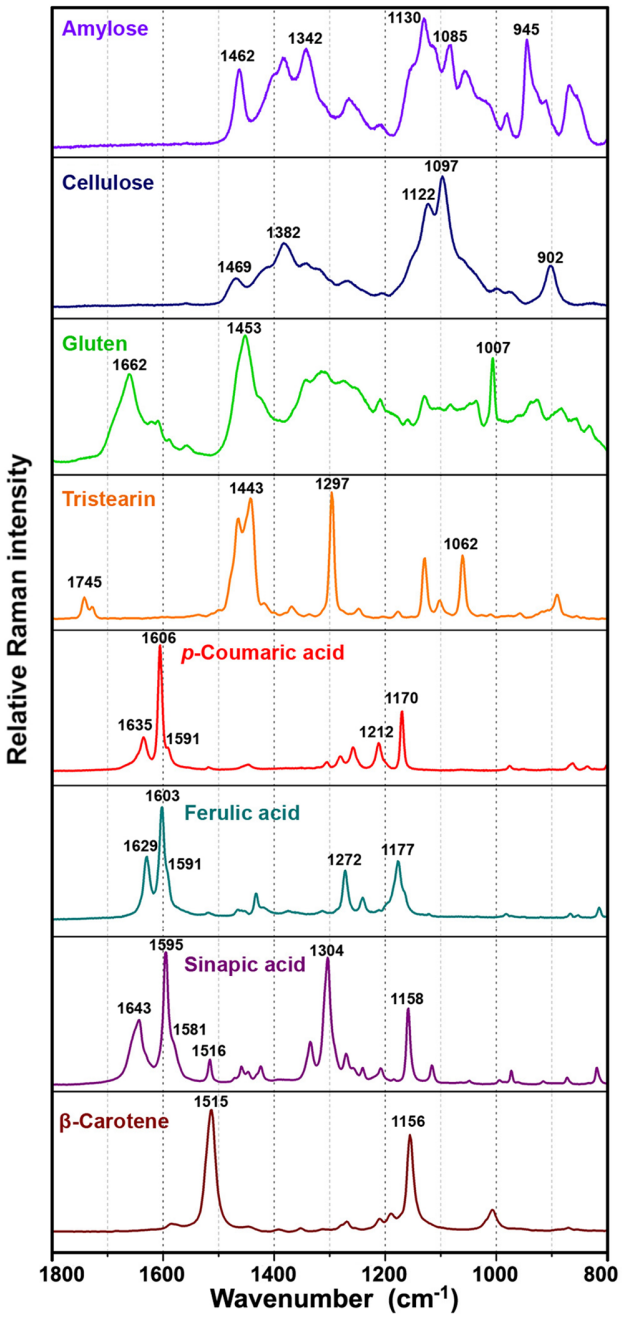


Fig 4. Raman spectra of biochemicals. carbohydrates (amylose and cellulose), protein (gluten), lipid (tristearin), phenylpropanoids (p-coumaric, ferulic and sinapic acids) and carotenoid (β-carotene); for better viewing the spectra are offset.

doi:10.1371/journal.pone.0137899.g004

is a pretty homogeneous region, Raman measurements of saccus substructures result in a visually pure spectrum of sporopollenins. All principal bands can be associated with phenylpropanoid building blocks of sporopollenins, as indicated by the similarity between the Raman spectrum of saccus region (Fig 1b) and the spectra of phenylpropanoid acids (Fig 4). Raman measurements of the corpus region may result in varying signals of lipids and proteins

strongly depending on the focus depth of the laser. Therefore, in contrast to findings of a previous Raman study [32], it is challenging to obtain reproducible Raman spectra of single pollen grains.

Consensus principal component analysis (CPCA)

The spectral data sets were pre-processed prior to CPCA in order to remove baseline variations and differences in the effective optical path length and in order to enhance the spectral signal [33]. These unwanted variations arise due to non-ideal instrument and sample properties, such as fluorescence backgrounds in Raman and variability between KBr sample pellets in FTIR due to slight differences in sample concentration. The spectra were smoothed and transformed to second derivative form by the SG algorithm followed by EMSC pre-processing. Second derivative spectra belonging to the different vibrational methodologies are given in the Supplementary information (Figs A–G in S1 File). In order to estimate variability within sample replicates, spectra from *Pinus* and *Picea* genera were used (Table A and Table B in S1 File). The variability within replicates shows that the highest reproducibility was obtained by infrared measurements on bulk samples (ATI), as well as by Raman microspectroscopy of measurements of corpus region of pollen grain (RMC). Moreover, the variability between KBR sample pellets (KBR), as well as between differently orientated single pollen grains (SGR), results with relatively low reproducibility.

The seven spectral data blocks were analysed by multiblock analysis (CPCA) in order to assess and compare the power of different vibrational methodologies for the analysis of different chemical aspects of pollen. The strength of a multiblock analysis compared to principal component analysis of each block separately is, that analytical aspects that are common and particular for the different spectroscopic methods can be revealed more straight-forward in CPCA. Further, spectral bands that are important for observed sample variation patterns in the different blocks can be visualized in a common multiblock correlation loading block. By CPCA so-called global and block scores are obtained. The global scores of the CPCA are representing the consensus variation between all vibrational methods, while the block scores are representing the consensus variation within each single block and are thus reflecting block-specific variation patterns (Fig 5). The fact that CPCA finds a common variation pattern for all blocks in each component, constitutes the major difference to a PCA analysis of each single block, separately: In PCA of separate blocks, the results obtained for each component are not comparable among the blocks. Therefore, by assessing global and block scores in CPCA, direct information about similarities and differences between the different vibrational methods can be obtained. The high fraction of explained variance for a single block indicates that the global pattern of the CPCA is providing good explanation for the variation within that block, and vice versa. In correlation loading plots the correlations between global scores (principal components) and vibrational bands are plotted [19, 23] (Fig 6). This allows the visualization of correlations between vibrational bands and principal components for all methods in one picture (blocks). Further, by including information about pollen genera in the correlation loading plots, correlations between vibrational bands and genus information can be studied (Fig 6). This is achieved by defining a matrix of indicator variables with zeros and ones that specifies if a sample belongs to a genus or not [23]. Variables that are well explained by respective components in CPCA are located within the outer ring, where the outer circle indicated complete correlation and the inner circle indicates 50% correlation. In addition to Figs 5 and 6, more detailed results are presented in the Supplementary Information (Figs A–G in S1 File). 3D presentations of global score plot are given in the SFig H in S1 File, alongside PCA score plots of individuals blocks (Fig I in S1 File). The score plots, obtained by a separate PCA of each individual block are

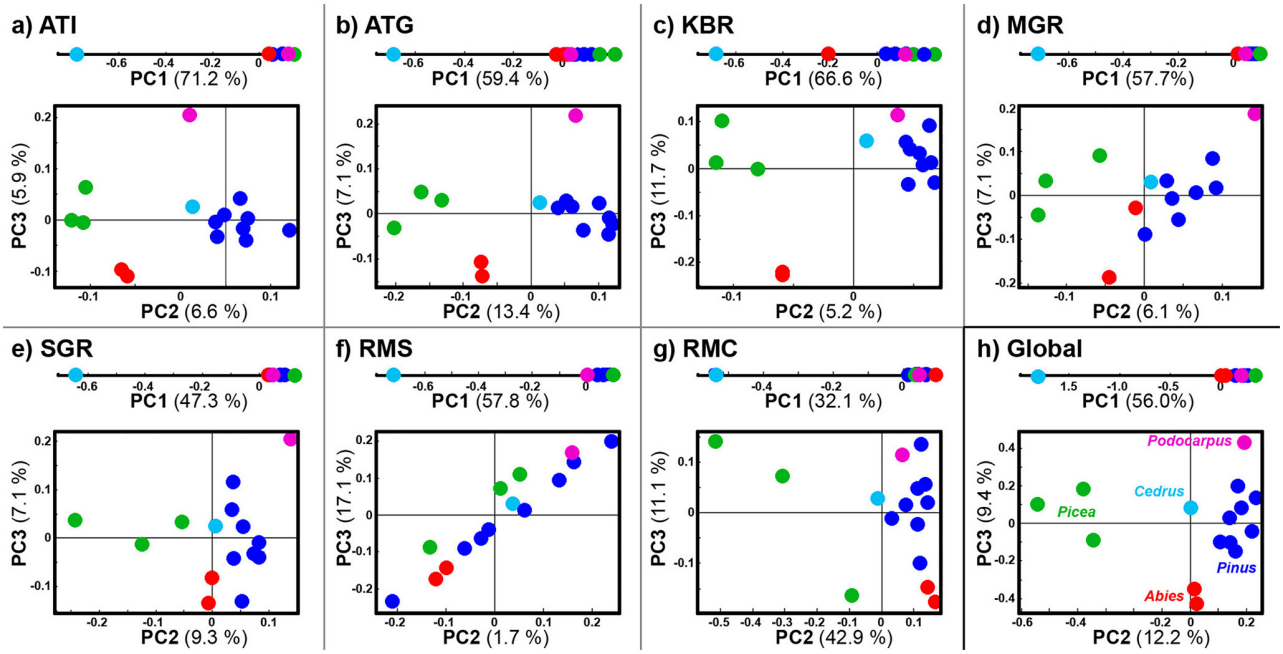


Fig 5. Score plots of individual blocks and global scores of consensus principal component analysis (CPCA). ATR-FTIR of **(a)** (ATI) intact pollen, **(b)** (ATG) ground pollen; **(c)** (KBR) Transmission FTIR of KBr pellets; Transmission IR microspectroscopy of **(d)** (MGR) multigrain and **(e)** (SGR) single grain; Raman spectroscopy of **(f)** (RMS) saccus region and **(g)** (RMC) corpus region of single pollen grain; **(h)** Global scores. Samples are labelled in accordance to pollen genus: Abies (red), Picea (green), Pinus (blue), Podocarpus (magenta), and Cedrus (cyan). The percent variances for the PCs are given in supplementary part.

doi:10.1371/journal.pone.0137899.g005

consistent with the CPCA results, and thus will not be discussed further in this study. Spectra of the same samples and the PCA score plots of individual blocks (with designation at species level), show good reproducibility of spectral measurements (Fig J and Fig K in S1 File).

The consensus variation pattern that was obtained in the first three components of the global score plot reflects the differences between the five different pollen genera as shown in Fig 5h. While PC1 is clearly separating *Cedrus* from the other genera, PC2 is separating *Picea* and *Pinus* and PC3 is separating *Abies* and *Podocarpus*. Therefore, the first three PCs are sufficient for separating the five Pinales genera. Similar grouping patterns as revealed by the global scores (Fig 5h) are found in the block score plots of the ATR block of intact pollen samples (Fig 5a), of the ATR block of ground pollen samples (Fig 5b), and in the block score plot of the KBr pellet block (Fig 5c). The microspectroscopy measurements, both Raman and FTIR, reveal different patterns (Fig 5d–5g). While the first principal components are as before separating *Cedrus* from the other genera, PC2 and PC3 scores are not providing a clear separation of pollen genera. Especially the Raman spectra obtained from the saccus region do not show any grouping with respect to genera, which is also reflected in the low explained variances in PC2 (Fig 5f).

Biochemical composition of pollen

The correlation loadings for the CPCA model are shown in Fig 6, and additional information with more detailed correlation loading plots and spectral data are presented in the Supplementary Information (Figs A–G in S1 File). Infrared bands associated with sporopollenins, lipids and proteins are highly positively or negatively correlated to *Cedrus* for PC1. More specific: In the ATG spectral block, strong positive correlation to *Cedrus* are observed for bands at 1381,

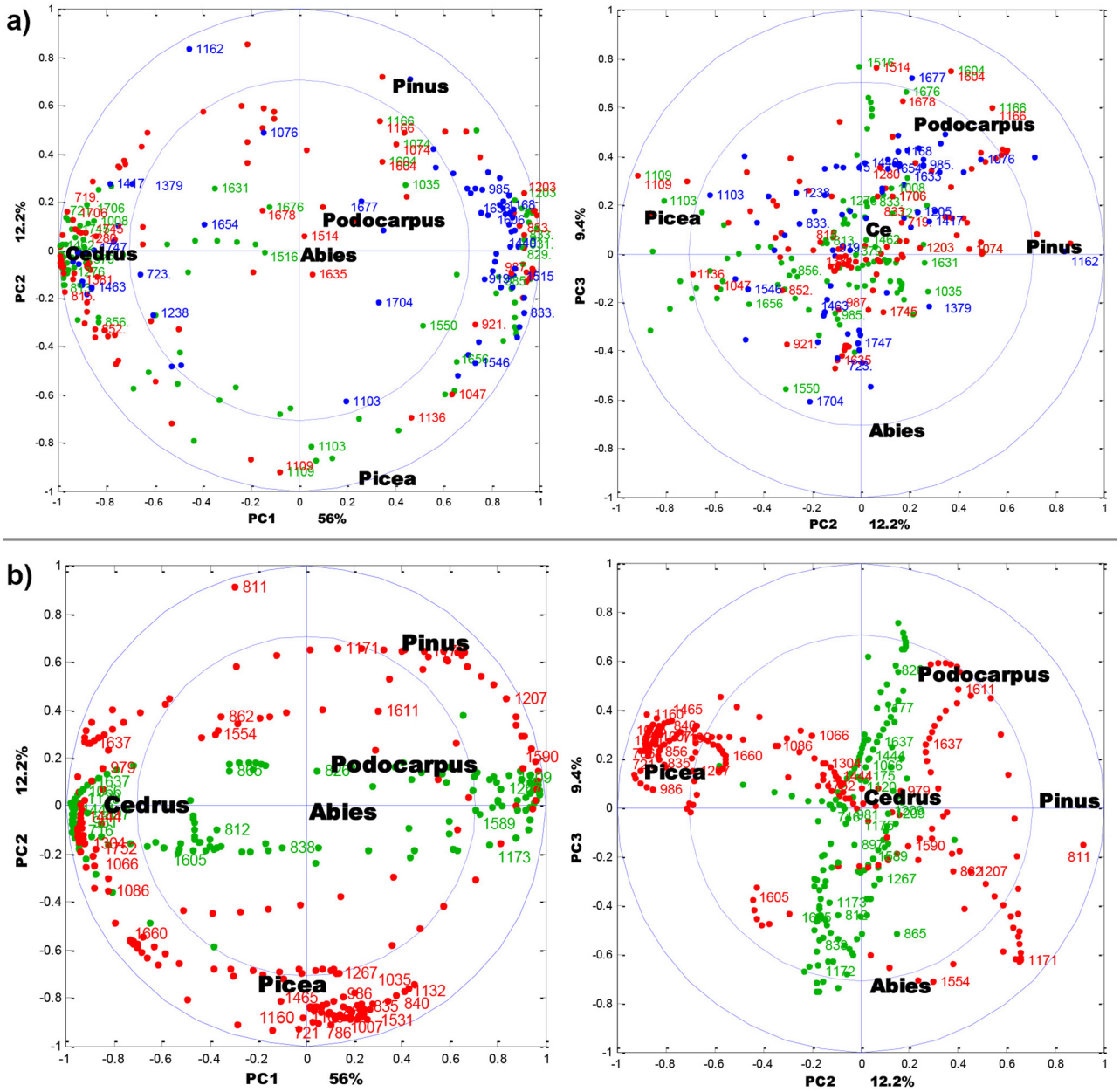


Fig 6. CPCA correlation loading plots for the first three principal components. (a) Correlation between pollen genera (black), FTIR of KBr pellets (blue), ATR of ground pollen (red) and ATR of intact pollen (green). **(b)** Correlation between pollen genera (black), Raman of corpus region (red) and Raman of saccus region (green). For the sake of clarity, only the selected variables are presented. The percent variances for the first five PCs are 56.0, 12.2, 9.4, 7.5 and 5.1.

doi:10.1371/journal.pone.0137899.g006

1270, 852 and 815 cm^{-1} referring to sporopollenin and at 1745, 1706 and 720 cm^{-1} referring to lipids. In the ATI spectral block, strong positive correlation to *Cedrus* are observed at bands at 1516, 1270, 856 and 813 cm^{-1} referring to sporopollenin, at 1745, 1462, and 721 cm^{-1} referring to lipids and at 1631 and 1550 cm^{-1} referring to proteins. In the KBR spectral block strong positive correlation to *Cedrus* are observed at bands at 1747, 1463, 1238 and 723 cm^{-1} (lipids) (Fig 6a). In addition, *Cedrus* has a strong negative correlation to the band at 833 cm^{-1} (sporopollenin)

in the ATG block, to the bands 1604 and 833 cm^{-1} (sporopollenin) in the ATI block, and to the bands at 1606, 1515 and 833 cm^{-1} (sporopollenin) and to the bands at 1633 and 1546 cm^{-1} (proteins) in the KBR block (Fig 6a). These results show that there are clear and interpretable chemical differences in the biochemical composition between the *Cedrus* pollen and the rest of Pinales pollens.

The relative intensities of lipid and protein bands indicate that *Cedrus atlantica* has significantly higher relative amount of lipids than average Pinales pollen. Even more interesting is the positive and negative correlation between *C. atlantica* and a number of sporopollenins bands. Although the exact chemical composition of sporopollenins has yet to be determined, it is known that that these complex biopolymers are composed of phenylpropanoid building blocks [26]. The phenylpropanoid building blocks, primarily derivatives of *p*-coumaric and ferulic acids, have characteristic vibrations, and thus can be determined by FTIR and Raman spectroscopy [34]., The vibrational spectra of phenylpropanoids are shown in Figs 3 and 4. The relative intensities of Raman and infrared bands indicate that *C. atlantica* pollen have a higher ratio of ferulic-to-*p*-coumaric acid derivatives in sporopollenin compared to other Pinales pollens. Specifically, this is revealed by the positive correlation between *C. atlantica* and the infrared bands of ferulic acid at 1267 cm^{-1} (alkyl rocking vibration), and at 852 and 815 cm^{-1} (ring vibrations), as well as by negative correlation with the bands of *p*-coumaric acid at 833 cm^{-1} (ring vibrations) present in the ATG block (Figs 6a and 2a). The Raman spectrum of *C. atlantica* pollen corroborates this notion by a higher ratio of ring stretches doublets (ratio of peaks at 1637 and 1610 cm^{-1}), as well as lower ratios of ring bending peaks at 1209 and 1173 cm^{-1} (Fig 2c). The strong positive correlation between *C. atlantica* and the band at 1637 cm^{-1}, and strong negative correlation to the bands at 1209 and 1173 cm^{-1} can be seen in the RMS block (Fig 6b). The band at 1209 cm^{-1} is characteristic for *p*-coumaric acid, while ferulic acid has a higher ratio of ring stretches doublets than *p*-coumaric acid. Therefore the Raman spectra are fully supporting the notion obtained by FTIR.

Sporopollenins have variable composition which is dependent on species specific pathways for the biosynthesis of sporopollenin from simple phenylpropanoids, such as cinnamic acid. Moreover, phenylpropanoids play important role in mitigating damage caused by UV-B radiation, and a variety of plant species are responding to increased UV-B exposure by increasing the concentrations of these compounds [35]. Since the concentration of phenylpropanoids is correlated with UV-B radiation, and since sporopollenins are extremely resilient to deterioration, quantitative measurement of these compounds in pollen microfossils can be used as UV-B proxy to track changes in the flux of UV-B radiation over geological time [15].

C. atlantica has survived the last glacial in multiple refuge areas along the Atlas Mountains [36]. Following the glacial, it has spread rapidly along the major mountain chains of northern Algeria and Morocco. *C. atlantica* is unique among other measured species by pollination season. *Cedrus* species shed pollen in autumn, a considerable difference to other Pinaceae that pollinate during spring. Pollen germination and rapid tube growth demand excessive energy reserves, particularly during pollination in colder conditions that is characteristic for *C. atlantica*. Moreover, the natural habitat of *C. atlantica* is at lower latitudes and higher altitudes than habitats of other measured Pinaceae species, both of which is connected with high exposure to UV-B radiation. Therefore, it is highly plausible that pollen traits of *C. atlantica*, in the form of high nutrient content and specific sporopollenins composition, are playing an important role in adaptation to extreme mountainous conditions of south Mediterranean.

The separation between *Pinus* and *Picea* genera, present in PC2, is predominantly based on different carbohydrate composition. This is indicated by the strong positive correlation between *Picea* and carbohydrate bands at 1136 and 1109 cm^{-1} in the ATG block, and at 1132 and 1035 cm^{-1} in the RMC block (Fig 6). Moreover, *Picea* genera have strong positive

correlation with bands specific for carotenoids at 1531 and 1160 cm^{-1} in the RMC block (Fig 6b). The two bands are associated with C-C single bond (1160 cm^{-1}) and double bond stretch (1531 cm^{-1}) of the polyene chain (Fig 4) [37]. An important feature of Raman technique is that carotenoids can be measured in complex matrices, such as pollen, by obtaining their Raman spectra under resonant excitation [38]. Carotenoids have an allowed π-π^* electronic transition which occurs in the visible region and which gives rise to their strong colours. Resonance Raman spectrum of carotenoids is obtained when the wavelength of the incident laser coincides with electronic transition causing strong enhancement of vibrational bands, particularly those at 1531 and 1160 cm^{-1} that have strong electron-phonon coupling. Even though pollen carotenoids are present at too low concentration to be observed by FTIR, they can be analysed by Raman due to resonant effect.

Compared with other Pinaceae pollens, pollen of *Picea* has significantly higher concentration of carotenoids. Similar to phenylpropanoids, pollen carotenoids are playing important role in photo protection from environmental stress by quenching free radicals [39]. Pinaceae pollen are anemophilous, and therefore photo protection is of outmost importance for survival if airborne grains are carried to higher layers of atmosphere. Studies have shown that pollination distances in Pinaceae can be tens of kilometres, and possibly even greater [40].

The separation between *Podocarpus neriifolius* and *Abies* genera, present in PC3, is based on sporopollenin composition as well. This is indicated by the positive correlation between *P. neriifolius* and sporopollenin bands at 1678, 1604 and 1514 cm^{-1} in the ATG block, at 1677 cm^{-1} in the ATI and KBR blocks, at 1611 cm^{-1} in the RMC block, and at 1521 and 826 cm^{-1} in the RMS block (Fig 6). Analogous to *C. atlantica*, the results indicate that the chemical composition of sporopollenin in *P. neriifolius*, specifically the ratio of phenylpropanoid building blocks, is different from composition in Pinaceae. Unfortunately, the listed spectral differences are not specific and conclusive enough. However they suggest that *P. neriifolius* sporopollenin could contain some minor phenylpropanoid building blocks, in addition to *p*-coumaric and ferulic acids. For instance, presence of species-specific Raman bands at 1521 and 1300 cm^{-1} is consistent with presence of sinapic acid (Figs 4 and 6).

The explained variances of CPCA indicate the advantages of different spectroscopic methodologies. The ATI and the RMS blocks are contributing most to the first principal component, regarding FTIR and Raman respectively. These two methodologies are focused on obtaining information associated with pollen grain wall. Therefore, they are best suited for obtaining variations between taxa with different chemical composition of sporopollenins, such is the case between *Cedrus* and the rest of Pinaceae pollens. In the same way, the ATG and the RMC blocks are contributing most to the second principal component, regarding FTIR and Raman respectively. Both of the methodologies are optimized for gathering information on chemical composition of pollen grain interior. For that reason, they are obtaining variations in carbohydrate composition, such is the case between *Picea* and *Pinus* genera.

Conclusions

The study demonstrated that the vibrational methodologies have great potential for systematic collection of data on ecosystems since phylogenetic variation is well represented in the biochemical composition of pollen. The differences can be clearly connected to specific chemical composition, such as nutrient storage or grain wall biopolymers. For example, the extensive spectral differences between pollen of *Cedrus atlantica* and the rest of Pinaceae family were unambiguously connected with molecular composition, revealing the unique composition of *C. atlantica*'s sporopollenin.

Out of the seven tested methodologies, the best taxonomy based differentiation of the pollen was obtained by reflectance (ATR) and transmission (KBr pelleting) IR measurements on bulk samples, as well as by Raman microspectroscopy measurements that are focused on the corpus region of pollen grains. Raman microspectroscopy measurements indicate that measurement area, as well as the depth of focus, can have crucial influence on the obtained data. The FTIR microspectroscopy methods have underperformed, compared to other FTIR methodologies. The predominant reason for this is low reproducibility and insufficient specificity of FTIR microspectroscopy data, caused by variations in spatial orientation of a measured grains and by scattering effects [16, 29]. Therefore, further studies are needed to in order to develop better experimental setting for FTIR microspectroscopy of individual pollen grains.

Supporting Information

S1 File. Contains the following files. Fig A CPCA correlation loading plots with markings related to ATI. **Fig B** CPCA correlation loading plots with markings related to ATG. **Fig C** CPCA correlation loading plots with markings related to KBR. **Fig D** CPCA correlation loading plots with markings related to MGR. **Fig E** CPCA correlation loading plots with markings related to SGR. **Fig F** CPCA correlation loading plots with markings related to RMS. **Fig G** CPCA correlation loading plots with markings related to RMC. **Fig H** 3D presentation of global score plot. **Fig I** PCA score plots for individual blocks, genus level. **Fig J** Raman spectra of corpus and saccus region of species. **Fig K** PCA score plots for individual blocks, species level. **Fig L** PCA score plots for ATI block, species level. **Fig M** PCA score plots for KBR block, species level. **Fig N** PCA score plots for RMC block, species level. **Table A** Variability measurements within replicates of *Pinus* genus. **Table B** Variability measurements within replicates of *Picea* genus.
(PDF)

Acknowledgments

We thank Elin Ørmen for the help with the SEM measurements, and Bjørg Narum, Karen Wahlstrøm Sanden, Ulrike Böcker and Nils Kristian Afseth for helpful discussions.

Author Contributions

Conceived and designed the experiments: MB BZ AK. Performed the experiments: MB BZ. Analyzed the data: MB BZ AK. Contributed reagents/materials/analysis tools: BZ AK. Wrote the paper: MB BZ AK.

References

1. Sivaguru M, Mander L, Fried G, Punyasena SW. Capturing the Surface Texture and Shape of Pollen: A Comparison of Microscopy Techniques. Plos One. 2012; 7(6). ARTN e39129 doi: 10.1371/journal.pone.0039129 WOS:000305340000066. PMID: 22720050
2. Benton MJH, D. A.T. Introduction to Paleobiology and the Fossil Record. United Kingdom: Wiley-Blackwell; 2009. 592 p.
3. Dell'Anna R, Lazzeri P, Frisanco M, Monti F, Malvezzi Campeggi F, Gottardini E, et al. Pollen discrimination and classification by Fourier transform infrared (FT-IR) microspectroscopy and machine learning. Anal Bioanal Chem. 2009; 394(5):1443–52. doi: 10.1007/s00216-009-2794-9 PMID: 19396429
4. Gottardini E, Rossi S, Cristofolini F, Benedetti L. Use of Fourier transform infrared (FT-IR) spectroscopy as a tool for pollen identification. Aerobiologia. 2007; 23(3):211–9. doi: 10.1007/s10453-007-9065-z
5. Pappas CS, Tarantilis PA, Harizanis PC, Polissiou MG. New method for pollen identification by FT-IR spectroscopy. Appl Spectrosc. 2003; 57(1):23–7. WOS:000180902000005. PMID: 14610932

6. Schulte F, Lingott J, Panne U, Kneipp J. Chemical characterization and classification of pollen. Analytical chemistry. 2008; 80(24):9551–6. Epub 2008/11/04. doi: 10.1021/ac801791a PMID: 18975984.
7. Zimmermann B, Kohler A. Infrared Spectroscopy of Pollen Identifies Plant Species and Genus as Well as Environmental Conditions. Plos One. 2014; 9(4):e95417. doi: 10.1371/journal.pone.0095417 PMID: 24748390
8. Zimmermann B. Characterization of Pollen by Vibrational Spectroscopy. Appl Spectrosc. 2010; 64 (12):1364–73. doi: 10.1366/000370210793561664 PMID: 21144154
9. Boyain-Goitia AR, Beddows DCS, Griffiths BC, Telle HH. Single-pollen analysis by laser-induced breakdown spectroscopy and Raman microscopy. Appl Optics. 2003; 42(30):6119–32. WOS:000186000200026.
10. Glauser AL, Harper CJ, Taylor TN, Taylor EL, Marshall CP, Marshall AO. Reexamination of cell contents in Pennsylvanian spores and pollen grains using Raman spectroscopy. Rev Palaeobot Palyno. 2014; 210:62–8. WOS:000343847100005.
11. Mularczyk-Oliwa M, Bombalska A, Kaliszewski M, Wlodarski M, Kopczynski K, Kwasny M, et al. Comparison of fluorescence spectroscopy and FTIR in differentiation of plant pollens. Spectrochim Acta A. 2012; 97:246–54. WOS:000310395800034.
12. Kano H, Hamaguchi HO. Vibrational imaging of a single pollen grain by ultrabroadband multiplex coherent anti-stokes Raman scattering microspectroscopy. Chem Lett. 2006; 35(10):1124–5. WOS:000242708800020.
13. Fraser WT, Scott AC, Forbes AE, Glasspool IJ, Plotnick RE, Kenig F, et al. Evolutionary stasis of sporopollenin biochemistry revealed by unaltered Pennsylvanian spores. The New phytologist. 2012; 196 (2):397–401. doi: 10.1111/j.1469-8137.2012.04301.x PMID: 22913758.
14. Willis KJ, Feurdean A, Birks HJB, Bjune AE, Breman E, Broekman R, et al. Quantification of UV-B flux through time using UV-B-absorbing compounds contained in fossil Pinus sporopollenin. New Phytologist. 2011; 192(2):553–60. WOS:000295282700027. doi: 10.1111/j.1469-8137.2011.03815.x PMID: 21810096
15. Lomax BH, Fraser WT, Harrington G, Blackmore S, Sephton MA, Harris NBW. A novel palaeoaltimetry proxy based on spore and pollen wall chemistry. Earth Planet Sc Lett. 2012; 353:22–8. doi: 10.1016/j.epsl.2012.07.039 WOS:000311014300003.
16. Lukacs R, Blumel R, Zimmerman B, Bagcioglu M, Kohler A. Recovery of absorbance spectra of micrometer-sized biological and inanimate particles. Analyst. 2015; 140(9):3273–84. WOS:000353154400044. doi: 10.1039/c5an00401b PMID: 25797528
17. Zimmermann B, Kohler A. Optimizing Savitzky-Golay Parameters for Improving Spectral Resolution and Quantification in Infrared Spectroscopy. Appl Spectrosc. 2013; 67(8):892–902. WOS:000322559700010. doi: 10.1366/12-06723 PMID: 23876728
18. Kohler A, Kirschner C, Oust A, Martens H. Extended Multiplicative Signal Correction as a Tool for Separation and Characterization of Physical and Chemical Information in Fourier Transform Infrared Microscopy Images of Cryo-sections of Beef Loin. Appl Spectrosc. 2005; 59(6):707–16. doi: 10.1366/0003702054280649 PMID: 16053536
19. Hassani S, Martens H, Qannari EM, Hanafi M, Borge GI, Kohler A. Analysis of-omics data: Graphical interpretation- and validation tools in multi-block methods. Chemometr Intell Lab. 2010; 104(1):140–53. doi: 10.1016/j.chemolab.2010.08.008 WOS:000284658300015.
20. Bocker U, Ofstad R, Wu ZY, Bertram HC, Sockalingum GD, Manfait M, et al. Revealing covariance structures in Fourier transform infrared and Raman microspectroscopy spectra: A study on pork muscle fiber tissue subjected to different processing parameters. Appl Spectrosc. 2007; 61(10):1032–9. WOS:000250162400003. PMID: 17958951
21. Perisic N, Afseth NK, Ofstad R, Hassani S, Kohler A. Characterising protein, salt and water interactions with combined vibrational spectroscopic techniques. Food Chem. 2013; 138(1):679–86. WOS:000314193800095. doi: 10.1016/j.foodchem.2012.10.117 PMID: 23265540
22. Westerhuis JA, Kourti T, MacGregor JF. Analysis of multiblock and hierarchical PCA and PLS models. J Chemometr. 1998; 12(5):301–21. WOS:000076411200001.
23. Kohler A, Hanafi M, Bertrand D, Qannari EM, Janbu AO, Møretrø T, et al. Interpreting Several Types of Measurements in Bioscience. Biomedical Vibrational Spectroscopy: John Wiley & Sons, Inc.; 2007. p. 333–56.
24. Karaman İ, Nørskov NP, Yde CC, Hedemann MS, Bach Knudsen KE, Kohler A. Sparse multi-block PLSR for biomarker discovery when integrating data from LC–MS and NMR metabolomics. Metabolomics. 2014; 11(2):367–79. doi: 10.1007/s11306-014-0698-y
25. Martens H, Næs T. Multivariate calibration. Chichester England; New York: Wiley; 1989. xvii, 419 p. p.

26. Blokker P, Boelen P, Broekman R, Rozema J. The Occurrence of p-coumaric Acid and Ferulic Acid in Fossil Plant Materials and their Use as UV-proxy. Plant Ecol. 2006; 182(1–2):197–207. doi: 10.1007/s11258-005-9026-y

27. Lin CP, Huang JP, Wu CS, Hsu CY, Chaw SM. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. Genome biology and evolution. 2010; 2:504–17. doi: 10.1093/gbe/evq036 PMID: 20651328; PubMed Central PMCID: PMC2997556.

28. Mirabella FM. Practical Spectroscopy Series: Internal ReflectionSpectroscopy: Theory and Applications. In: Dekker M, editor. Practical Spectroscopy. New York: CRC Press; 1993. p. 17–52.

29. Zimmermann B, Bağcıoğlu M, Sandt C, Kohler A. Vibrational microspectroscopy enables chemical characterization of single pollen grains as well as comparative analysis of plant species based on pollen ultrastructure. Planta. 2015:1–14. doi: 10.1007/s00425-015-2380-7

30. Zimmermann B, Tkalčec Z, Mešić A, Kohler A. Characterizing Aeroallergens by Infrared Spectroscopy of Fungal Spores and Pollen. Plos One. 2015; 10(4):e0124240. doi: 10.1371/journal.pone.0124240 PMID: 25867755

31. Ivleva NP, Niessner R, Panne U. Characterization and discrimination of pollen by Raman microscopy. Anal Bioanal Chem. 2005; 381(1):261–7. doi: 10.1007/s00216-004-2942-1 PMID: 15605236.

32. Pummer BG, Bauer H, Bernardi J, Chazallon B, Facq S, Lendl B, et al. Chemistry and morphology of dried-up pollen suspension residues. J Raman Spectrosc. 2013; 44(12):1654–8. doi: 10.1002/Jrs.4395 WOS:000328210500003.

33. Afseth NK, Kohler A. Extended multiplicative signal correction in vibrational spectroscopy, a tutorial. Chemometr Intell Lab. 2012; 117:92–9. doi: 10.1016/j.chemolab.2012.03.004 WOS:000311010900012.

34. Zuk M, Dyminska L, Kulma A, Boba A, Prescha A, Szopa J, et al. IR and Raman studies of oil and seedcake extracts from natural and genetically modified flax seeds. Spectrochim Acta A. 2011; 78(3):1080–9. doi: 10.1016/j.saa.2010.12.054 WOS:000288046600023.

35. Newsham KK, Robinson SA. Responses of plants in polar regions to UVB exposure: a meta-analysis. Global Change Biol. 2009; 15(11):2574–89. doi: 10.1111/j.1365-2486.2009.01944.x WOS:000270662000002.

36. Terrab A, Hampe A, Lepais O, Talavera S, Vela E, Stuessy TF. Phylogeography of North African Atlas Cedar (Cedrus Atlantica, Pinaceae): Combined Molecular and Fossil Data Reveal a Complex Quaternary History. Am J Bot. 2008; 95(10):1262–9. doi: 10.3732/Ajb.0800010 WOS:000260234900008. PMID: 21632331

37. de Oliveira VE, Castro HV, Edwards HGM, de Oliveira LFC. Carotenes and carotenoids in natural biological samples: a Raman spectroscopic analysis. J Raman Spectrosc. 2009; 41(6):642–50. doi: 10.1002/jrs.2493

38. Merlin JC. Resonance Raman-Spectroscopy of Carotenoids and Carotenoid-Containing Systems. Pure Appl Chem. 1985; 57(5):785–92. doi: 10.1351/pac198557050785 WOS:A1985AGR5200022.

39. Garcia-Asua G, Lang HP, Cogdell RJ, Hunter CN. Carotenoid diversity: a modular role for the phytoene desaturase step. Trends Plant Sci. 1998; 3(11):445–9. doi: 10.1016/S1360-1385(98)01329-6 WOS:000077252100011.

40. Robledo-Arnuncio JJ, Gil L. Patterns of pollen dispersal in a small population of Pinus sylvestris L. revealed by total-exclusion paternity analysis. Heredity. 2005; 94(1):13–22. doi: 10.1038/sj.hdy.6800542 WOS:000225772400003. PMID: 15292910

Supporting Information

A multiscale vibrational spectroscopic approach for identification and biochemical characterization of pollen

Murat Bağcıoğlu[1*], Boris Zimmermann[1], Achim Kohler[1,2]

[1]Department of Mathematical Sciences and Technology, Faculty of Environmental Science and Technology, Norwegian University of Life Sciences, Ås, Norway

[2]Nofima AS, Ås, Norway

*** Corresponding author**

E-mail: murat.bagcioglu@nmbu.no (MB)

Table of Contents **Page**

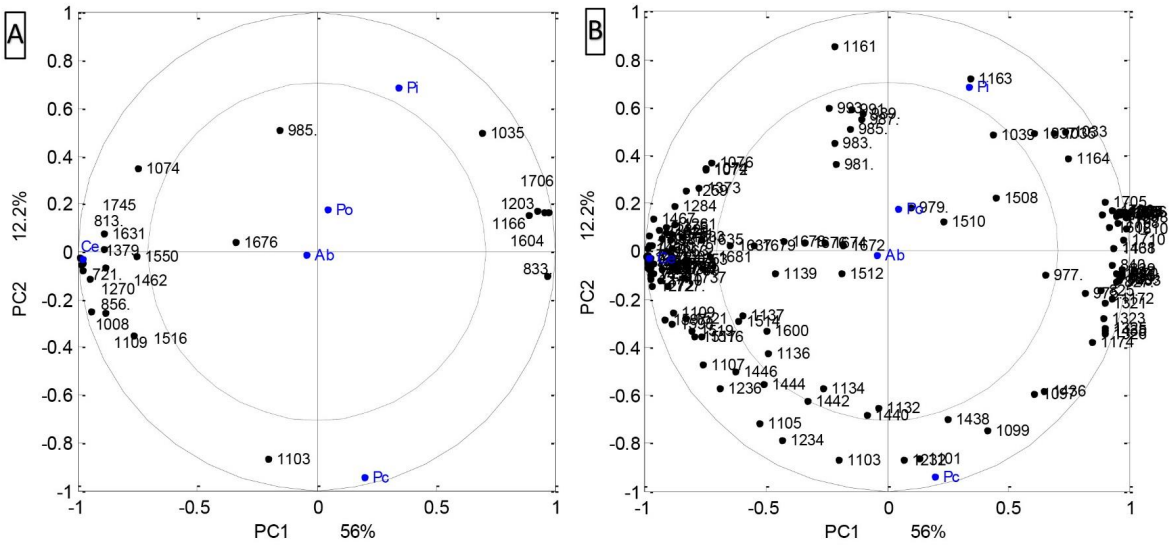


Fig. A (Part 1) CPCA correlation loading plots for the first two principal components. Variables related to ATR IR measurements of intact pollen grains (**ATI**) are presented in black color; **(A)** plot for the selected variables, and **(B)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

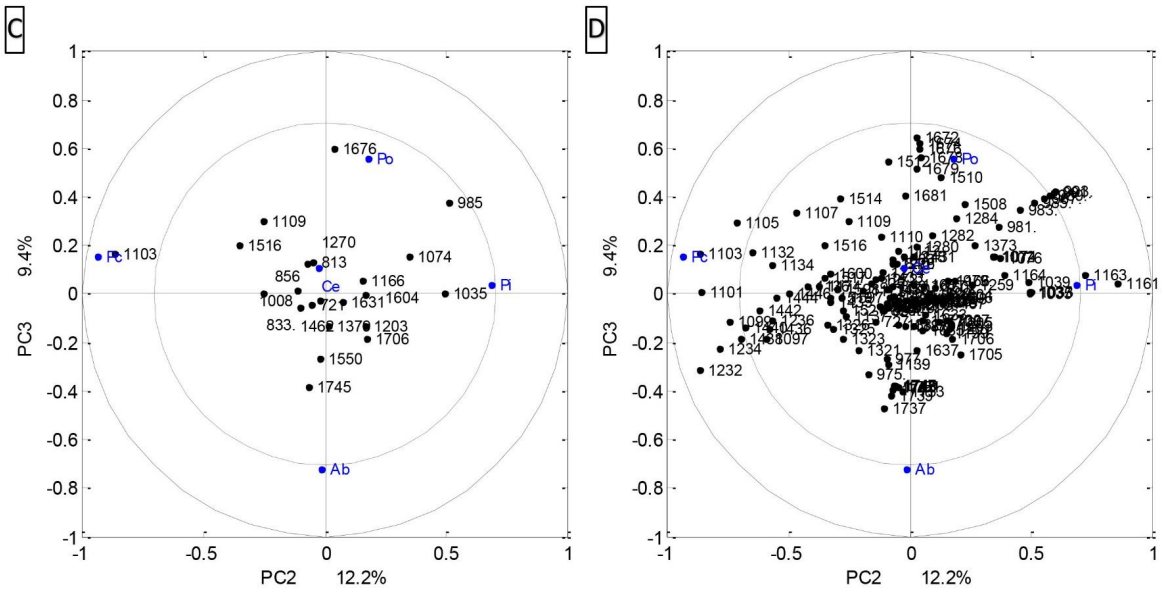


Fig. A (Part 2) CPCA correlation loading plots for the second and the third principal components. Variables related to ATR IR measurements of intact pollen grains (**ATI**) are presented in black color; **(C)** plot for the selected variables, and **(D)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

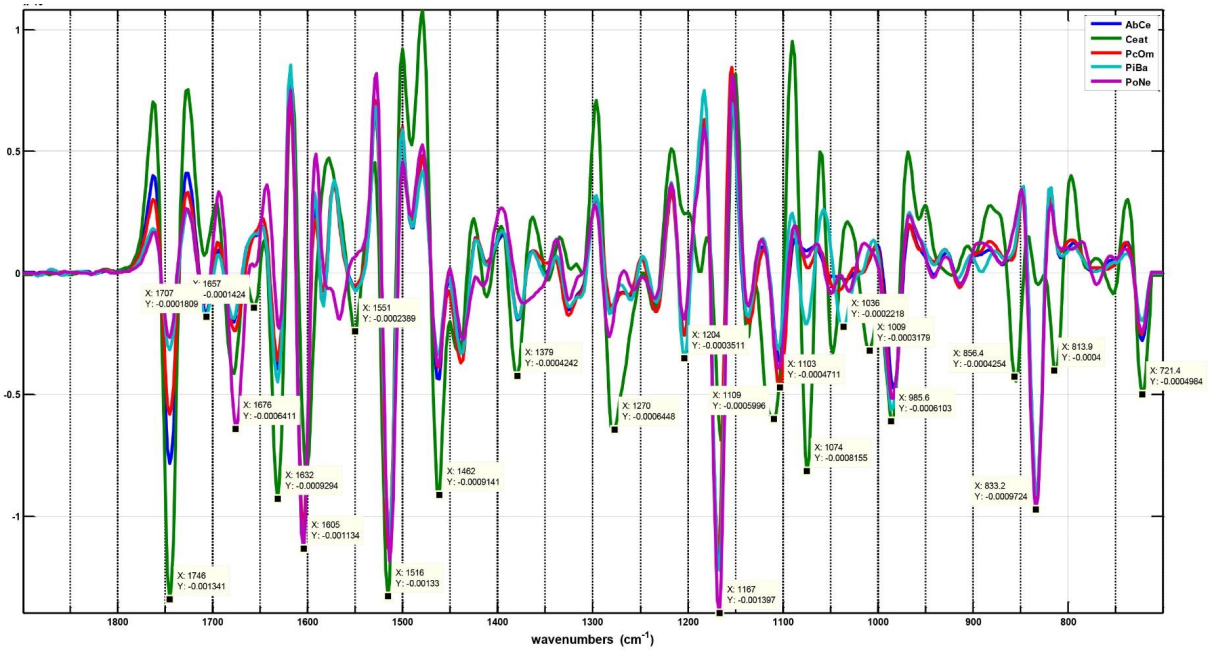


Fig. A (Part 3) Second-derivative and EMSC corrected spectra obtained by ATR IR measurements of intact pollen grains (**ATI**). Average spectra of five representative species are presented: *Abies cephalonica* (AbCe), *Cedrus atlantica* (CeAt), *Picea omorika* (PcOm), *Pinus banksiana* (PiBa), *Podocarpus neriifolius* (PoNe). The selected vibrational bands associated to the CPCA correlation loading plots on Figures S1 are marked.

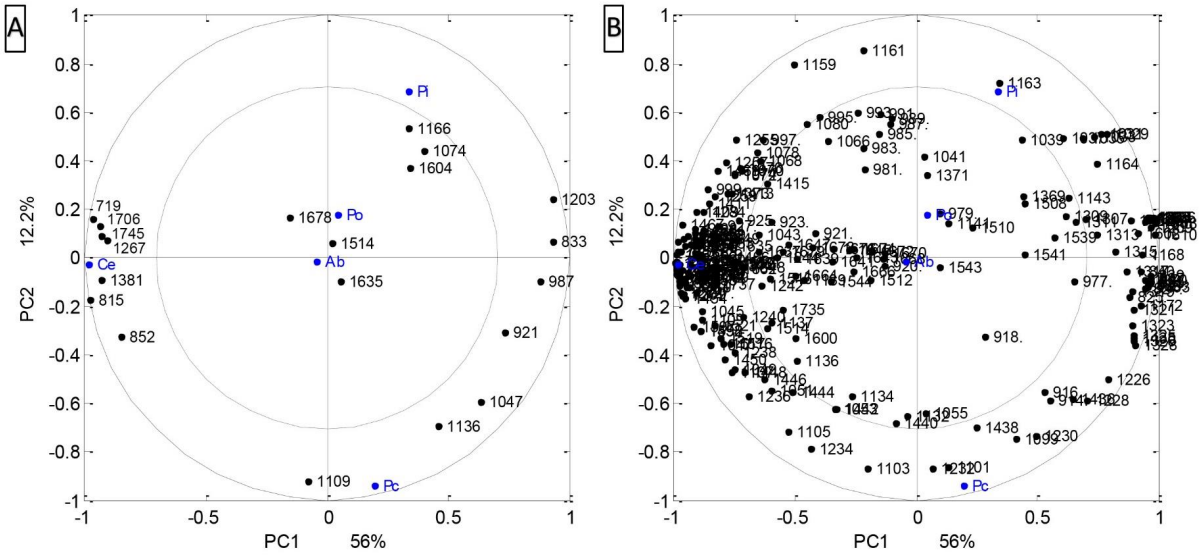


Fig. B (Part 1) CPCA correlation loading plots for the first two principal components. Variables related to ATR IR measurements of ground pollen grains (**ATG**) are presented in black color; **(A)** plot for the selected variables, and **(B)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

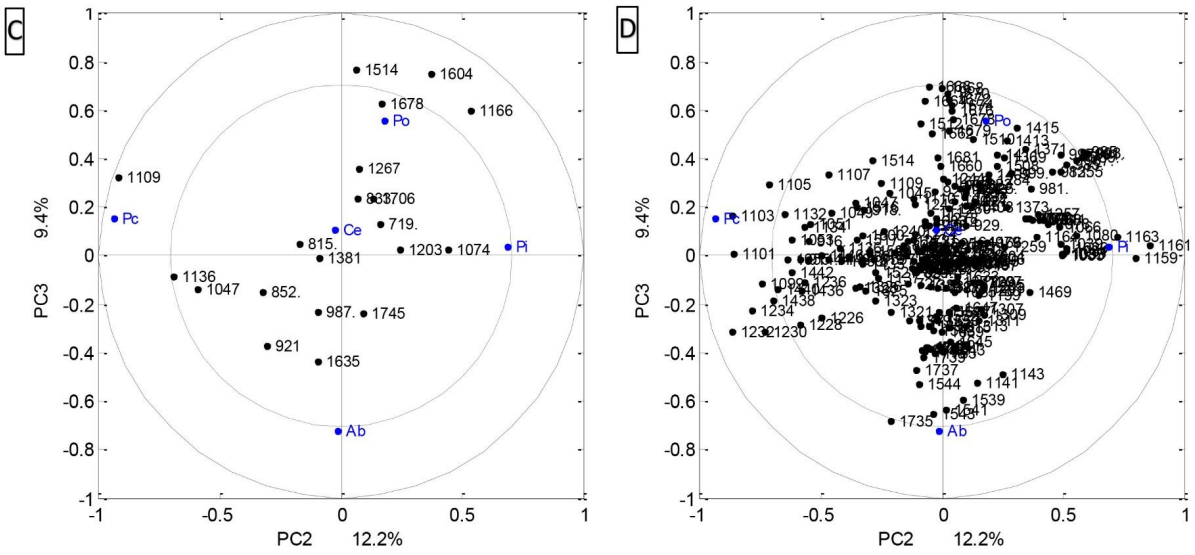


Fig. B (Part 2) CPCA correlation loading plots for the second and the third principal components. Variables related to ATR IR measurements of ground pollen grains (**ATG**) are presented in black color; **(C)** plot for the selected variables, and **(D)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

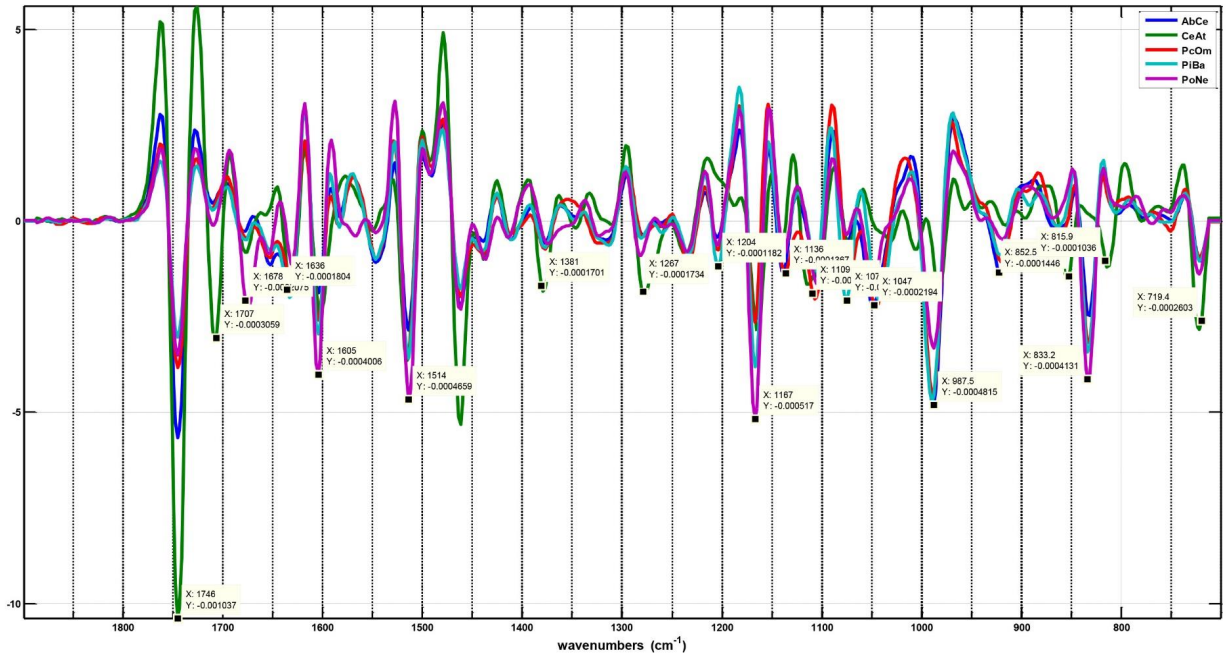


Fig. B (Part 3) Second-derivative and EMSC corrected spectra obtained by ATR IR measurements of ground pollen grains (**ATG**). Average spectra of five representative species are presented: *Abies cephalonica* (AbCe), *Cedrus atlantica* (CeAt), *Picea omorika* (PcOm), *Pinus banksiana* (PiBa), *Podocarpus neriifolius* (PoNe). The selected vibrational bands associated to the CPCA correlation loading plots on Figures S-2 are marked.

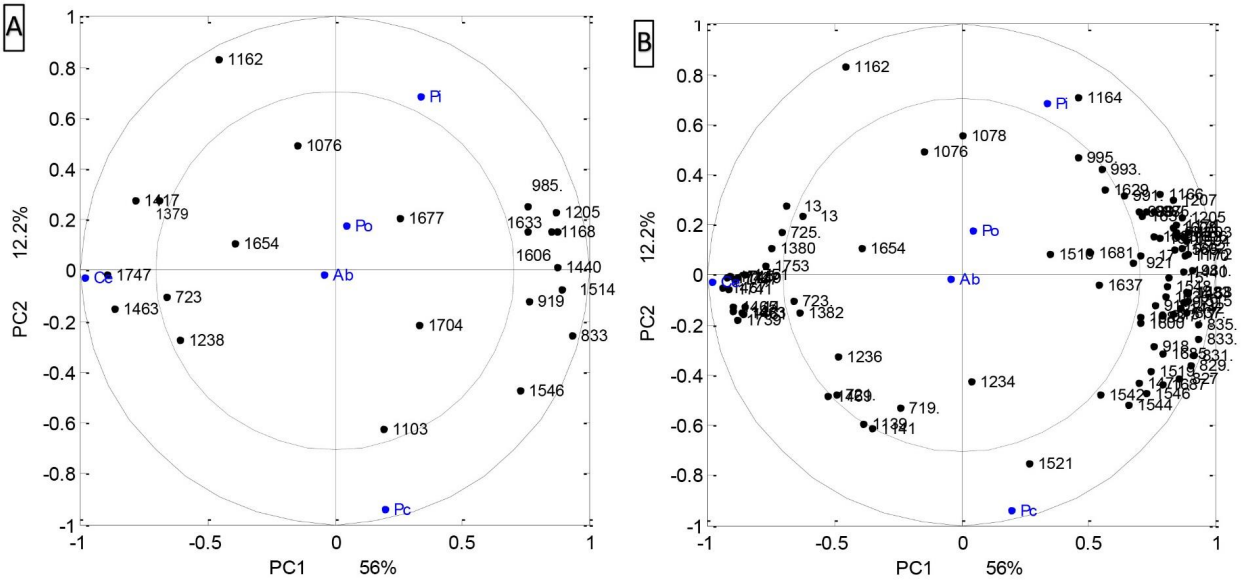


Fig. C (Part 1) CPCA correlation loading plots for the first two principal components. Variables related to transmission FTIR spectroscopy of KBr pellets (**KBR**) are presented in black color; **(A)** plot for the selected variables, and **(B)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

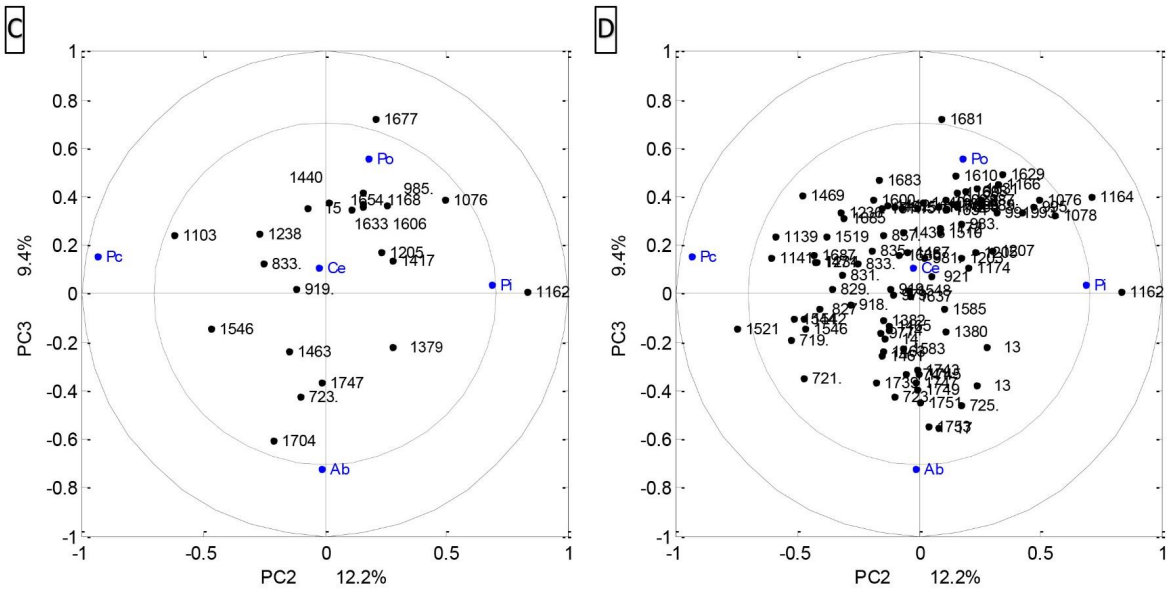


Fig. C (Part 2) CPCA correlation loading plots for the second and the third principal components. Variables related to transmission FTIR spectroscopy of KBr pellets (**KBR**) are presented in black color; **(C)** plot for the selected variables, and **(D)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

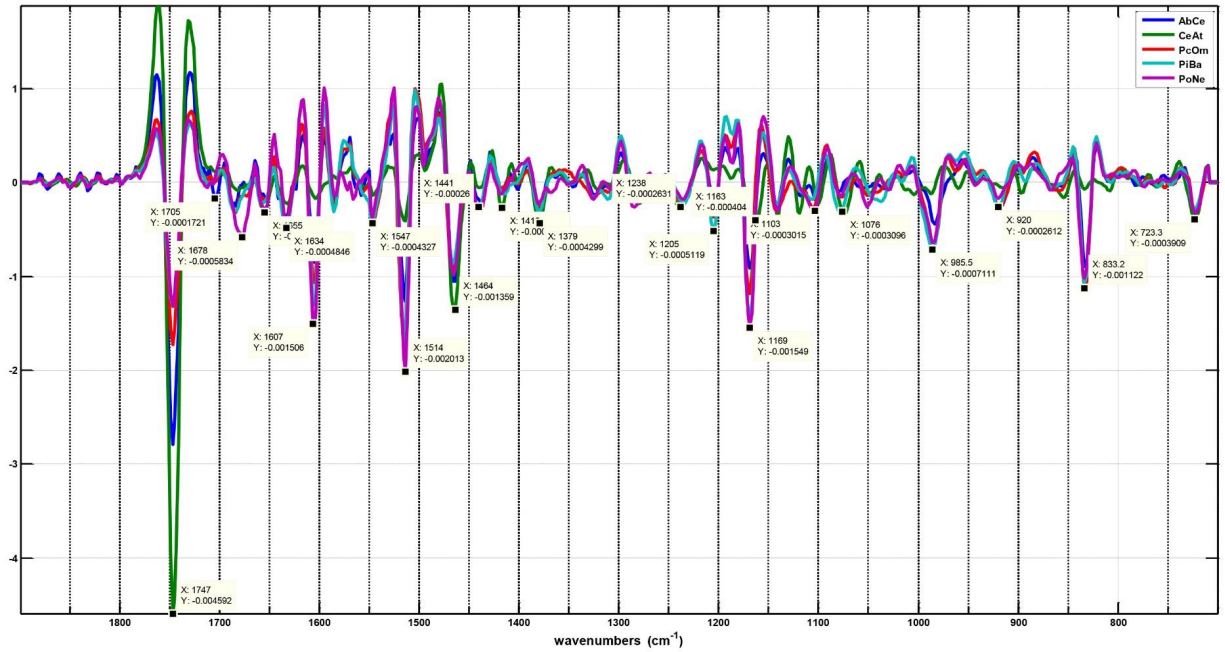


Fig. C (Part 3) Second-derivative and EMSC corrected spectra obtained by transmission FTIR spectroscopy of KBr pellets (**KBR**). Average spectra of five representative species are presented: *Abies cephalonica* (AbCe), *Cedrus atlantica* (CeAt), *Picea omorika* (PcOm), *Pinus banksiana* (PiBa), *Podocarpus neriifolius* (PoNe). The selected vibrational bands associated to the CPCA correlation loading plots on Figures S3 are marked.

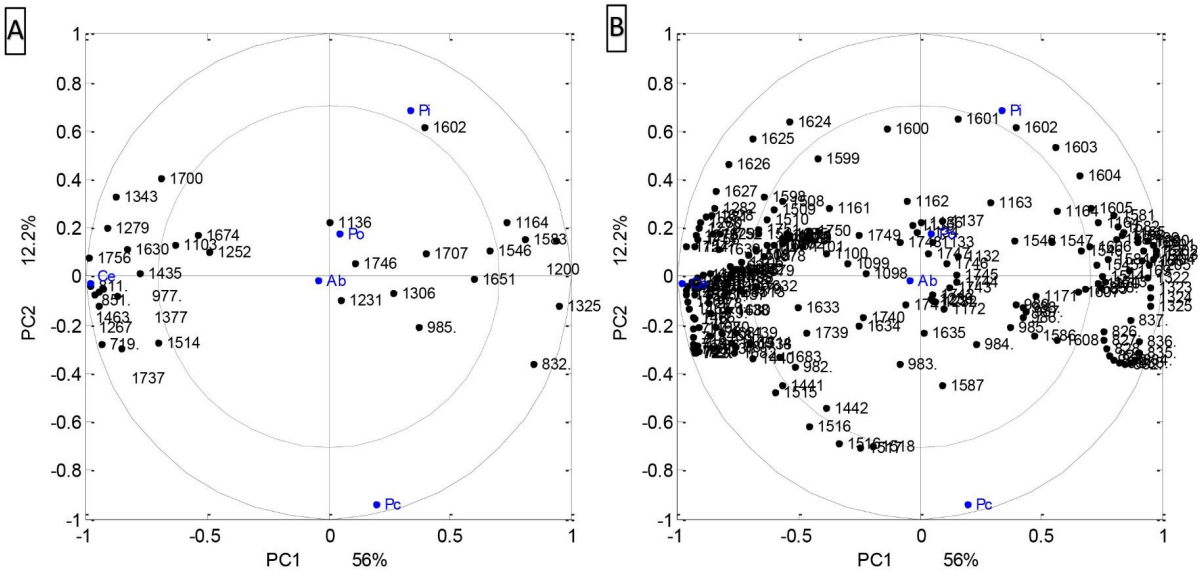


Fig. D (Part 1) CPCA correlation loading plots for the first two principal components. Variables related to transmission FTIR microspectroscopy of pollen multigrain (**MGR**) are presented in black color; **(A)** plot for the selected variables, and **(B)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

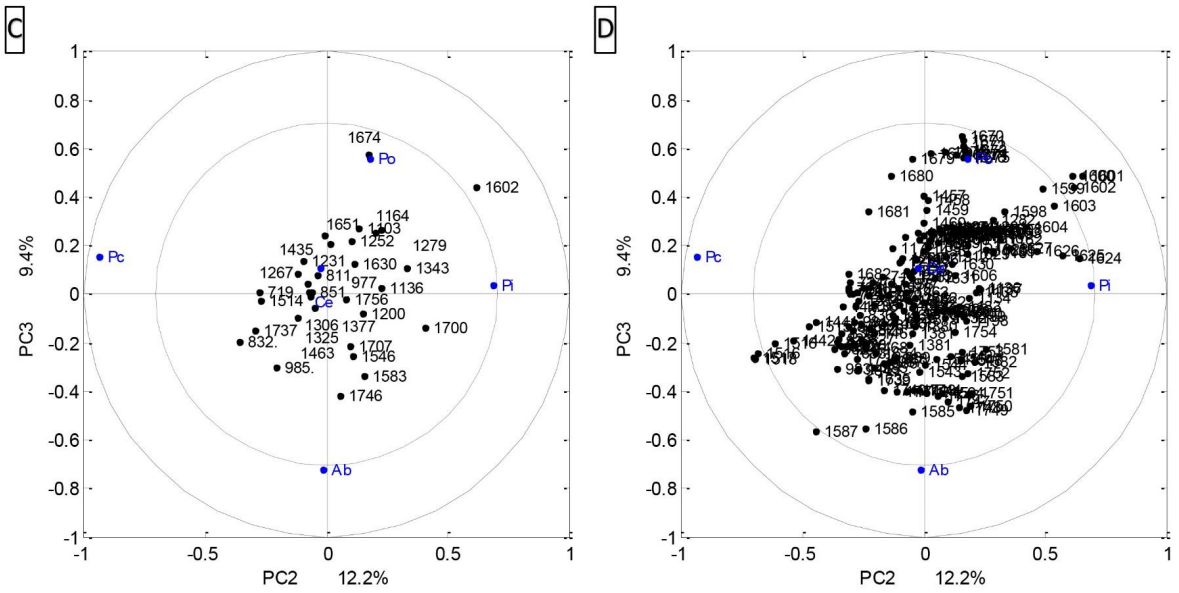


Fig. D (Part 2) CPCA correlation loading plots for the second and the third principal components. Variables related to transmission FTIR microspectroscopy of pollen multigrain (**MGR**) are presented in black color; **(C)** plot for the selected variables, and **(D)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

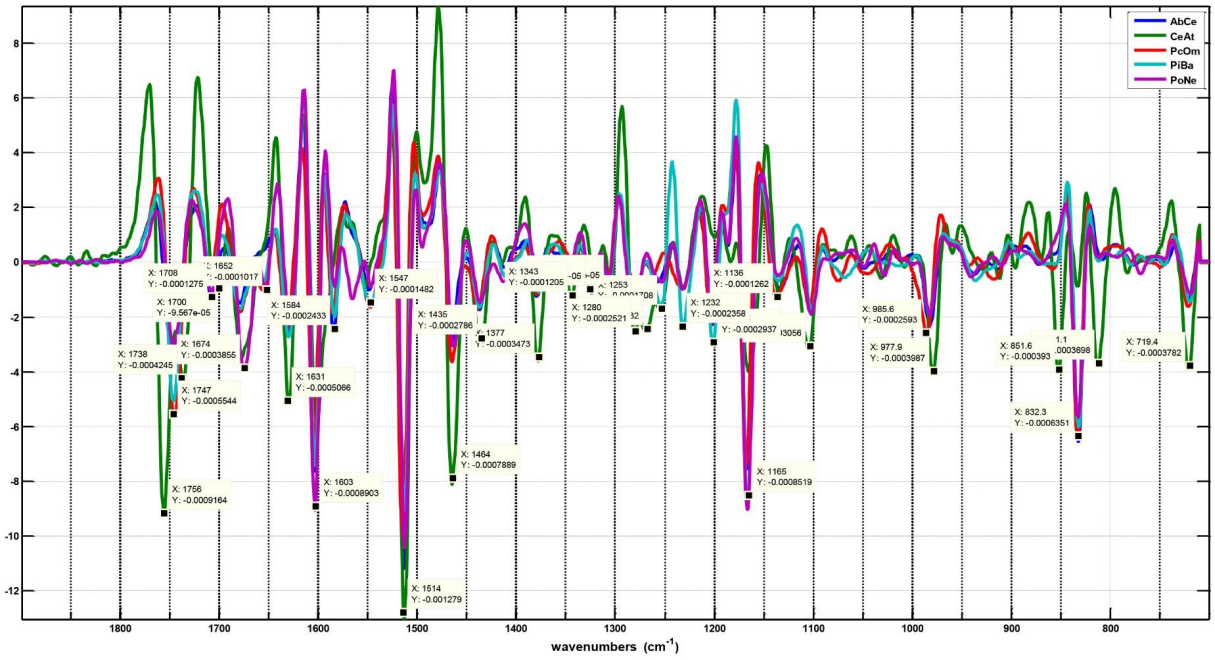


Fig. D (Part 3) Second-derivative and EMSC corrected spectra obtained by transmission FTIR microspectroscopy of pollen multigrain (**MGR**). Average spectra of five representative species are presented: *Abies cephalonica* (AbCe), *Cedrus atlantica* (CeAt), *Picea omorika* (PcOm), *Pinus banksiana* (PiBa), *Podocarpus neriifolius* (PoNe). The selected vibrational bands associated to the CPCA correlation loading plots on Figures S4 are marked.

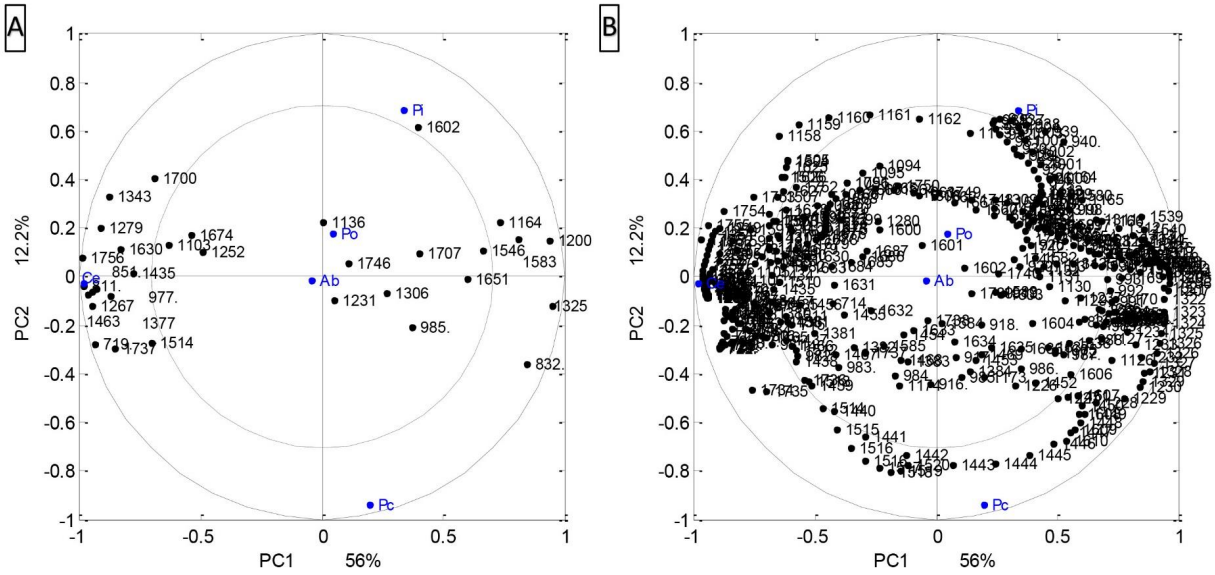


Fig. E (Part 1) CPCA correlation loading plots for the first two principal components. Variables related to transmission FTIR microspectroscopy of single pollen grain (**SGR**) are presented in black color; **(A)** plot for the selected variables, and **(B)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

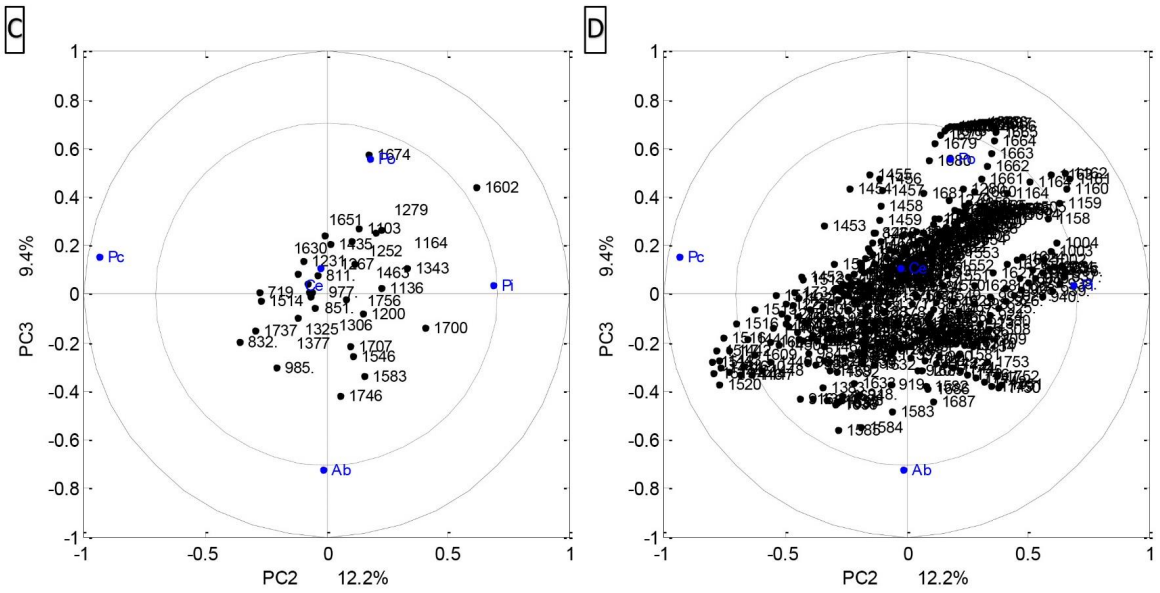


Fig. E (Part 2) CPCA correlation loading plots for the second and the third principal components. Variables related to transmission FTIR microspectroscopy of single pollen grain (**SGR**) are presented in black color; **(C)** plot for the selected variables, and **(D)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

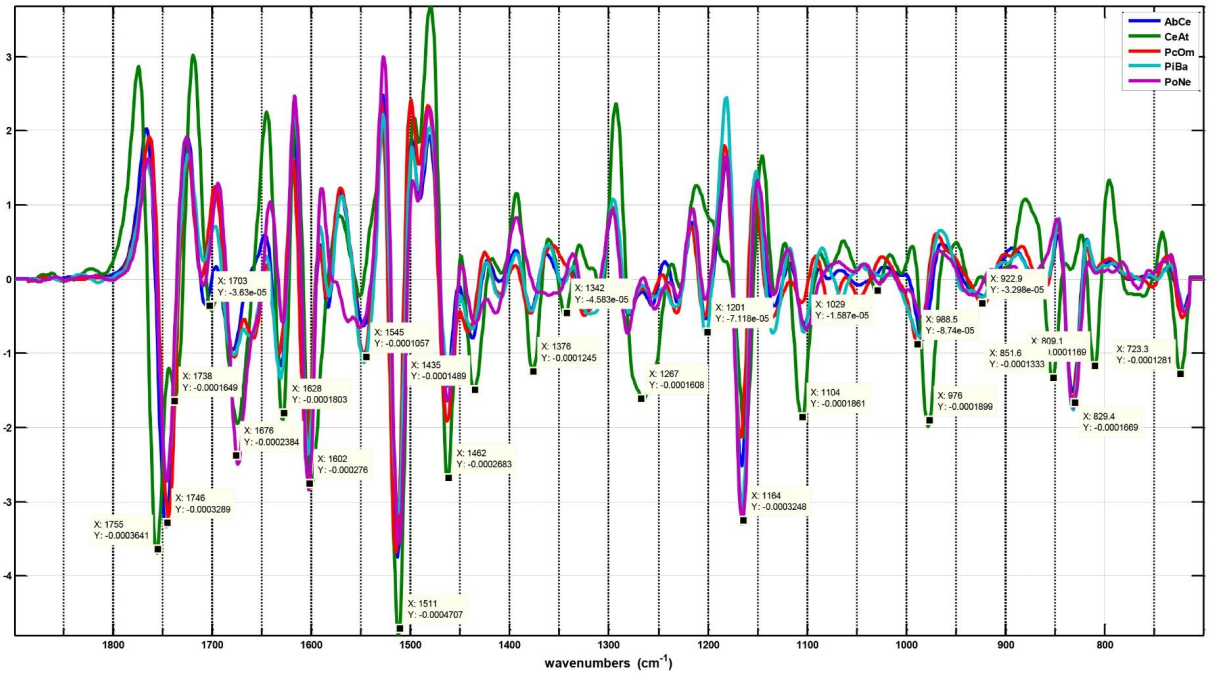


Fig. E (Part 3) Second-derivative and EMSC corrected spectra obtained by transmission FTIR microspectroscopy of single pollen grain (**SGR**). Average spectra of five representative species are presented: *Abies cephalonica* (AbCe), *Cedrus atlantica* (CeAt), *Picea omorika* (PcOm), *Pinus banksiana* (PiBa), *Podocarpus neriifolius* (PoNe). The selected vibrational bands associated to the CPCA correlation loading plots on Figures S5 are marked.

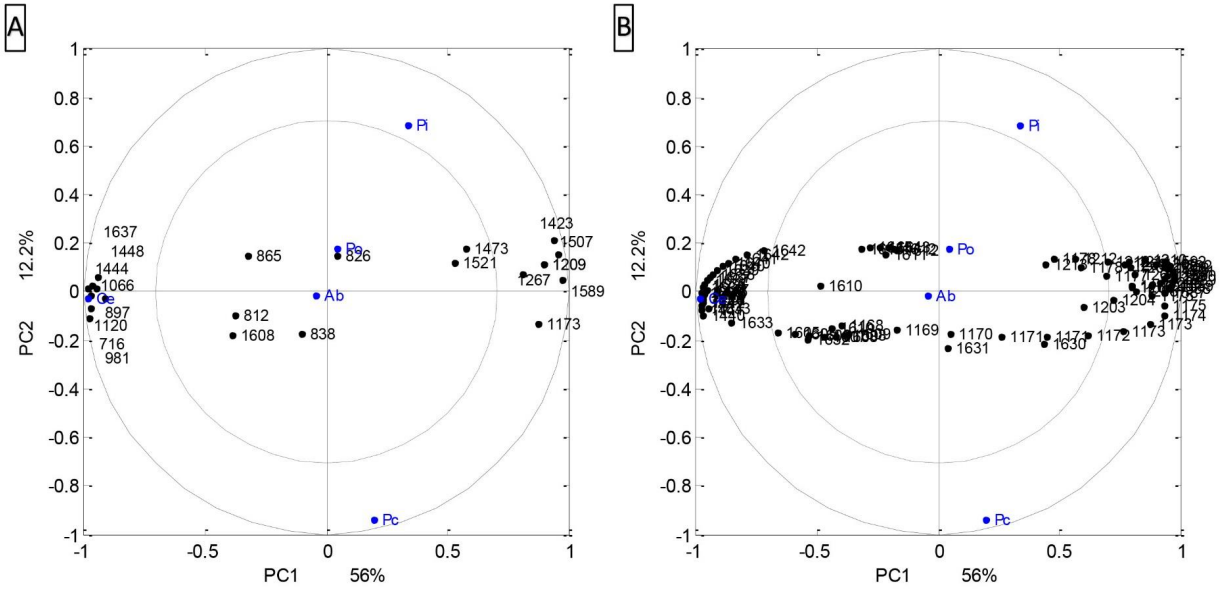


Fig. F (Part 1) CPCA correlation loading plots for the first two principal components. Variables related to Raman microspectroscopy measurements of saccus part of pollen grain (**RMS**) are presented in black color; **(A)** plot for the selected variables, and **(B)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

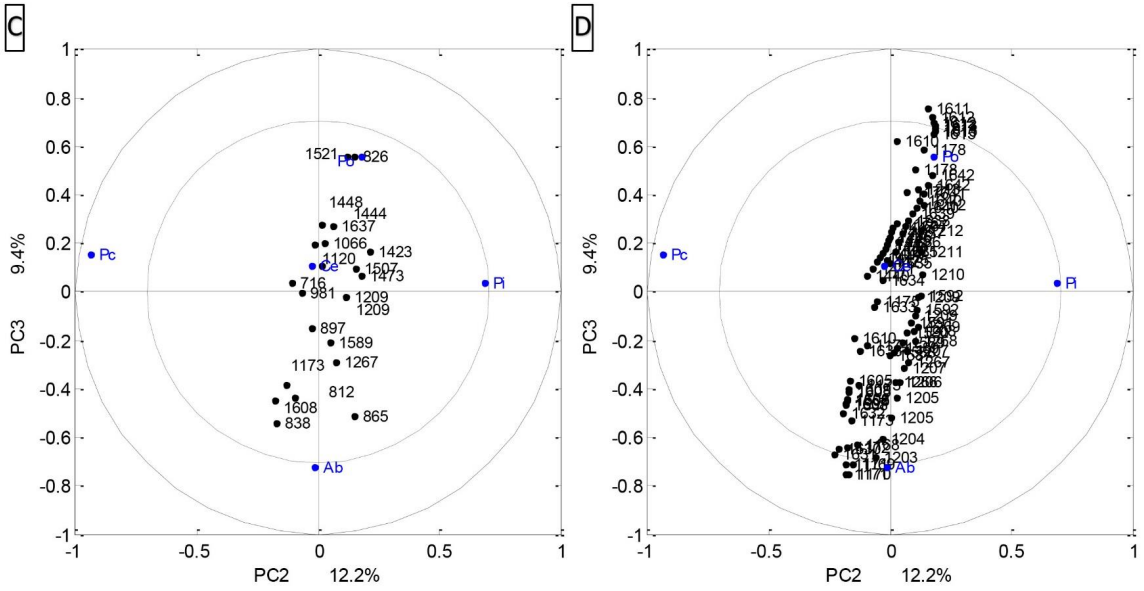


Fig. F (Part 2) CPCA correlation loading plots for the second and the third principal components. Variables related to Raman microspectroscopy measurements of saccus part of pollen grain (**RMS**) are presented in black color; **(C)** plot for the selected variables, and **(D)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

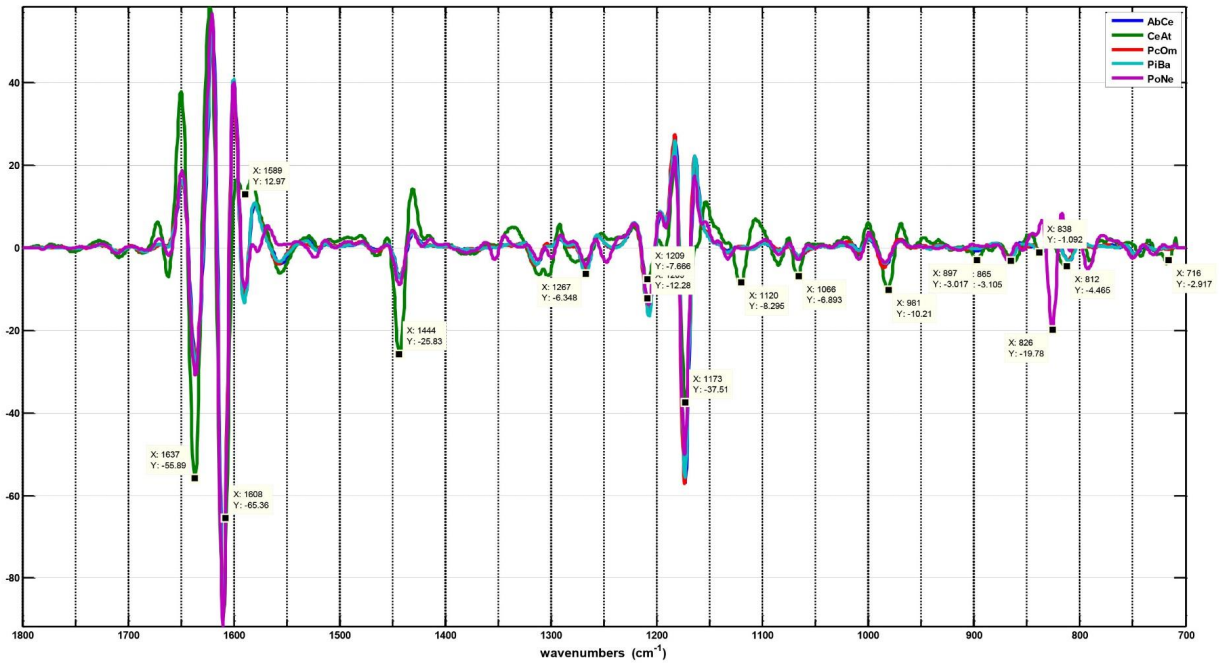


Fig. F (Part 3) Second-derivative and EMSC corrected spectra obtained by Raman microspectroscopy measurements of saccus part of pollen grain (**RMS**). Average spectra of five representative species are presented: *Abies cephalonica* (AbCe), *Cedrus atlantica* (CeAt), *Picea omorika* (PcOm), *Pinus banksiana* (PiBa), *Podocarpus neriifolius* (PoNe). The selected vibrational bands associated to the CPCA correlation loading plots on Figures S6 are marked.

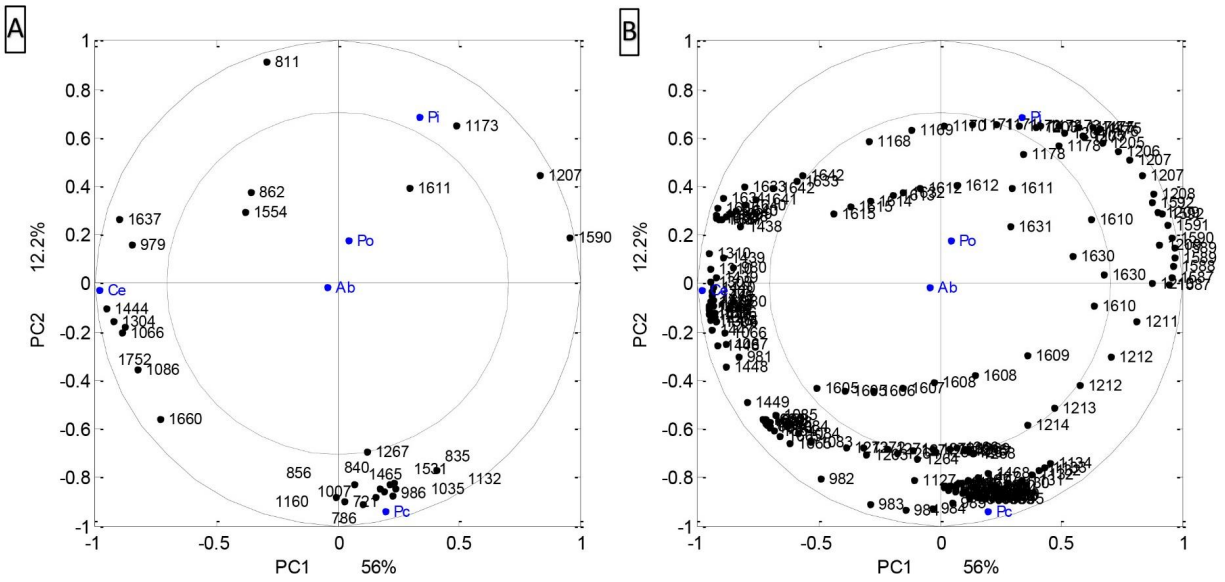


Fig. G (Part 1) CPCA correlation loading plots for the first two principal components. Variables related to Raman microspectroscopy measurements of corpus part of pollen grain (**RMC**) are presented in black color; **(A)** plot for the selected variables, and **(B)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

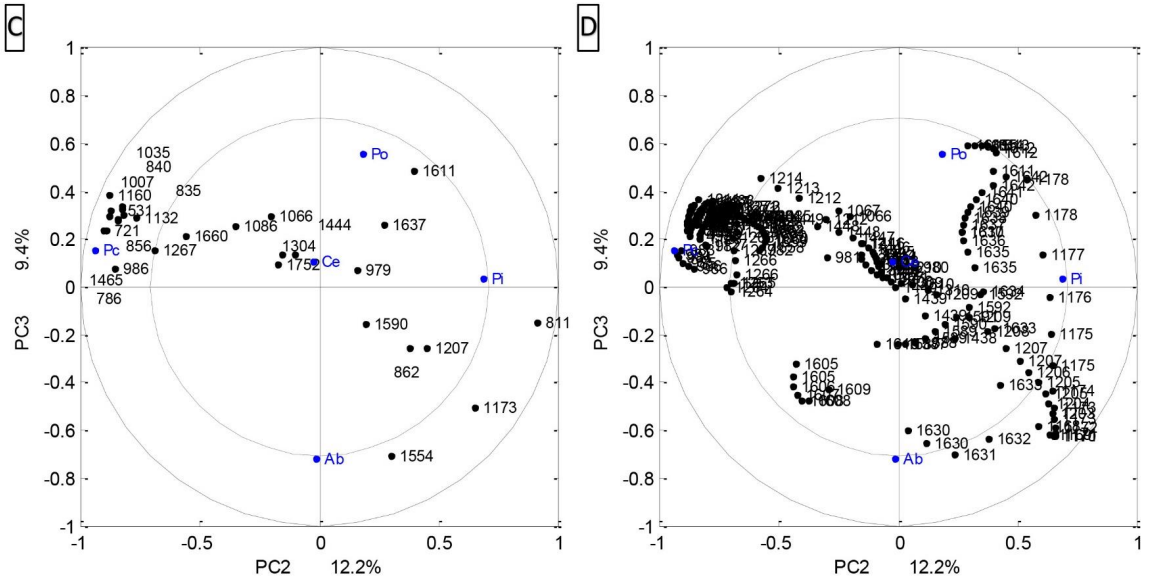


Fig. G (Part 2) CPCA correlation loading plots for the second and the third principal components. Variables related to Raman microspectroscopy measurements of corpus part of pollen grain (**RMC**) are presented in black color; **(C)** plot for the selected variables, and **(D)** plot for all variables. Design variables related to plant genera are presented in blue color: *Abies* (Ab), *Cedrus* (Ce), *Picea* (Pc), *Pinus* (Pi), *Podocarpus* (Po).

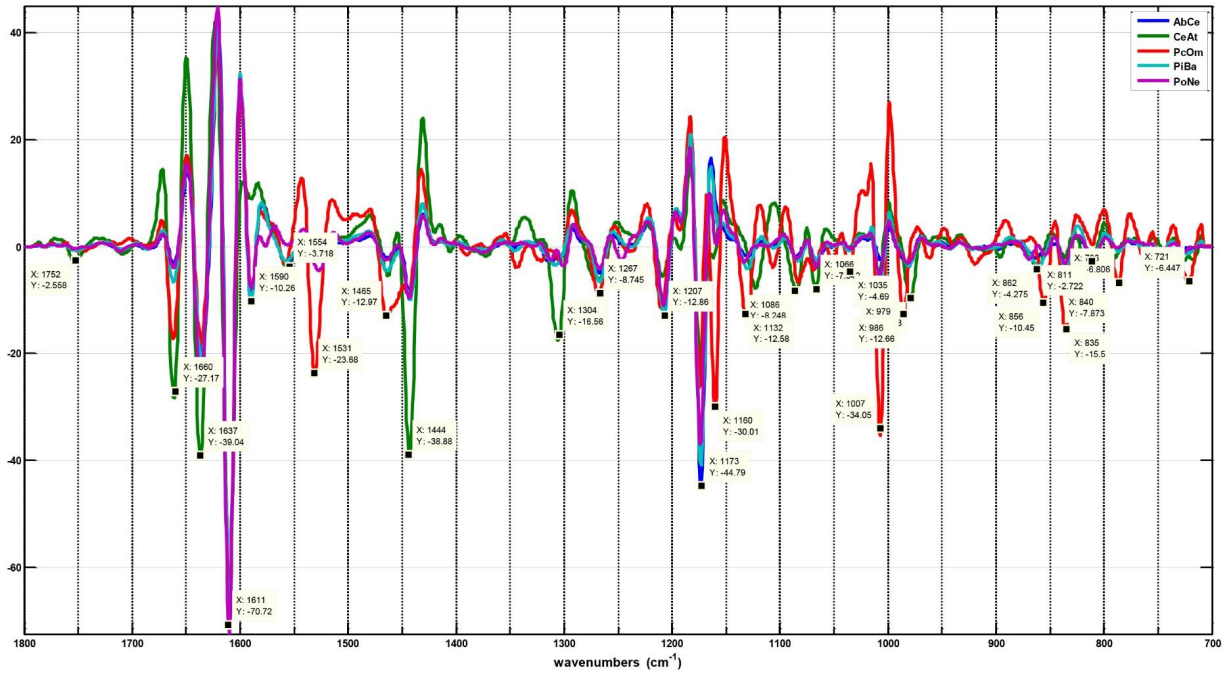


Fig. G (Part 3) Second-derivative and EMSC corrected spectra obtained by Raman microspectroscopy measurements of corpus part of pollen grain (**RMC**). Average spectra of five representative species are presented: *Abies cephalonica* (AbCe), *Cedrus atlantica* (CeAt), *Picea omorika* (PcOm), *Pinus banksiana* (PiBa), *Podocarpus neriifolius* (PoNe). The selected vibrational bands associated to the CPCA correlation loading plots on Figures S7 are marked.

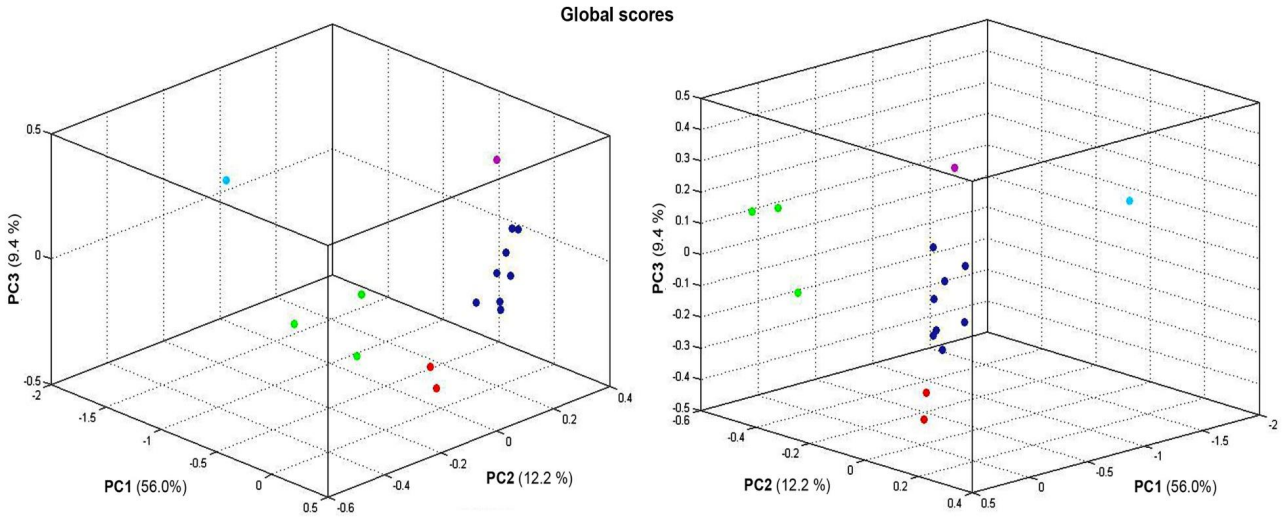


Fig. H 3D presentation of global scores of consensus principal component analysis (CPCA). Markings are labelled in accordance to pollen genus: *Abies* (red), *Cedrus* (blue), *Picea* (green), *Pinus* (dark yellow), and *Podocarpus* (violet). The percent variances for the first five PCs are 56.0, 12.2, 9.4, 7.5 and 5.1.

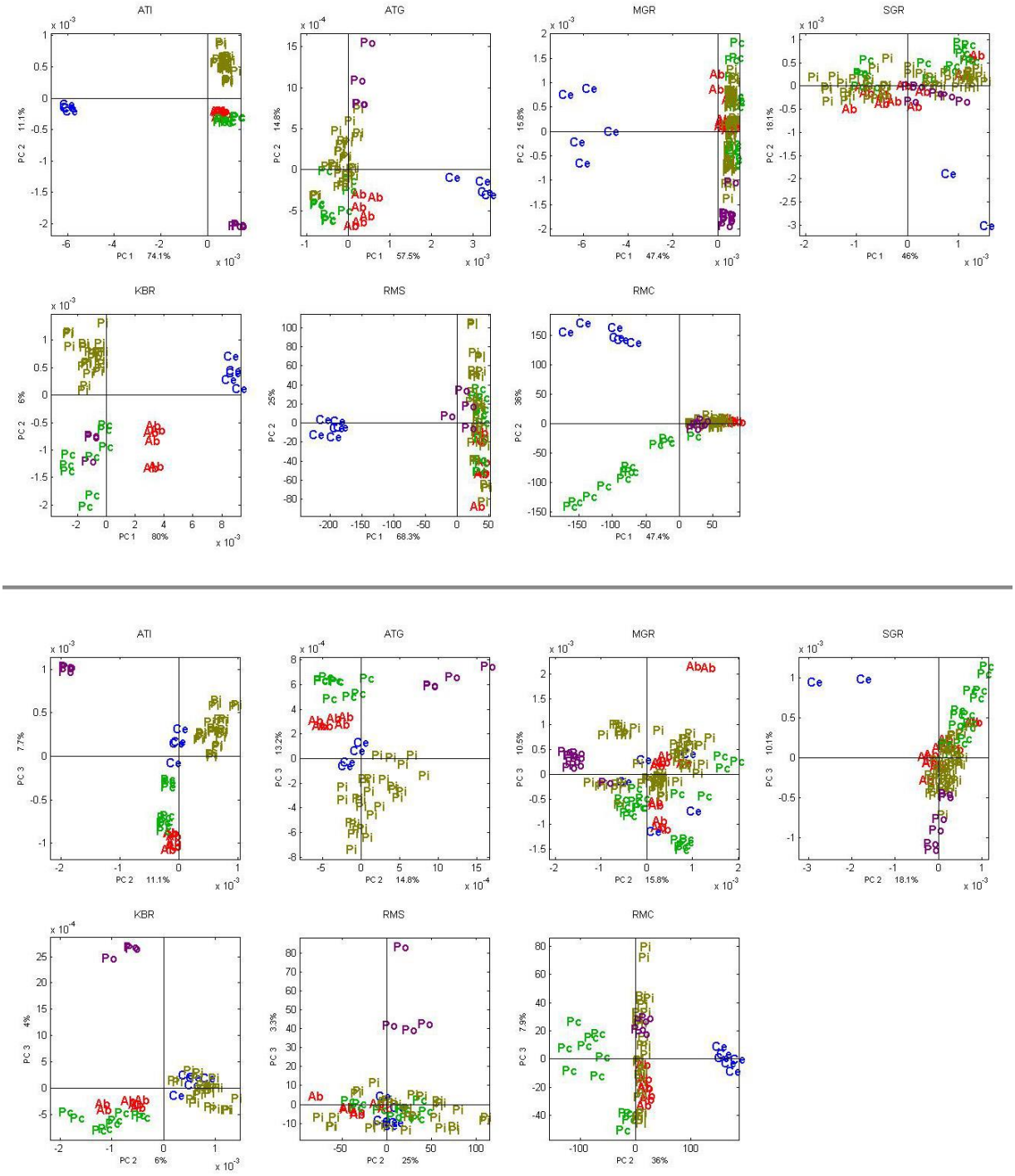


Fig. I PCA score plots performed on individual spectral data blocks: 1) ATI (ATR-FTIR of intact pollen), 2) ATG (ATR-FTIR of ground pollen), 3) MGR (transmission FTIR microspectroscopy of multigrain), 4) SGR (transmission FTIR microspectroscopy of single grain), 5) KBR (transmission FTIR of KBr pellets), 6) RMC (Raman of corpus region) and 7) RMS (Raman of saccus region). Samples are labelled in accordance to pollen genus: *Abies* (Ab, red), *Cedrus* (Ce, blue), *Picea* (Pc, green), *Pinus* (Pi, dark yellow), *Podocarpus* (Po, violet).

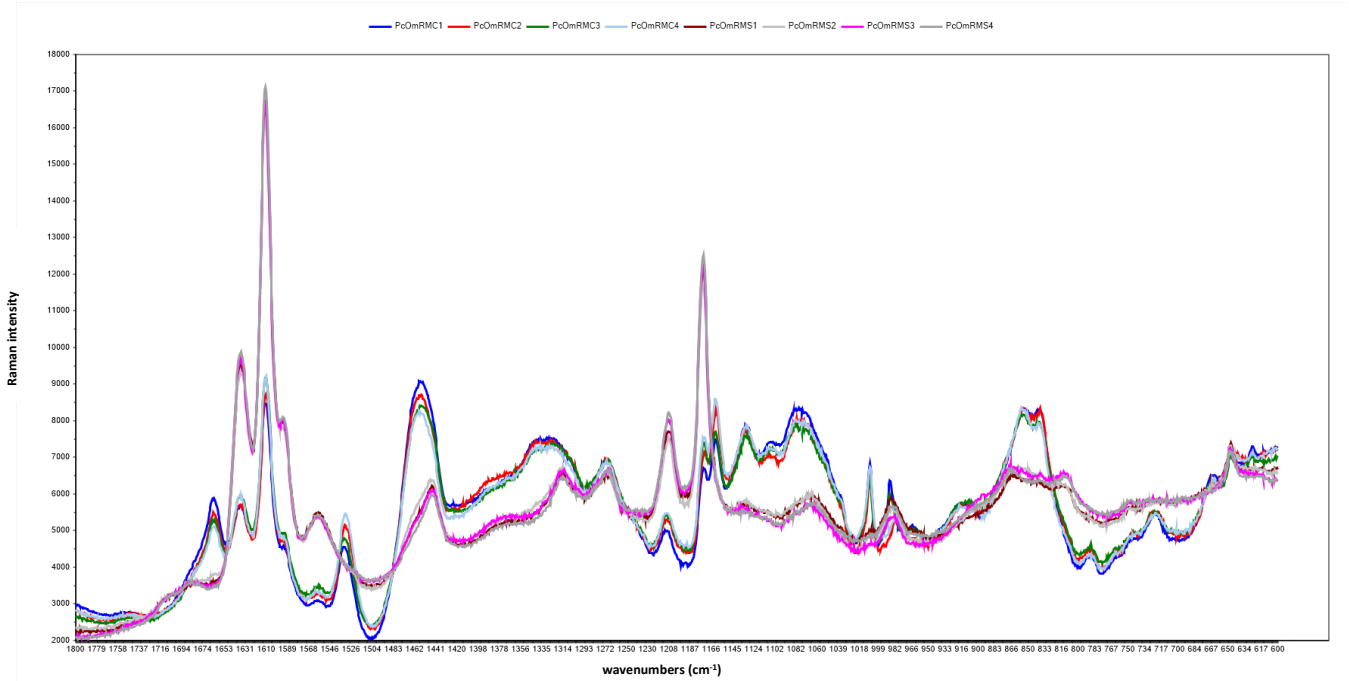


Fig. J Raman spectra of representative samples of *Picea omorika* pollen. The spectral set consists of EMSC normalized spectra of measurements of corpus region (RMC), and saccus region (RMS), with four spectra per region.

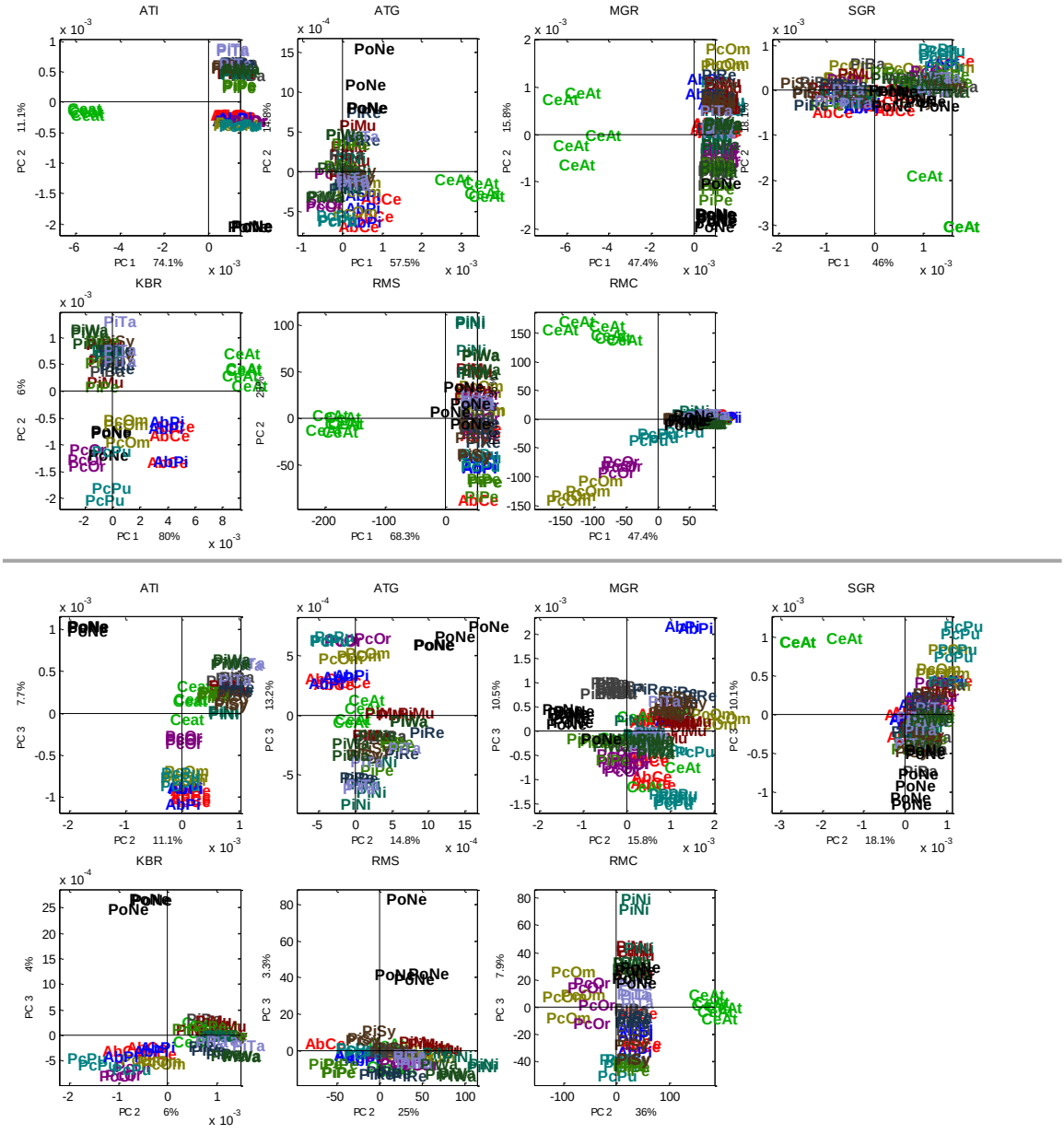


Fig. K PCA score plots performed on individual spectral data blocks: 1) ATI (ATR-FTIR of intact pollen), 2) ATG (ATR-FTIR of ground pollen), 3) MGR (transmission FTIR microspectroscopy of multigrain), 4) SGR (transmission FTIR microspectroscopy of single grain), 5) KBR (transmission FTIR of KBr pellets), 6) RMC (Raman of corpus region) and 7) RMS (Raman of saccus region). Samples are labelled in accordance to pollen species: *Abies cephalonica* (AbCe), *Cedrus atlantica* (CeAt), *Picea omorika* (PcOm), *Picea orientalis* (PcOr), *Picea pungens* (PcPu), *Pinus banksiana* (PiBa), *Pinus mugo* (PiMu), *Pinus nigra* (PiNi), *Pinus peuce* (PiPe), *Pinus resinosa* (PiRe), *Pinus sylvestris* (PiSy), *Pinus tabuliformis* (PiTa), *Pinus wallichiana* (PiWa),*Podocarpus neriifilious* (PoNe).

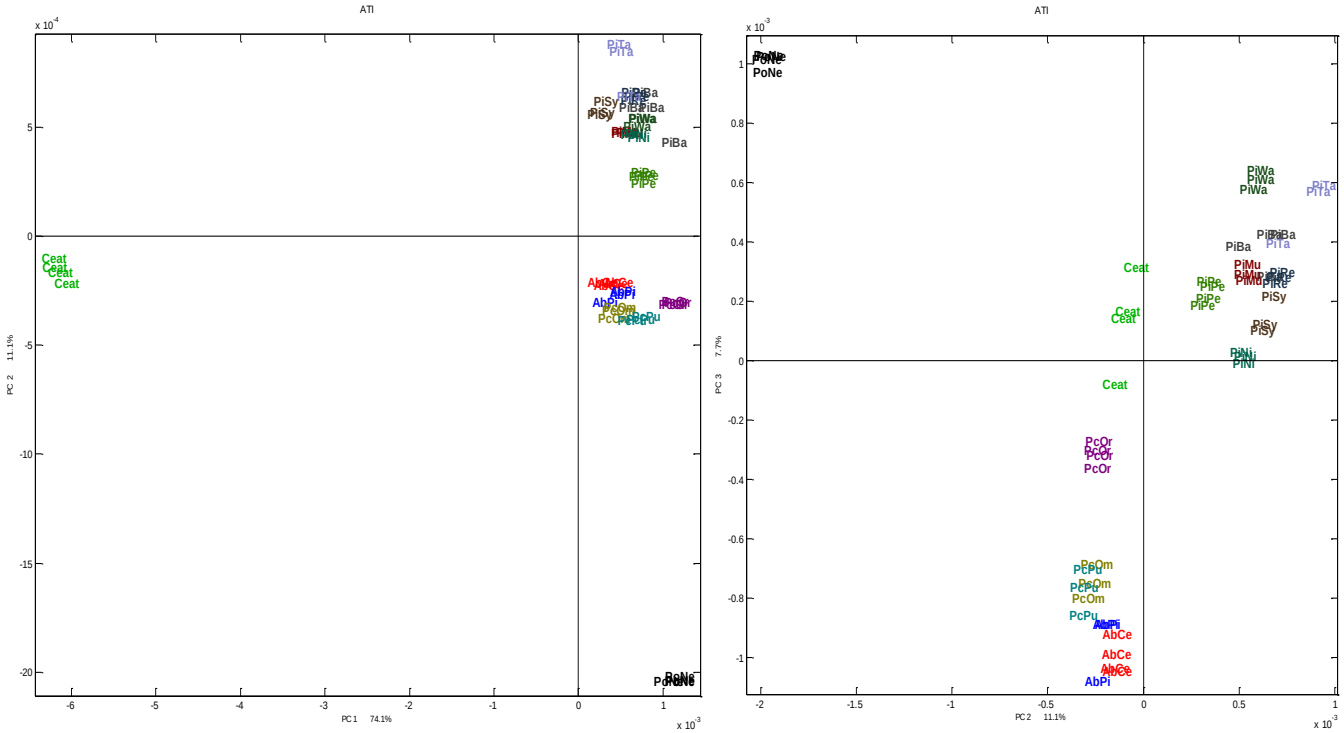


Fig. L PCA score plots performed on individual spectral data blocks: ATI (ATR-FTIR of intact pollen). Samples are labelled in accordance to pollen species: *Abies cephalonica* (AbCe), *Cedrus atlantica* (CeAt), *Picea omorika* (PcOm), *Picea orientalis* (PcOr), *Picea pungens* (PcPu), *Pinus banksiana* (PiBa), *Pinus mugo* (PiMu), *Pinus nigra* (PiNi), *Pinus peuce* (PiPe), *Pinus resinosa* (PiRe), *Pinus sylvestris* (PiSy), *Pinus tabuliformis* (PiTa), *Pinus wallichiana* (PiWa),*Podocarpus neriifilious* (PoNe).

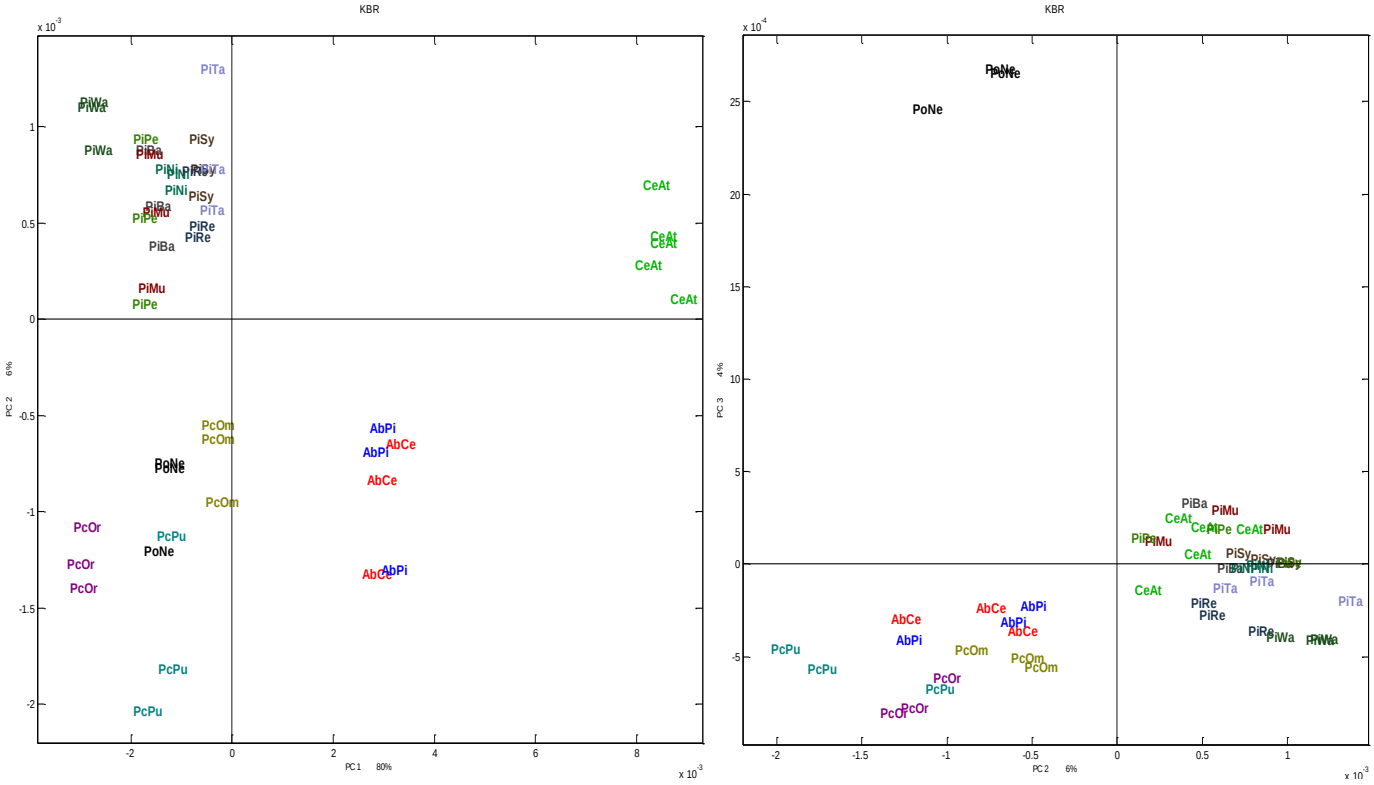


Fig. M PCA score plots performed on individual spectral data blocks: KBR (transmission FTIR of KBr pellets). Samples are labelled in accordance to pollen species: *Abies cephalonica* (AbCe), *Cedrus atlantica* (CeAt), *Picea omorika* (PcOm), *Picea orientalis* (PcOr), *Picea pungens* (PcPu), *Pinus banksiana* (PiBa), *Pinus mugo* (PiMu), *Pinus nigra* (PiNi), *Pinus peuce* (PiPe), *Pinus resinosa* (PiRe), *Pinus sylvestris* (PiSy), *Pinus tabuliformis* (PiTa), *Pinus wallichiana* (PiWa),*Podocarpus neriifilious* (PoNe).

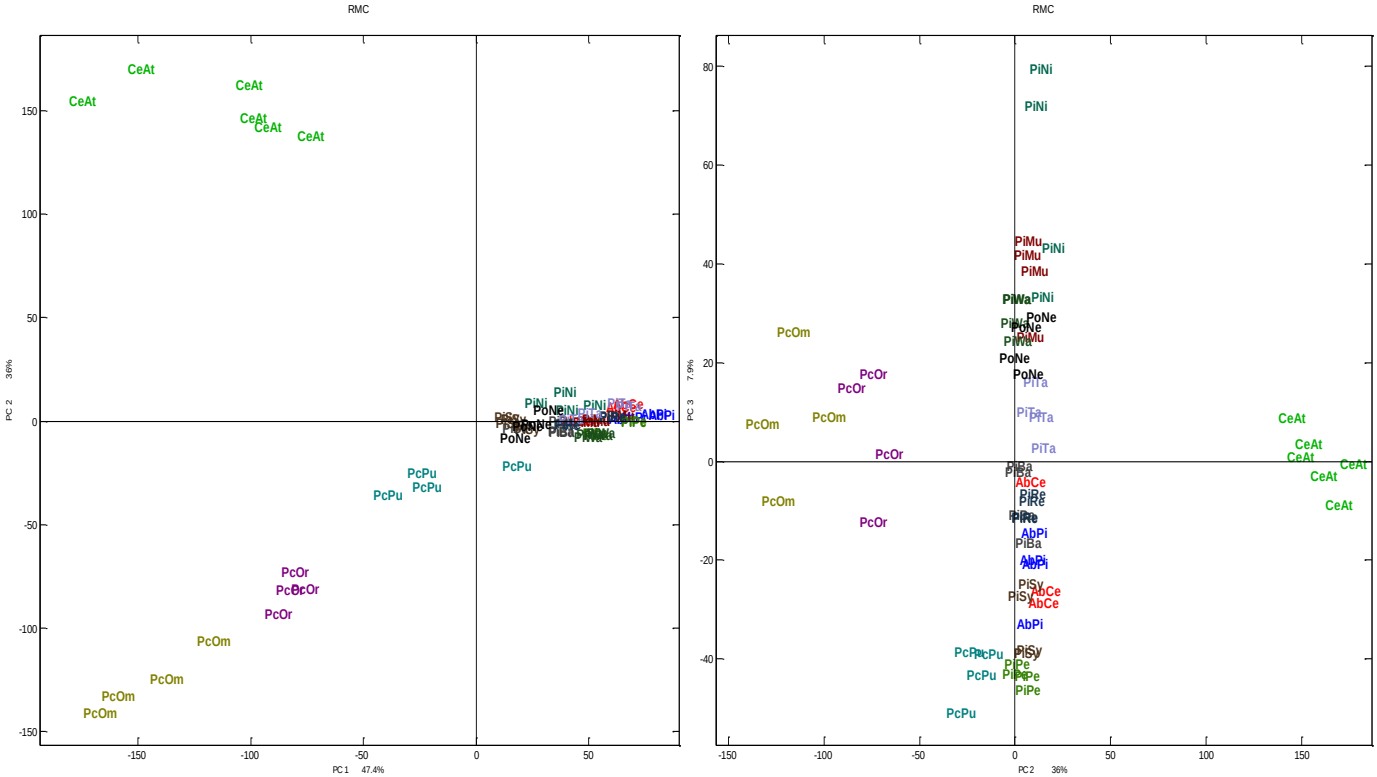


Fig. N PCA score plots performed on individual spectral data blocks: RMC (Raman spectra of corpus region). Samples are labelled in accordance to pollen species: *Abies cephalonica* (AbCe), *Cedrus atlantica* (CeAt), *Picea omorika* (PcOm), *Picea orientalis* (PcOr), *Picea pungens* (PcPu), *Pinus banksiana* (PiBa), *Pinus mugo* (PiMu), *Pinus nigra* (PiNi), *Pinus peuce* (PiPe), *Pinus resinosa* (PiRe), *Pinus sylvestris* (PiSy), *Pinus tabuliformis* (PiTa), *Pinus wallichiana* (PiWa),*Podocarpus neriifilious* (PoNe).

Table A Variability measurements within replicates[§] of *Pinus* genus.

IR Region	**Species**								**Genus**
1200-700 cm^{-1} (1-*PCC)x10^{-4}**	***Pinus banksiana***	***Pinus peuce***	***Pinus mugo***	***Pinus nigra***	***Pinus resinosa***	***Pinus sylvetris***	***Pinus tabuliformis***	***Pinus, wallichiana***	***Pinus***
ATI	0.003681	0.000889	0.001097	0.001414	0.001061	0.001272	0.001957	0.000922	0.0043
ATG	0.008495	0.005373	0.024527	0.030032	0.019066	0.010331	0.038434	0.015859	0.0381
KBR	0.074957	0.096663	0.029511	0.168076	0.177112	0.07285	0.088509	0.064469	0.1760
MGR	0.006705	0.002553	0.009297	0.001983	0.012672	0.006115	0.001965	0.004316	0.0180
SGR	0.269506	0.074953	0.235117	0.080506	0.11851	0.254767	0.085362	0.068036	0.1927
RMS	0.00051	0.000404	0.001007	0.002186	0.000739	0.001633	0.000143	0.000233	0.0070
RMC	0.002131	0.001705	0.001667	0.004697	0.002735	0.001638	0.00268	0.000619	0.0108

IR Region	**Species**								**Genus**
1800-1500 cm^{-1} (1-*PCC)x10^{-4}**	***Pinus banksiana***	***Pinus peuce***	***Pinus mugo***	***Pinus nigra***	***Pinus resinosa***	***Pinus sylvetris***	***Pinus tabuliformis***	***Pinus wallichiana***	***Pinus***
ATI	0.005322	0.000278	0.000382	0.009551	0.000311	0.003848	0.009442	0.000768	0.0067
ATG	0.026859	0.099302	0.095689	0.122511	0.031509	0.002885	0.026689	0.00311	0.0698
KBR	0.040537	0.073159	0.046467	0.00788	0.021484	0.019477	6.60E-02	0.01467	0.0487
MGR	0.008991	0.038083	0.004665	0.020438	0.023404	0.006171	0.012604	0.009468	0.0293
SGR	0.134267	0.043698	0.068565	0.109443	0.023178	0.107776	0.076228	0.016804	0.1057
RMS	0.000186	0.000219	0.000585	0.001669	0.000926	0.001032	8.40E-05	0.000381	0.0081
RMC	0.000538	0.000187	0.000365	0.002432	0.000268	0.000659	0.000346	0.000137	0.0074

*Pearson Correlation Coefficient (PCC)

[§]Replicate spectra were obtained from same sample measured as at least three times.

Table B Variability measurements within replicates[§] of *Picea* genus

IR Region	Species			Genus
1200-700 cm^{-1} (1-*PCC)x10^{-4}**	***Picea omorika***	***Picea oritentalis***	***Picea pungens***	***Picea***
ATI	0.00109	0.00165	0.00308	0.00360
ATG	0.00858	0.01830	0.01088	0.00287
KBR	0.26673	0.06914	0.25413	0.37640
MGR	0.00377	0.00272	0.00659	0.01660
SGR	0.08598	0.05455	0.01603	0.06600
RMS	0.00164	0.00078	0.00030	0.00380
RMC	0.02821	0.01491	0.01526	0.06600

IR Region	Species			Genus
1800-1500 cm^{-1} (1-*PCC)x10^{-4}**	***Picea omorika***	***Picea oritentalis***	***Picea pungens***	***Picea***
ATI	0.00065	0.00097	0.00223	0.00460
ATG	0.00537	0.22381	0.02451	0.10210
KBR	0.02929	0.01872	0.11992	0.09470
MGR	0.00152	0.27308	0.01386	0.11870
SGR	0.05143	0.04629	0.00413	0.05790
RMS	0.00069	0.00103	0.00016	0.00350
RMC	0.00303	0.00325	0.00145	0.00960

*Pearson Correlation Coefficient (PCC)

[§]Replicate spectra were obtained from same sample measured as at least three times.

Reproducibility and variability of vibrational spectroscopic measurements was estimated by Pearson's correlation coefficient (*PCC*). For the sake of clarity, due to attained *PCC*-values, which were very close to one, results were displayed as 1-PCC.The closer the obtained values are to zero, the lower the variability in the respective data set. The variability test was performed for the two different spectral regions. The spectra were transformed to second derivative form by Savitzky-Golay (SG) algorithm algorithm followed by Extended multiplicative signal correction (EMSC) pre-processing, same as performed in Consensus principal component analysis (CPCA).

6.2 Paper II

Planta (2015) 242:1237–1250
DOI 10.1007/s00425-015-2380-7

EMERGING TECHNOLOGIES

Vibrational microspectroscopy enables chemical characterization of single pollen grains as well as comparative analysis of plant species based on pollen ultrastructure

Boris Zimmermann[1] · Murat Bağcıoğlu[1] · Christophe Sandt[2] · Achim Kohler[1,3]

Received: 2 April 2015 / Accepted: 7 August 2015 / Published online: 20 August 2015

Abstract
Main conclusion **Chemical imaging of pollen by vibrational microspectroscopy enables characterization of pollen ultrastructure, in particular phenylpropanoid components in grain wall for comparative study of extant and extinct plant species.**

A detailed characterization of conifer (Pinales) pollen by vibrational microspectroscopy is presented. The main problems that arise during vibrational measurements were scatter and saturation issues in Fourier transform infrared (FTIR), and fluorescence and penetration depth issues in Raman. Single pollen grains larger than approx. 15 μm can be measured by FTIR microspectroscopy using conventional light sources, while smaller grains may be measured by employing synchrotron light sources. Pollen grains that were larger than 50 μm were too thick for FTIR imaging since the grain constituents absorbed almost all infrared light. Chemical images of pollen were obtained on sectioned samples, unveiling the distribution and concentration of proteins, carbohydrates, sporopollenins and lipids within pollen substructures. The comparative analysis of pollen species revealed that, compared with other Pinales pollens, *Cedrus atlantica* has a higher relative amount of lipid nutrients, as well as different chemical composition of grain wall sporopollenin. The pre-processing and data analysis, namely extended multiplicative signal correction and principal component analysis, offer simple estimate of imaging spectral data and indirect estimation of physical properties of pollen. The vibrational microspectroscopy study demonstrates that detailed chemical characterization of pollen can be obtained by measurement of an individual grain and pollen ultrastructure. Measurement of phenylpropanoid components in pollen grain wall could be used, not only for the reconstruction of past environments, but for assessment of diversity of plant species as well. Therefore, analysis of extant and extinct pollen species by vibrational spectroscopies is suggested as a valuable tool in biology, ecology and palaeosciences.

Keywords FTIR microspectroscopy · Raman microspectroscopy · Pinales · Imaging · Cell wall

✉ Boris Zimmermann
boris.zimmermann@nmbu.no

Murat Bağcıoğlu
murat.bagcioglu@nmbu.no

Christophe Sandt
christophe.sandt@synchrotron-soleil.fr

Achim Kohler
achim.kohler@nmbu.no

[1] Department of Mathematical Sciences and Technology, Faculty of Environmental Science and Technology, Norwegian University of Life Sciences, Drøbakveien 31, 1430 Ås, Norway

[2] Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin, BP 48, 91192 Gif-sur-Yvette, France

[3] Nofima AS, Osloveien 1, 1430 Ås, Norway

Introduction

Fourier transform infrared (FTIR) and Raman spectroscopy are powerful techniques for studying inorganic, organic and biological samples. During recent years, they have become very popular for the analysis of plants. The two techniques, commonly referred to as vibrational or molecular spectroscopy, offer chemical characterization of plant samples via identifiable spectral features. A vibrational spectrum of a plant sample contains specific

signatures of the constituent compounds, such as water, lipids, proteins, carbohydrates, pigments and complex biopolymers (Schulz and Baranska 2007). A major advantage of these techniques is that they can provide economical high-throughput analysis of biological samples without any chemical pre-treatment.

In vibrational microspectroscopy, infrared and Raman spectrometers are combined with a microscope to measure spectra of samples on a microscopic level. In microspectroscopy, spatially resolved measurements of a sample are obtained either by point-by-point measurements employing a single element detector ('mapping') or by the simultaneous measurement of the whole area by a focal plane array detector ('imaging'). The obtained image is often referred to as hyperspectral image since it contains information on hundreds of spectral channels (wavelengths). The image enables chemical interpretation of spatially resolved spectral signals. Thus, based on these specific signals, chemical images with spatial distribution of chemical components in the sample can be obtained. The spatial resolution of FTIR and Raman microspectroscopy is defined by the diffraction limit, which depends on the wavelength of the light of the source and the objective used. In Raman measurements, the lateral resolution is typically within the range of 0.3–1.0 μm, which is comparable to microscopes that are using light in the visible range. In FTIR spectroscopy, the diffraction limit is within the range of 2–20 μm due to the long wavelength of mid-infrared light (Lasch and Naumann 2006). However, such spatial resolution is rarely obtained with conventional IR sources and thus the typical lateral resolution is within the range of 20–50 μm. Regarding that plant cells are on average some of the largest eukaryotic cells, in general both techniques have sufficient spatial resolution for the analysis of plant cells and tissues. FTIR and Raman have been applied extensively on various plant structures, such as cell walls, seeds, and leaves, for research on plant physiology, developmental biology, genetics and ecology (Gorzsas et al. 2011; Barron et al. 2005; Dokken et al. 2005; Chen et al. 2013; Yu 2011; Gierlinger and Schwanninger 2006; Chylińska et al. 2014; Agarwal 2006; Fackler and Thygesen 2013; Gierlinger et al. 2012).

Plant pollen is microscopic and, therefore, its characterization and analysis has been almost exclusively performed within the domain of optical and electron microscopy. Pollen grains are multicellular male gametophytes of seed plants, that have vegetative and generative cells enclosed in a highly resistant cell wall. Pollen grains of various species can vary quite a lot in grain shape and size, number, type and position of apertures, and structure and sculpture of the grain wall. In the last decade, vibrational spectroscopy of pollen has been performed in macroscopic mode on bulk samples, resulting in the clear identification of main chemical constituents (Zimmermann and Kohler 2014; Zimmermann 2010; Parodi et al. 2013; Pappas et al. 2003; Gottardini et al. 2007; Schulte et al. 2008, 2009, 2010; Ivleva et al. 2005).

Unfortunately, although microspectroscopy of pollen appears as a natural extension of vibrational studies, the measurements are hampered by technical difficulties. Raman measurements of pollen are hindered by laser-induced degradation, as well as by the strong fluorescence background that often masks any underlying Raman spectra (Laucks et al. 2000; Boyain-Goitia et al. 2003; Ivleva et al. 2005). The problem of strong fluorescence in the Raman spectra of pollen can be circumvented by photobleaching, i.e., photodepletion of the carotenoid molecules (Ivleva et al. 2005; Schulte et al. 2008, 2009). However, the process not only destructs the sample partially, it also significantly increases the spectral acquisition time. An alternative solution is the application of FT-Raman spectroscopy employing a near infrared (NIR, 1064 nm) laser excitation (Zimmermann 2010). Unfortunately, the use of a near infrared laser excitation in FT-Raman spectroscopy results in a lower spatial resolution and sensitivity compared to dispersive Raman where lasers with wavelengths in the visible region are employed. In FTIR microspectroscopy of pollen we encounter strong scattering, since pollen are due to their size and shape highly scattering samples in the infrared. The strong light scattering results in anomalous spectral features that can significantly interfere with and distort the signals of chemical absorption (Lukacs et al. 2015; Dell'Anna et al. 2009). Since scatter issues are a direct result of pollen size and morphology, the same problems can be expected in FTIR measurements of other microorganisms that are enclosed in a hard shell, such as plant and fungal spores and diatom algae. Despite these difficulties, vibrational spectroscopy of pollen has been providing valuable new data, and can thus serve as a powerful biophysical technique particularly in environmental and plant development studies (Zimmermann and Kohler 2014; Schulte et al. 2010).

In the paper at hand, we present the study on chemical imaging and characterization of pollen grains by FTIR microspectroscopy brining pollen substructures and their chemical components, such as grain wall biopolymers, into focus. This is of importance since the concentration of phenylpropanoid components in wall biopolymers can be used as UV-B proxy and changes in the flux of UV-B radiation can be tracked over geological time (Lomax et al. 2012). Therefore, the aim of this study was to investigate to what extend FTIR and Raman microspectroscopy measurements can reveal spatially resolved chemical information of pure grain biopolymers of Pinaceae pollen. This would facilitate the analysis of these important compounds, which according to state-of-the-art technology are analyzed

following harsh chemical pre-treatment, which could alter their chemical composition (Wehling et al. 1989; Ahlers et al. 1999; Jungfermann et al. 1997). The study covers single grain FTIR measurements employing both globar and synchrotron light sources. The level of chemical information that can be obtained by spatially resolved FTIR spectroscopy is discussed and compared to Raman microspectroscopic measurements. Different methodologies for obtaining vibrational microspectroscopic data are discussed together with methods for multivariate analysis of the microspectroscopic data. The potential of vibrational imaging as a tool for studying pollen physiology is elucidated.

Results

FTIR measurements of intact pollen grains

Spectral anomalies hinder spectral reproducibility

In infrared microspectroscopy, pollen, cells and tissues are highly scattering samples since the sizes of samples are at the same order as the wavelength range employed in the infrared microspectroscopic measurements (Lukacs et al. 2015; Bassan et al. 2010a, b). For instance, Cupressaceae pollen grains have a uniform spheroidal morphology with diameters varying between 15 µm (*Juniperus*) to 35 µm (*Cunninghamia*), while the wavelength range of the mid-infrared light employed in the FTIR spectrometers is approx. 2.5–25.0 µm. Thus, spectra of single pollen grains, presented in Fig. 1, show substantial and complex anomalies. Compared with the spectra of bulk samples (multigrain samples), the single grain spectra of *Juniperus chinensis* have extensive anomalous signals. The strongest scattering anomalies were obtained in the microspectroscopy of single *Juniperus* grains while the weakest ones were obtained for *Cunninghamia* grains. This is as expected, since the sizes of *Juniperus* grains are exactly within the magnitude of the mid-IR light (Fig. 1). Since these anomalies origin from morphological and optical properties of the samples, infrared microspectroscopy of single grains having different morphological and/or optical properties, result in general in spectra with different scatter distortions. Since the same species can exhibit a variety of morphological characteristics, infrared spectra of single pollen grains from the same species, are not reproducible (Fig. 1c). The anomalous signals are distorting chemical absorbance signals, thus significantly hampering characterization of the sample. An additional problem is that the absorbance values of single pollen grains are often at the detection limit, as defined by signal-to-noise ratio. This means that conventional light sources (globar) cannot provide quality spectra for small pollen grains, such is the case for measured species of *Juniperus* (Fig. 1).

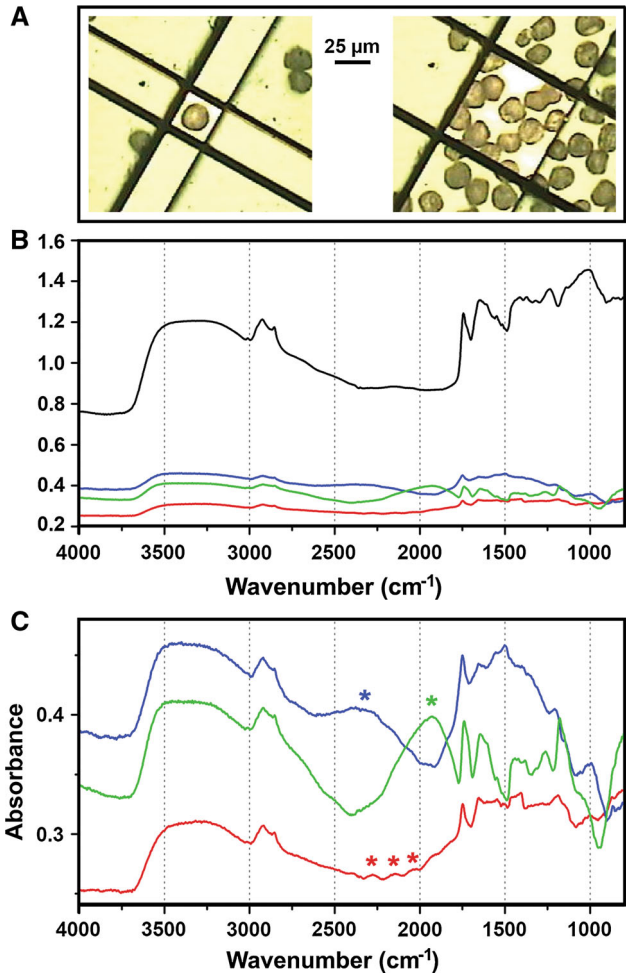


Fig. 1 Scattering anomalies in infrared spectra of pollen grains. **a** Optical microscope images of the measured *Juniperus chinensis* pollen grains on 3 mm ZnSe slide, with a 25 × 25 µm aperture for single grain measurement (*left*) and a 75 × 75 µm aperture for multigrain measurement (*right*). **b** Globar µFTIR spectra of *Juniperus chinensis*: multigrain spectrum (*black*), and single grain spectra of different grains (*blue*, *green* and *red*). **c** Enlarged single grain spectra from Fig. 1b; spectral artifacts due to scattering are designated with *asterisk*

Spatial oriented pollen samples hinder reproducibility

As opposed to Cupressaceae pollen grains, grains of Pinaceae have significantly larger diameters, ranging from 60 µm (*Podocarpus*) to 120 µm (*Abies*). In addition, their morphology is very complex: Pinaceae are described as saccate pollen due to a large hollow projection (saccus) from the central body of pollen grain (corpus). All measured Pinaceae species, as well as the Podocarpaceae species, have two sacci. The saccus regions, that can be easily depicted in the equatorial view (grain *a* in Fig. 2a),

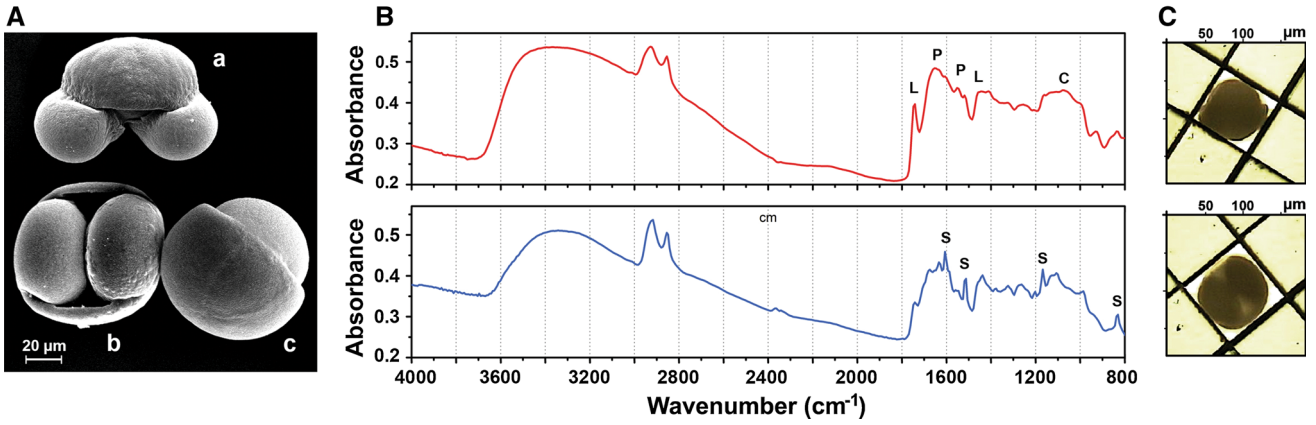


Fig. 2 Influence of spatial orientation of pollen samples on their infrared spectra. **a** SEM image of the intact *Abies cephalonica* pollen grains in: *a* equatorial front view, *b* distal polar view, and *c* equatorial profile view. **b** Globar µFTIR spectra of single pollen grains with equatorial profile orientation (*red*), and distal polar orientation (*blue*); the marked bands are associated with lipids (L), proteins (P), carbohydrates (C) and sporopollenins (S). **c** Corresponding optical microscope images of the measured pollen grains on 3 mm ZnSe slide, with depicted 100 × 100 µm aperture

are composed of sporopollenins only, a complex biopolymer. On the other hand, the corpus region contains nucleic acids and proteins, as well as nutrients in the form of lipids and sugars. As can be seen in Fig. 2b, measurements of Pinaceae single grains result in high-quality IR spectra that are devoid of scattering signals. However, we observe, that Pinaceae spectra are not reproducible due to differences in chemical absorption bands. The reason is that, Pinaceae samples do not have pseudo-radial symmetry as Cupressaceae grains, but rather distinct spatial orientations due to bilateral symmetry. The different orientations, corresponding to the infrared spectra in Fig. 2b, are shown in Fig. 2c (optical microscopy) and a (SEM measurement). It can be seen that the spectrum of the grain that has equatorial profile orientation has strong signals of lipids and proteins (Fig. 2b). Lipids are characterized by the strong vibrational band at 1745 cm^{-1} (C=O stretch), as well as by weaker bands at 1462 cm^{-1} (CH_2 deformation) (Gottardini et al. 2007; Pappas et al. 2004; Zimmermann 2010; Zimmermann and Kohler 2014). Proteins are characterized by two strong and broad bands at 1650 cm^{-1} (amide I: C=O stretch) and 1550 cm^{-1} (amide II: NH deformation and C–N stretch), while carbohydrates have strong absorption in 1200–900 cm^{-1} region (C–O–C and C–OH stretch) (Gottardini et al. 2007; Pappas et al. 2004; Schulte et al. 2008; Zimmermann 2010; Zimmermann and Kohler 2014). On the other hand, the spectrum of the grain that has distal polar orientation has strong signals of sporopollenins. The sporopollenin bands at 1605, 1515, 1171 and 833 cm^{-1} can be associated with the vibrations of aromatic rings (Zimmermann 2010; Schulte et al. 2008; Zimmermann and Kohler 2014). The reason for this discrepancy in spectral signatures of pollen for the two orientations is the spatial arrangement of the main biochemicals in Pinaceae pollen grains. While the saccus regions are predominantly composed of sporopollenins, the corpus region is mainly composed of proteins, lipids and carbohydrates.

FTIR measurements of pollen substructures

For bigger pollen grains, even substructures on pollen grains can be measured by FTIR microspectroscopy. For instance, Pinaceae grains have sufficient sizes for more detailed measurements of grain substructures even by conventional IR light sources (globar). By choosing the desired aperture, IR microspectroscopy enables spatially resolved measurements of saccus and corpus substructures (Fig. 3a). The microspectroscopy data reveals the substantial biochemical differences between related Pinaceae genera *Cedrus* and *Abies*. The measurements show elevated relative lipid content in pollen interior belonging to *Cedrus* species, as compared to *Abies* (and other Pinaceae) species (Fig. 3b). More intriguing, though, is the extreme differences in biochemical composition of the grain wall (Fig. 3c). While *Abies* species show spectral features that are characteristic for the rest of the Pinaceae family (*Pinus* and *Picea* genera), *Cedrus* species display significant chemical deviation from this common wall composition. The difference indicates different chemical composition of sporopollenin biopolymers that build the exine layer of grain wall. The grain wall spectrum of *C. atlantica* shows bands at 852 and 815 cm^{-1}, which are both associated with ring vibrations of phenylpropanoid building blocks in sporopollenin (Schulte et al. 2008). The presence of these infrared bands indicates that *C. atlantica* has significant amount of ferulic building blocks in sporopollenins. In contrast, other measured Pinales pollens have sporopollenins composed almost exclusively of *p*-coumaric

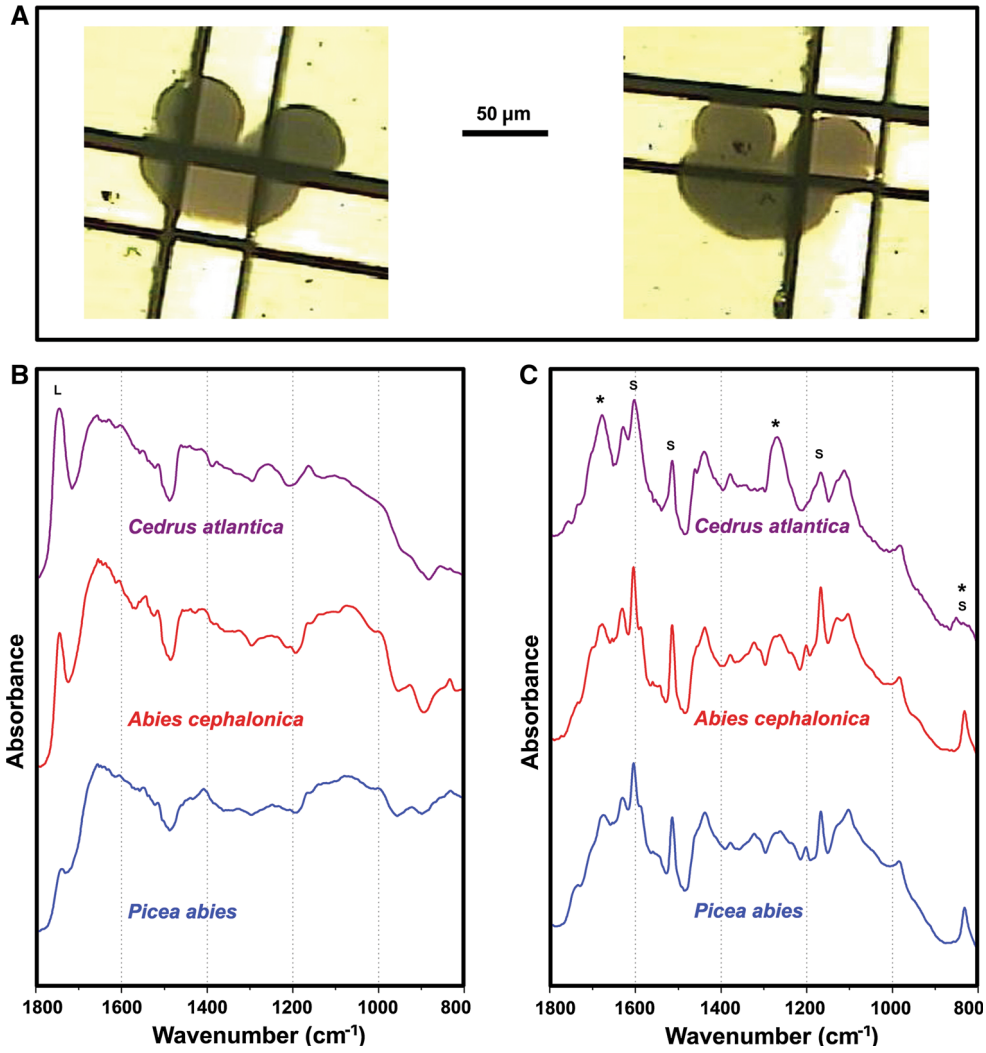


Fig. 3 Measurement of infrared spectra of ultrastructure of pollen grain. **a** Optical microscope images of the measured *Abies cephalonica* pollen grain substructures, corpus (*left*) and saccus (*right*), on 3 mm ZnSe slide, with depicted 40 × 40 µm aperture. **b** Globar µFTIR spectra of corpus regions, and **c** saccus regions of *Cedrus atlantica*, *Abies cephalonica*, and *Picea abies*. For better viewing the spectra are offset. The marked bands are associated with lipids (L) and sporopollenins (S); the spectral regions of interest are denoted with *asterisks*

building blocks, as indicated by the strong band at 833 cm^{-1}. Phenylpropanoids play important role in mitigating damage caused by UV-B radiation, and thus their direct measurement from pollen grains, as described here, could be used to monitor effect of solar radiation on conifers (Blokker et al. 2006).

High-spatial resolution spectra of pollen substructures by synchrotron infrared spectroscopy

Synchrotron radiation light sources allow a more precise FTIR measurement of Pinaceae grains, such as line and image maps. As an example we show the line map of an intact *Abies cephalonica* pollen grain along the equatorial transect in Fig. 4. As can be seen, the chemical composition as well as the total absorption is changing substantially when moving from the hollow saccus region to the denser corpus region. Two phenomena can be observed when moving to the denser corpus region (red spectrum in Fig. 4b): (1) the scatter contributions increase considerable, and (2) the optical path length (thickness) of the sample becomes so high that the major part of the infrared light is completely absorbed. The spatially highly resolved synchrotron infrared spectra show that the central area of the corpus part of the grain is opaque for a large part of the infrared spectrum of the light. The small fraction of light that reaches the detector has passed through the periphery areas of corpus part, and thus cannot provide objective information about the grain interior. This allows us to draw an important conclusion for the spectra obtained from single Pinaceae pollen grains by conventional infrared microspectroscopy using globar sources as presented in Fig. 2: we expect that the globar measurements of the whole Pinaceae pollen grains presented above, are only in a little degree containing information about the corpus region. Spectral information is predominantly obtained from the saccus regions, since the saccus regions are transmitting more light than corpus region. This also

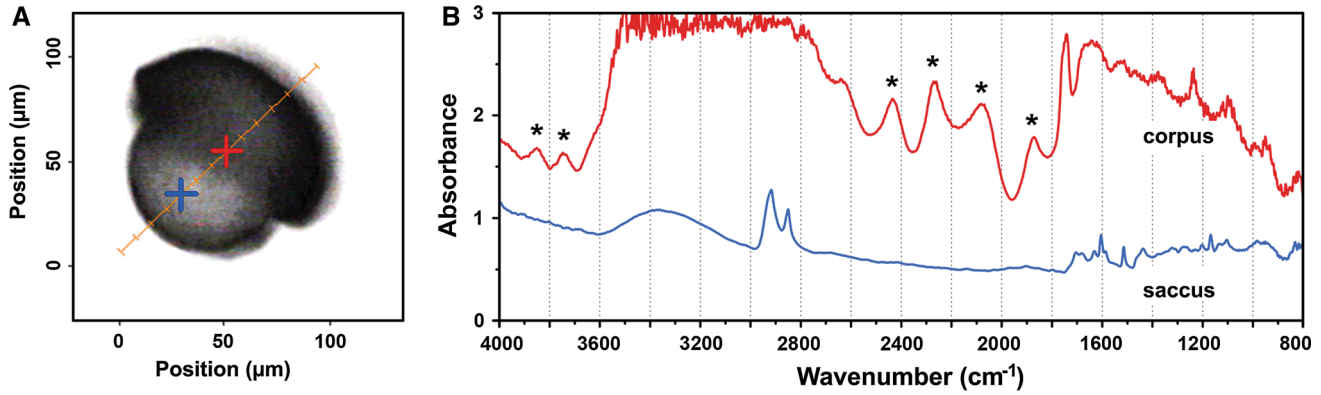


Fig. 4 Linear mapping of pollen grain by FTIR microspectroscopy. **a** Optical microscope image of the intact *Abies cephalonica* pollen grain in equatorial profile view, with positions of spectra acquisition: saccus (*blue*) and corpus regions (*red*). **b** Synchrotron µFTIR spectra, obtained with 20 × 20 µm aperture, of saccus (*blue*) and corpus regions (*red*) from Fig. 1a; spectral artifacts due to scattering are designated with *asterisk*

explains why the single grain spectra of Pinaceae pollen (see Fig. 2) are highly dependent on the grain spatial orientation during the measurement. The saturation effects revealed by synchrotron infrared spectroscopy with high brilliance employing small apertures, must be valid also for conventional infrared microscopy employing globar sources.

Vibrational measurements of sectioned pollen grains

As demonstrated above, large pollen grains, such as those of the family Pinaceae, are totally absorbing in the IR region. Thus, in order to obtain infrared spectra of such and similar-sized single pollen grains, the grains were sectioned to provide shorter optical path lengths. Following this strategy, FTIR images were obtained, where an image with a full FTIR spectrum in each pixel was obtained by mapping a whole image plane (Fig. 5) within 10 µm sections of Pinaceae grains embedded in optical cutting temperature compound (OCT) (Fig. 5). FTIR images, thus, represent large data sets, which can be displayed as pseudo color chemical images and associated with molecular composition of the sample. Chemical images are constructed by plotting intensity values of wavelength of interest (such as maxima of amide band), or alternatively by plotting ratio of band areas of signals of interest (such as ratio of amide band and sporopollenins ring vibration band in Fig. 5c, d). Construction of chemical image requires data processing, such as extended multiplicative signal correction (EMSC), for separation of chemical and physical spectral variations. For example, spectral variations can be caused by differences in the sample thickness, refractive index and other physical properties.

In addition to the infrared chemical images, Raman chemical images where obtained (Fig. 5). The quality of the Raman spectra obtained from thin sections was considerably better than previously obtained Raman spectra of intact grains (Zimmermann 2010): Since Raman spectroscopy is a scattering technique, spectra of intact grains show mainly signals from pollen grain wall components (Zimmermann 2010), while the Raman spectra of sections reveal molecular information about substructures of the cell interior. Another advantage of Raman spectroscopy of thin sections as opposed to whole grains was that grain sections suppressed strong fluorescence backgrounds in the Raman that may arise due to aromatic and polyenic compounds in grain wall.

It is worth noting that FTIR and Raman spectroscopy differ considerably in the spatial resolution they provide. For instance, in the µRaman measurement shown in Fig. 5, the acquisition laser light was focused on an area of 1 µm diameter, while the step size of motorized microscope stage in x and y directions was 2 µm. In the corresponding µFTIR measurement of the same sample, the synchrotron light was focused on an area of 10 × 10 µm due to resolution limitations. However, the small step size of microscope stage with 5 µm resulted in spatial oversampling and has improved the resolution. Regarding the nominal resolution, the µRaman measurement in Fig. 5c, has approximately seven times better nominal resolution than the µFTIR measurement in Fig. 5d, e (307 pixels per 1000 µm^2 in µRaman, versus 46 pixels per 1000 µm^2 in µFTIR).

As a general rule, the two methods obtain similar but complementary chemical information, since strong bands in the IR spectrum of a sample often correspond to weak bands in the Raman and vice versa. This is exactly what we observe in the vibrational spectroscopy of pollen, where strong Raman signals of sporopollenins are complemented by week signals in the infrared, while week Raman signals of lipids and proteins are countered by strong infrared signals of these compounds.

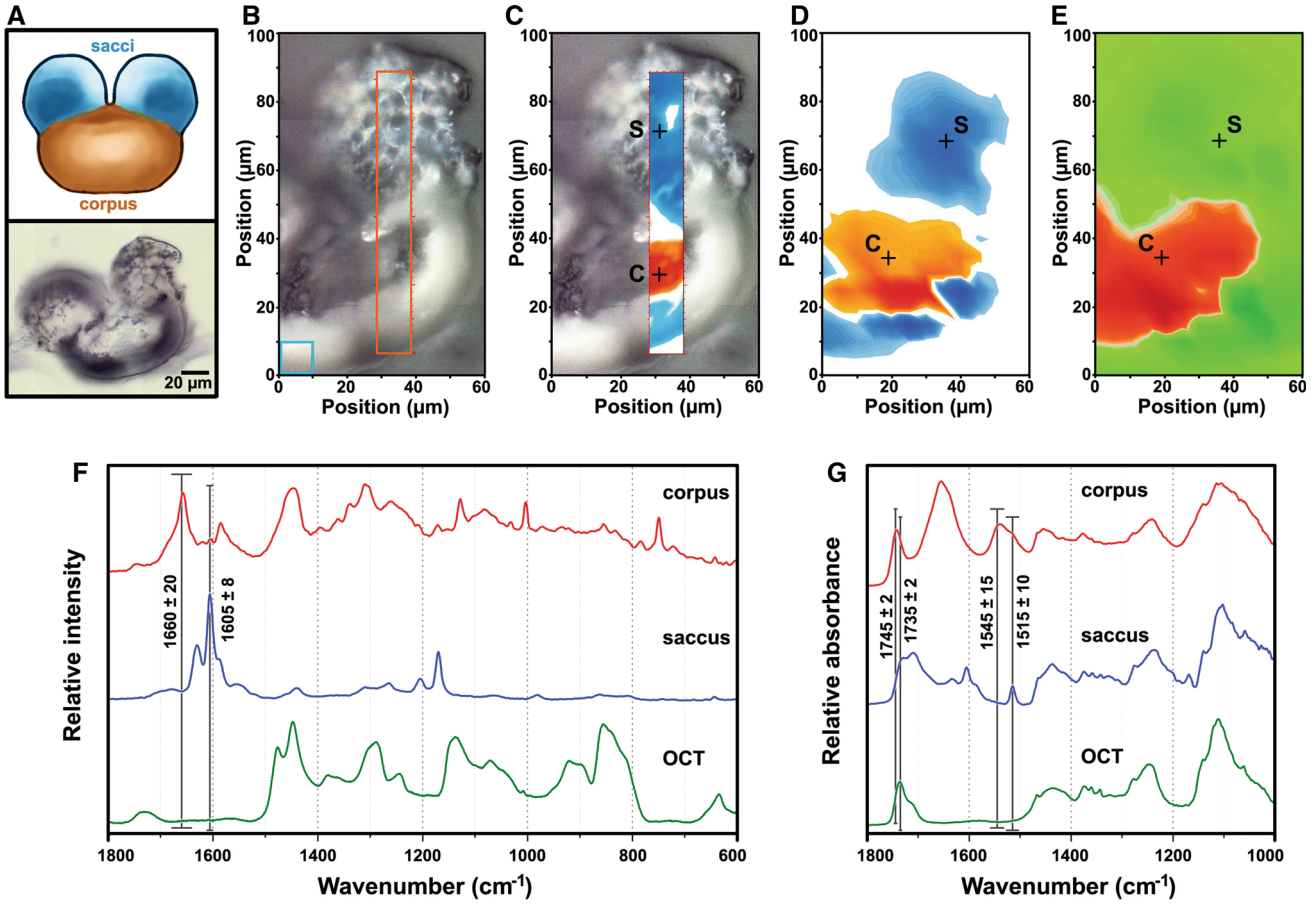


Fig. 5 Chemical images of pollen grain obtained by FTIR and Raman microspectroscopy. **a** Sketch in equatorial front view of a bisaccate Pinaceae pollen grain (*up*), and optical microscope image of the 10 µm section of *Abies cephalonica* pollen grain embedded in OCT on 1 mm ZnS slide (*down*). **b** Optical microscope image of the pollen grain showing the area (60 × 100 µm) for synchrotron µFTIR imaging (blue square represents 10 × 10 µm aperture; 5 µm FTIR step size); the area (10 × 82 µm) covered by µRaman imaging is depicted by orange rectangle (2 µm Raman step size). **c** µRaman image (overlaid on visible image) generated by the ratio of the areas under the amide I protein band (1660 ± 20 cm^{-1}) and the sporopollenin band (1605 ± 8 cm^{-1}); high ratio is depicted by *red* and low ratio by *blue colors* (see position of bands in Fig. 1f). **d** µFTIR image generated by the ratio of the areas under the amide II protein band (1545 ± 15 cm^{-1}) and the sporopollenin band (1515 ± 10 cm^{-1}); high ratio is depicted by *red* (high protein content) and low ratio by *blue colors* (high sporopollenin content); see position of bands in Fig. 5g. **e** µFTIR image generated by the ratio of the areas under the carbonyl lipid band (1745 ± 2 cm^{-1}) and the carbonyl OCT band (1735 ± 2 cm^{-1}); high ratio is depicted by *red* (high lipid content) and low ratio by *green colors* (high OCT content); see position of bands in Fig. 5g. **f** EMSC corrected Raman spectra of corpus and saccus pollen grain substructures (designated on Fig. 5c by C and S, respectively), and of OCT (see ratio of bands in Fig. 5c). **g** EMSC corrected IR spectra of corpus and saccus pollen grain substructures (designated on Fig. 5d, e by C and S, respectively), and of OCT (see ratios of bands in Fig. 5d, e). For better viewing the Raman and IR spectra are offset

The Raman chemical image in Fig. 5c shows clear spatial separation of sporopollenin constituents, characterized by the peak at 1605 cm^{-1} (aromatic ring stretches), and proteins, characterized by the peak at 1660 cm^{-1} (amide I band) (Schulte et al. 2008; Zimmermann 2010). Our results are corroborating findings obtained in the previous Raman chemical imaging, obtained on pollen of rye (*Secale cereal*) (Schulte et al. 2008). Due to larger radius and different symmetry, *Abies* pollen, compared to *Secale* pollen, shows clearer distinction between nutrient-rich regions (corpus) and sporopollenin-rich regions (grain wall, including saccus). However, the results are difficult to compare since the two species belong to far related plant lineages with significant differences in chemical composition of pollen (Zimmermann and Kohler 2014).

A special attention has to be paid to the potential presence of signals belonging to the embedding OCT material in the measured Raman and FTIR spectra. Raman spectra of OCT exhibit week signals, and thus OCT signals are not interfering significantly with Raman signals of chemical constituents of pollen (Fig. 5f). In contrast, OCT has strong absorption bands in the infrared. Hence, OCT signals are present in the spectra of saccus and corpus regions. Although infrared signals of OCT interfere with the signals of

sporopollenins, lipids and proteins, there is still a sufficient number of specific peaks for the separation of information related to chemical pollen constituents. This is illustrated by the infrared chemical image presented in Fig. 5d. It was obtained by calculating the ratio of the protein band at 1545 cm^{-1} (amide II) and the sporopollenin band at 1515 cm^{-1} (aromatic ring vibration) (Zimmermann 2010; Zimmermann and Kohler 2014). The FTIR chemical image in Fig. 5d is complementary to Raman chemical image in Fig. 5c. In addition, an FTIR image was obtained by the ratio of the lipid band at 1745 cm^{-1} (carbonyl stretch) and the OCT band at 1735 cm^{-1} (carbonyl stretch) illustrating the spatial distribution of lipids (Fig. 5e).

Imaging and sample morphology

Vibrational imaging measurements are often affected by a number of unwanted phenomena, such as interfering effects of undesired physical variations. Regarding FTIR measurement of thin pollen sections (Fig. 5), there are two main sources of physical variations: changes in thickness within the section lead to variation in the effective optical path length, while chemical and physical inhomogeneity within the section lead to variations in the optical properties such as refraction and scattering. EMSC pre-processing enables estimation and separation of the undesired physical variations from the chemical information in the spectral data set (Kohler et al. 2005). For example, the chemical images in Fig. 5 were obtained after pre-processing with EMSC. In the EMSC model the unwanted physical information is explicitly parameterized (see Eq. 1), thus enabling detailed analysis of physical properties of a measured sample. Regarding that pollen grains have interesting physical properties, characterized by complex morphology of pollen ultrastructure, analysis of EMSC parameters offers an alternative approach for imaging of pollen grains.

Figure 6 illustrates the separation of physical and chemical information in FTIR images of pollen. The separation was obtained by using an EMSC model with constant, linear and quadratic parameters. The EMSC separates the recorded (original) spectrum (Fig. 6a) into a "corrected chemical spectrum" including chemical absorption bands (black spectrum in Fig. 6b), and physical information described by EMSC parameters (blue and red lines in Fig. 6b).

As can be seen from Fig. 6a, b the absorbance values of saccus and corpus regions differ substantially, with significantly weaker absorbance in the saccus region. The reason is aforementioned strong vibrational bands associated with lipids (carbonyl stretch) and proteins (amide I and amide II bands). On the other hand, sporopollenins are devoid of strong bonds, although they have characteristic bands at 1605 and 1515 cm^{-1}. The difference in total absorbance between the two regions is mirrored by the difference in multiplicative parameters (b_i in Eq. 2) for the two regions: lower values of b for the saccus region and higher values for the corpus region. Moreover, the spectra belonging to different regions differ by other calculated EMSC parameters as well. For example, values for constant parameters (a_i in Eq. 2) have high positive values for the corpus region, low positive values for saccus region, and high negative values for the empty (OTC-only) region (Fig. 6d). In addition, values for the quadratic parameter (e_i in Eq. 2) have negative values for corpus region, and positive values for saccus region (Fig. 6e). These parameters correspond to baseline variations and may be explained by scattering phenomena. Thus, it is possible to construct scattering image of pollen's substructures by using only physical information, as described by EMSC parameters, and obtaining an image of morphological properties of substructures.

PCA imaging of pollen

Principal component analysis (PCA) is a multivariate data analysis method and is often the first method carried out on a multivariate spectral data set, such as vibrational spectra, for explorative data analysis. Since PCA allows interpretation based on all variables simultaneously, it enables more comprehensive visualization of spectral data. PCA transforms a spectral set from original variables (wavenumbers) into latent variables (principal components). The first component maximizes the co-variance in variable and sample space. Then, this first principal component is removed from the data set and the second component is found by maximizing the remaining co-variance and so on. The obtained principal components are uncorrelated. Typically, analysis of first few principal components is sufficient for obtaining a deeper understanding of the main variation patters that are present in spectral data since they capture the predominant part of the variance in the data. The relationship between original and latent variables is visualized by loading plots, which show how the different wavelengths contribute to the PCA model.

A PCA of a spectral image is shown in Fig. 7. As can be seen from Fig. 7b, c, PC scores can be associated with different chemical constituents of pollen grain. The loadings plot of the PC1 (Fig. 7d) has high negative values for signals associated with lipids (1745 cm^{-1}; carbonyl stretch) and proteins (1650 and 1550 cm^{-1}; amide I and amide II, respectively), and positive signals associated with the embedding OCT material (1500–1200 cm^{-1} region). In the PC1 image (Fig. 7b), the corpus region is characterized by high negative PC1 values, thus indicating that this region has a higher content of lipids and proteins. On the other hand, the loadings plot of the PC2 (Fig. 7d) shows high positive values for signals associated with sporopollenins (1605, 1515 and

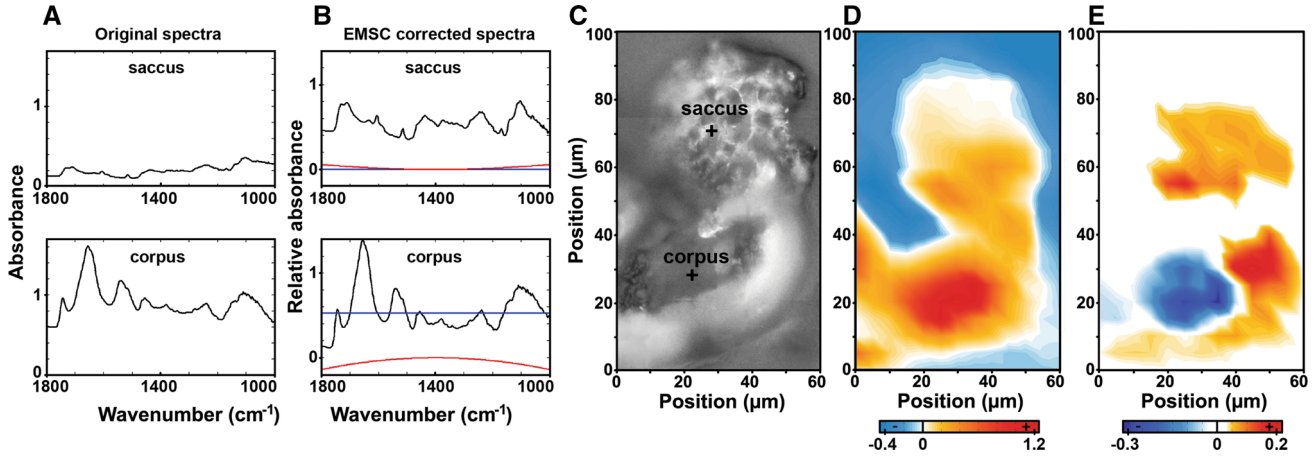


Fig. 6 Images obtained by separating chemical from physical information in the infrared spectra of *Abies cephalonica* pollen grain. **a** Original and **b** EMSC corrected IR spectra of corpus and saccus pollen grain substructures (designated on Fig. 5c); baseline (*a*) and quadratic (*e*) parameters are represented by *blue* and *red colors*, respectively. **c** Optical microscope image showing the area (60 × 100 µm) for µFTIR synchrotron imaging and the position of Fig. 2a, b spectra. Images of scattering properties: **d** constant baseline (*a*), and **e** quadratic baseline (*e*); negative and positive values are depicted by *blue* and *red colors*, respectively

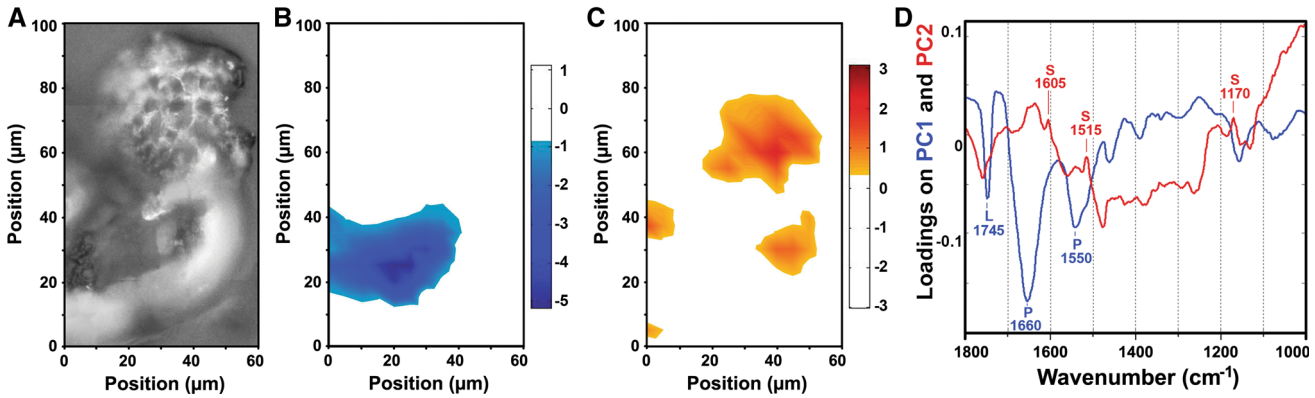


Fig. 7 Images obtained by principal component analysis of infrared spectral data. **a** Optical microscope image showing the area (60 × 100 µm) for µFTIR synchrotron imaging of the 10 µm section of *Abies cephalonica* pollen grain embedded in OCT on 1 mm ZnS slide. **b** Image of PC1 scores; high negative values are depicted by *blue color*. **c** Image of PC2 scores; high positive values are depicted by *red color*. **d** Loadings plot on the first two principal components of the PCA; the marked signals are associated with lipids (L), proteins (P) and sporopollenins (S). The percent variances for the first five PCs are 75.7, 9.8, 5.2, 2.5, and 1.9 %

1170 cm^{-1}; aromatic rings vibrations), and negative signals associated with OCT. In the PC2 image (Fig. 7c), the saccus region is characterized by high positive PC2 values; thus, indicating that this region has higher content of sporopollenins. Therefore, pseudo-chemical images of pollen grain sections can be obtained by using PCA information.

Discussion

Advantages and limitations of vibrational microspectroscopies of pollen

Different chemical constituents of pollen have different absorption coefficients, and thus different sub-regions of pollen have different propagation depths of infrared light. For example, proteins and lipids have higher absorption in the infrared region than grain wall components such as sporopollenins and cellulose. Therefore, grain regions with high concentration of grain wall components (such as saccus of Pinaceae grains) transmit more infrared light than grain regions rich in lipid nutrients (such as corpus of Pinaceae grains). This difference in infrared light absorption between different chemicals is mostly irrelevant for pollen grains with radial symmetry. However, for grains with bilateral symmetry, such as saccate pollen of Pinaceae family, it results in single grain spectra that are strongly dependent on grain spatial orientation (Fig. 2).

The biggest challenge in FTIR measurement of single pollen grains is the presence of anomalous signals due to

scattering (Figs. 1, 4). These signals are a major obstacle in the chemical interpretation of infrared spectra of single pollen grains. Spectral signatures of single pollen grains show strong Mie scattering due to their nearly perfect spherical shape (Lukacs et al. 2015). In comparison, nearly perfect absorbance spectra have been obtained, when plant materials have been probed by FTIR microspectroscopy as thin sections (Fackler and Thygesen 2013; Yu 2011). Mie scattering signatures are difficult to remove by spectral pre-processing methods. In an infrared spectrum, scattering and absorption signatures are highly entangled. The real and the imaginary part of the refractive index, that describe the optical and absorptive properties of the respective material, are connected through the Kramers–Kronig relation (Bassan et al. 2010a). This connection entangles the signatures of physical properties of the material and the chemical properties of the material in the measured absorbance spectrum. While the scatter signatures can give valuable information about size and density of pollen, the absorption signatures constitute a fingerprint of the biochemical components. Unfortunately, analytical solutions for complex structures, such as pollen grains, that would offer separation of scatter and absorbance components have not yet been developed (Lukacs et al. 2015). Therefore, spectra obtained from highly scattering samples contain strongly entangled chemical and physical information hampering the interpretation of the spectral signals.

Raman measurements of intact grains result in excessive representation of an external part of a grain due to a greater probability of multiplicative scattering in the grain interior (Zimmermann 2010). Therefore, Raman measurements of intact grains have disproportional bias toward pollen grain wall. For large pollen grains with thick grain walls, such as Pinaceae, it is unlikely that any constituent from the grain interior can be manifested in a spectrum. However, by sectioning a grain, pollen grain composition of the corpus region interior can be measured in detail (Fig. 5c).

Spectral pre-processing as a pre-requisite for chemical imaging

Light scattering is directly related to morphology, which can vary substantially in shape, size, and texture. Such morphology variation is typical for the measured Pinaceae samples: Saccus region of Pinaceae is characterized by sponge-like morphology of significantly lower density than one found within densely packed corpus region. Therefore, scattering features may serve as valuable information source for characterization of pollen, alongside chemical information. Figure 6 shows the separation of physical and chemical information in an FTIR image of pollen by using an EMSC model. The EMSC parameters are used to estimate non-chemical absorption, reflection and scattering of light that is caused by irregularities within the pollen section. The pseudo-structural images, based on the representation of EMSC parameters (Fig. 6d, e), clearly show that their values are not random but rather specific for pollen ultrastructure. Therefore, alongside measurement of chemical composition of pollen ultrastructure, the FTIR methodology offers indirect estimation of physical properties. The indirect estimation of physical properties is not possible by filter-based pre-processing methods such as standard normal variate (SNV) (Barnes et al. 1989). EMSC, the model-based pre-processing method, allows the estimation of physical parameters, before removing them from spectra. While Multiplicative Signal Correction (MSC) only estimates a constant baseline (Ilari et al. 1988), EMSC in its basic form estimates in addition a linear and non-linear baseline. Compared to SNV and MSC, EMSC enables correction of wavelength-dependent scattering effects and thus provides better separation of physical and chemical information in vibrational spectra. The correlation of non-constant baseline variations with the physical properties of the sample is clearly demonstrated in Fig. 6e, where the quadratic baseline parameter reflects pollen ultrastructure by separating saccus and corpus regions.

Alongside EMSC-based imaging, PCA-based imaging of spectral data offers yet another rapid estimate of pollen chemical composition. As can be seen from Fig. 7, the most significant chemical information on pollen sections can be obtained by using PCA data analysis. The first two principal components are associated with lipids and proteins, as well as with sporopollenins (Fig. 7d). The PCA-based pseudo-chemical images of pollen grain sections show clear spatial distribution of these principal pollen constituents in either corpus or saccus substructures. The major value of such approach is significant reduction of data set by two orders of magnitude, from 400 frequency variables (wavenumbers) to only two latent variables (PCs 1 and 2) per spatial variable (measurement point).

Vibrational imaging as a tool for studying ultrastructure and physiology of pollen

The comparative analysis of Pinaceae genera shows substantial biochemical differences in the composition of grain wall, indicating different chemical composition of sporopollenins. The main difference in chemical composition of sporopollenins was obtained for *Cedrus*. In addition, *Cedrus* is unusual by quite high relative amount of lipid nutrients in grain corpus region. In this regard, *Cedrus* pollen have unique chemical composition with substantial deviation from *Abies*, its closest lineage. It should be noted however that *Abies* and *Cedrus* genera have diverged early in their evolutionary history, during Jurassic ($\sim$183 Ma), approximately 30 Ma after the two branched from the rest

of Pinaceae lineage (Lin et al. 2010). Considering that *Abies* species have significantly larger grains (approx. 120 μm diameter for *Abies*, approx. 75 μm for *Cedrus*), this difference in nutrient concentration is possibly an adaptation of *Cedrus* to compensate the effect of relatively small grain size. *Cedrus* has one other unique feature connected to pollen: As opposed to other Pinaceae species that pollinate during spring, *Cedrus* shed pollen in autumn. It can be assumed that the large amount of lipid nutrients in grain interior, and the exceptional composition of grain wall, could play significant role in this unique pollination strategy. Measurement of sporopollenins can provide other valuable information as well. Grain wall sporopollenins are extremely resilient, thus preserving clear chemical fingerprint for at least several thousand years. It has been shown that this chemical fingerprint can be used for reconstruction of past climate, such as quantification of UV-B radiation flux through time using *Pinus* pollen from sediments (Willis et al. 2011). In this regard, vibrational microspectroscopy offers an alternative methodology for measurement of such chemical information, and thus has potential to significantly improve palaeoecological data.

Detailed chemical images of pollen can be obtained by cryosectioning and measurement either by Raman or synchrotron-based FTIR microspectroscopy. As demonstrated by chemical imaging of *Abies* in Fig. 5, spatial chemical information can be extracted by using characteristic bands of proteins, lipids, and sporopollenins. The FTIR chemical images were obtained even though OCT embedding matrix is not an optimal material for infrared spectroscopy (Fig. 5g). However, by carefully selecting ratios of characteristic bands it is possible to obtain clear chemical image of pollen OCT sections (Fig. 5d, e). On the other hand, OCT is suitable matrix material for Raman measurement of pollen due to weak Raman intensities of OCT bands (Fig. 5f). By measuring the same pollen section with the two vibrational microspectroscopies one can obtain complementary chemical images of pollen grain interior (Fig. 5). The acquired images, showing chemical heterogeneity of pollen interior with saccus and corpus regions, is a clear demonstration of the methods' potential for determining spatially resolved biochemical fingerprint of pollen ultrastructure.

In summary, in this paper, we presented an approach for chemical imaging of pollen grains by FTIR and Raman microspectroscopy. The main problems arising during measurements are scatter and saturation issues in FTIR, and fluorescence and penetration depth issues in Raman. Clear chemical images of pollen grain can be obtained by careful spectral pre-processing and analysis. Vibrational microspectroscopy studies can be used for measurement of an individual pollen grain, including measurement of grain ultrastructure. In theory, such studies would enable measurement of within-plant and even within-flower biochemical variations of pollen grains in botany, as well as measurements of fossilized pollen samples from sediments. Therefore, chemical characterization of individual pollen grains by vibrational spectroscopies is suggested as a valuable tool in biology, ecology and palaeosciences.

Materials and methods

Plant material

Pollen samples were obtained through fieldwork at the Botanical Garden of the Faculty of Science, University of Zagreb. In total 24 samples were collected, belonging to 12 Pinales species during 2011 and 2012 pollination seasons. The sample set included six Pinaceae species: *Abies cephalonica* Loudon (Greek fir), *Cedrus atlantica* (Endl.) Manetti ex Carrière (Atlas cedar), *Picea omorika* (Pancic) Purk. (Serbian spruce), *Picea abies* (L.) H.Karst. (Norway spruce), *Pinus resinosa* Aiton (Red pine), *Pinus sylvestris* L. (Scots pine); five Cupressaceae species: *Cunninghamia lanceolata* (Lamb.) Hook. (China fir), *Cupressus sempervirens* L. (Mediterranean cypress), *Juniperus chinensis* L. (Chinese juniper), *Juniperus communis* L. (Common juniper), *Juniperus excelsa* M.Bieb. (Greek juniper); and one Podocarpaceae species: *Podocarpus neriifolius* D.Don (Brown pine). The pollen samples were collected directly from plants at flowering time by shaking male cones (strobili). The samples were kept in paper bags at r.t. for 24 h, and afterwards transferred to vials and stored at -15 °C.

Sample preparation

Spectra of intact pollen grains were measured on ZnSe slides. Sectioned samples of pollen grains were obtained by embedding pollen in an optical cutting temperature compound (OCT, Tissue-Tek; Sakura Finetek, USA). The OCT samples were snap-frozen and sectioned at −20 °C using a cryostat (Leica CM 3050 S, Nussloch, Germany) obtaining sections of 10 μm thickness. The pollen sections were transferred to ZnS slides for measurements.

FTIR spectroscopy

FTIR microspectroscopy measurements were performed with globar and synchrotron sources. For each measured sample an image was recorded with an optical microscope.

Measurements with a globar source were performed on intact pollen grains in transmission mode with a spectral resolution of 4 cm^{-1}, using an Equinox 55 FTIR spectrometer with an IRScopeII IR microscope (Bruker

Optik GmbH, Germany). The system was equipped with a globar mid-IR source and a liquid nitrogen-cooled mercury cadmium telluride (MCT) detector. The spectra were measured in the 4000–600 cm^{-1} spectral range, with 128 scans for both background and sample spectra, using a 15× objective, with different aperture sizes, depending on the size of the sample. Background (reference) spectra were recorded immediately before starting each measurement using the sample-free setup. The Bruker system was controlled with OPUS 6.0 software (Bruker Optik GmbH, Germany).

Measurements with a synchrotron source were performed on intact and sectioned pollen grains at the SOLEIL synchrotron facility, on the SMIS infrared beamline, details of which can be found elsewhere (Dumas et al. 2006). The transmission spectra were recorded with a resolution of 4 cm^{-1} by using the synchrotron radiation coupled to a Nicolet 5700 FTIR spectrometer with a Nicolet Continuum XL IR microscope (Thermo Scientific, USA), equipped with a liquid nitrogen cooled mercury cadmium telluride detector. The spectra were measured in the 8000–650 cm^{-1} spectral range, with 128 scans for both background and sample spectra, using 15 × and 32 × objectives with different aperture sizes, depending on the size of the sample. The numerical aperture of the microscope was 0.65. Spectral maps were acquired as both linear maps and spectral images. For linear maps and images, steps sizes of 5 or 10 µm were used in a linear direction or in x and y directions, respectively. Background (reference) spectra were recorded immediately before starting each measurement using the sample-free setup. The Thermo system was controlled with OMNIC 8.1 software.

Raman spectroscopy

The Raman spectra were recorded using a DXR Raman Instrument (Thermo Fisher Scientific, Waltham, USA). Raman measurements were performed in a closed chamber using a stabilized 532 nm laser. The laser signal was generated by a Diode-pumped, solid state (DPSS) laser with 10 mW power, which was focused onto a 1.0 µm diameter spot, and measured in the spectral range of 125–3500 cm^{-1}. A 100 × objective (Thermo Fisher Scientific Waltham, USA) was used for focusing and collecting scattered Raman light. The confocal hole was set to 25 µm with a spectral resolution of 3.0 cm^{-1}. An exposure time of 2 × 10 s with 512 numbers of background exposures was used. A spectral map was acquired as spectral image, with a step size of 2 µm in x and y directions. Data acquisition and instrument control was carried out using the OMNIC 8.1 software (Thermo Fisher Scientific Inc, USA).

Data analysis

Data processing for chemical imaging was performed on the reduced wavenumber range, 1800–600 and 1800–1000 cm^{-1} for Raman and FTIR measurements, respectively. Each spectrum $\mathbf{z}_i$ (where $i = 1,\ldots, N$; N is number of spectra in an image) was pre-processed by EMSC (Martens and Stark 1991). EMSC estimates and corrects baseline effects and multiplicative effects using the model.

$$z_i = b_i\boldsymbol{m} + a_i\mathbf{1} + d_i\widetilde{\boldsymbol{\nu}} + e_i\widetilde{\boldsymbol{\nu}}^2 + \boldsymbol{\varepsilon}_i \quad (1)$$

where $\boldsymbol{m}$ is the reference spectrum (typically an average spectrum or a representative spectrum), $\boldsymbol{1}$ is a vector containing ones as entries referring to a spectrum describing a constant baseline, $\widetilde{\boldsymbol{\nu}}$ a linear spectrum and $\widetilde{\boldsymbol{\nu}}^2$ a quadratic spectrum. The term $\boldsymbol{\varepsilon}_i$ describes the residual spectrum. The parameters a_i, b_i, d_i and e_i refer to constant, multiplicative, linear, and quadratic effects, respectively, and are estimated by least squares for each spectrum.

After having estimated the parameters, the spectra are corrected according to

$$\boldsymbol{z}_{i,Corr} = (\boldsymbol{z}_i - a_i\mathbf{1} - d_i\widetilde{\boldsymbol{\nu}} - e_i\widetilde{\boldsymbol{\nu}}^2)/b_i \quad (2)$$

where $\mathbf{z}_{i,\mathrm{Corr}}$ are the corrected spectra.

Pre-processed spectra were used to evaluate biochemical similarities between pollen samples by using principal component analysis (PCA). The spectral pre-processing and data analyses were performed by algorithms developed in house in the environment of Matlab, R2014a (The Mathworks Inc., Natick, United States) and The Unscrambler X 10.3 (CAMO Process AS, Oslo, Norway).

Scanning electron microscopy

The microscope images were recorded with a Zeiss EVO 50 Extended Pressure scanning electron microscope (Carl Zeiss AG, Jena, Germany). The desiccated pollen samples were attached to SEM stubs (covered with double-stick tape) without prior pre-treatment with chemicals. The samples were coated with gold–palladium and measured with SEM.

Author contribution statement BZ and AK conceived and designed the experiments. BZ, MB and CS performed the experiments. BZ, MB and AK analyzed the data. BZ, MB, CS and AK wrote the paper.

Acknowledgments We thank Elin Ørmen from the Microscopy lab, Imaging Centre—NMBU, for the help with the SEM measurements, and Paul Dumas from the SMIS beamline, Synchrotron SOLEIL, for the help with the FTIR synchrotron measurements. The research was supported by the Norwegian Research Council (ISP project No. 216687), by the SOLEIL, French national synchrotron facility (project No. 20120345), and by the European Commission through the Seventh Framework Programme (FP7-PEOPLE-2012-IEF project No. 328289).

References

Agarwal UP (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(5):1141–1153. doi:10.1007/s00425-006-0295-z

Ahlers F, Thom I, Lambert J, Kuckuk R, Wiermann R (1999) H-1 NMR analysis of sporopollenin from Typha angustifolia. Phytochemistry 50(6):1095–1098

Barnes RJ, Dhanoa MS, Lister SJ (1989) Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl Spectrosc 43(5):772–777. doi:10.1366/0003702894202201

Barron C, Parker ML, Mills EN, Rouau X, Wilson RH (2005) FTIR imaging of wheat endosperm cell walls in situ reveals compositional and architectural heterogeneity related to grain hardness. Planta 220(5):667–677. doi:10.1007/s00425-004-1383-6

Bassan P, Kohler A, Martens H, Lee J, Byrne HJ, Dumas P, Gazi E, Brown M, Clarke N, Gardner P (2010a) Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples. Analyst 135:268–277

Bassan P, Kohler A, Martens H, Lee J, Jackson E, Lockyer N, Dumas P, Brown M, Clarke N, Gardner P (2010b) RMieS-EMSC correction for infrared spectra of biological cells: extension using full Mie theory and GPU computing. J Biophotonics 3:609–620

Blokker P, Boelen P, Broekman R, 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(1–2):197–207. doi:10.1007/s11258-005-9026-y

Boyain-Goitia AR, Beddows DCS, Griffiths BC, Telle HH (2003) Single-pollen analysis by laser-induced breakdown spectroscopy and Raman microscopy. Appl Opt 42(30):6119. doi:10.1364/ao.42.006119

Chen J, Sun S, 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(29):9385–9400. doi:10.1007/s00216-013-7366-3

Chylińska M, Szymańska-Chargot M, Zdunek A (2014) Imaging of polysaccharides in the tomato cell wall with Raman microspectroscopy. Plant Methods 10(14):1–9

Dell'Anna R, Lazzeri P, Frisanco M, Monti F, Malvezzi Campeggi F, Gottardini E, Bersani M (2009) Pollen discrimination and classification by Fourier transform infrared (FT-IR) microspectroscopy and machine learning. Anal Bioanal Chem 394(5):1443–1452. doi:10.1007/s00216-009-2794-9

Dokken KM, Davis LC, Marinkovic NS (2005) Use of infrared microspectroscopy in plant growth and development. Appl Spectrosc Rev 40(4):301–326. doi:10.1080/05704920500230898

Dumas P, Polack F, Lagarde B, Chubar O, Giorgetta JL, Lefrançois S (2006) Synchrotron infrared microscopy at the French Synchrotron Facility SOLEIL. Infrared Phys Technol 49:152–160

Fackler K, Thygesen LG (2013) Microspectroscopy as applied to the study of wood molecular structure. Wood Sci Technol 47(1):203–222. doi:10.1007/s00226-012-0516-5

Gierlinger N, Schwanninger M (2006) Chemical imaging of poplar wood cell walls by confocal Raman microscopy. Plant Physiol 140(4):1246–1254. doi:10.1104/pp.105.066993

Gierlinger N, Keplinger T, Harrington M (2012) Imaging of plant cell walls by confocal Raman microscopy. Nat Protoc 7(9):1694–1708. doi:10.1038/nprot.2012.092

Gorzsas A, Stenlund H, Persson P, Trygg J, 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. Plant J Cell Mol Biol 66(5):903–914. doi:10.1111/j.1365-313X.2011.04542.x

Gottardini E, Rossi S, Cristofolini F, Benedetti L (2007) Use of Fourier transform infrared (FT-IR) spectroscopy as a tool for pollen identification. Aerobiologia 23(3):211–219

Ilari JL, Martens H, Isaksson T (1988) Determination of particle-size in powders by scatter correction in diffuse near-infrared reflectance. Appl Spectrosc 42(5):722–728. doi:10.1366/0003702884429058

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

Jungfermann C, Ahlers F, Grote M, Gubatz S, Steuernagel S, Thom I, Wetzels G, Wiermann R (1997) Solution of sporopollenin and reaggregation of a sporopollenin-like material: a new approach in the sporopollenin research. J Plant Physiol 151(5):513–519

Kohler A, Kirschner C, Oust A, Martens H (2005) Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform infrared microscopy images of cryo-sections of beef loin. Appl Spectrosc 59(6):707–716. doi:10.1366/0003702054280649

Lasch P, Naumann D (2006) Spatial resolution in infrared microspectroscopic imaging of tissues. Biochim Biophys Acta 1758(7):814–829. doi:10.1016/j.bbamem.2006.06.008

Laucks ML, Roll G, Schweigers G, Davis EJ (2000) Physical and chemical (Raman) characterization of bioaerosols-pollen. J Aerosol Sci 31:307–319

Lin CP, Huang JP, Wu CS, Hsu CY, Chaw SM (2010) Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. Genome Biol Evol 2:504–517. doi:10.1093/gbe/evq036

Lomax BH, Fraser WT, Harrington G, Blackmore S, Sephton MA, Harris NBW (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

Lukacs R, Blumel R, Zimmermann B, Bagcoglu M, Kohler A (2015) Recovery of absorbance spectra of micrometer-sized biological and inanimate particles. Analyst. doi:10.1039/C5AN00401B

Martens H, Stark E (1991) Extended multiplicative signal correction and spectral interference subtraction—new preprocessing methods for near-infrared spectroscopy. J Pharm Biomed 9(8):625–635. doi:10.1016/0731-7085(91)80188-F

Pappas CS, Tarantilis PA, Harizanis PC, Polissiou MG (2003) New method for pollen identification by FT-IR spectroscopy. Appl Spectrosc 57(1):23–27

Pappas D, Smith BW, Winefordner JD (2004) Raman imaging for two-dimensional chemical analysis. Appl Spectrosc Rev 35(1–2):1–23. doi:10.1081/asr-100101219

Parodi G, Dickerson P, Cloud J (2013) Pollen identification by Fourier transform infrared photoacoustic spectroscopy. Appl Spectrosc 67(3):342–348. doi:10.1366/12-06622

Schulte F, Lingott J, Panne U, Kneipp J (2008) Chemical characterization and classification of pollen. Anal Chem 80(24):9551–9556. doi:10.1021/Ac801791a

Schulte F, Mader J, Kroh LW, Panne U, Kneipp J (2009) Characterization of pollen carotenoids with in situ and high-performance thin-layer chromatography supported resonant Raman spectroscopy. Anal Chem 81(20):8426–8433

Schulte F, Panne U, Kneipp J (2010) Molecular changes during pollen germination can be monitored by Raman microspectroscopy. J Biophotonics 3(8–9):542–547. doi:10.1002/jbio.201000031

Schulz H, Baranska M (2007) Identification and quantification of valuable plant substances by IR and Raman spectroscopy. Vib Spectrosc 43(1):13–25. doi:10.1016/j.vibspec.2006.06.001

Wehling K, Niester C, Boon JJ, Willemse MTM, Wiermann R (1989) P-coumaric acid—a monomer in the sporopollenin skeleton. Planta 179(3):376–380

Willis KJ, Feurdean A, Birks HJ, Bjune AE, Breman E, Broekman R, Grytnes JA, New M, Singarayer JS, Rozema J (2011) Quantification of UV-B flux through time using UV-B-absorbing compounds contained in fossil Pinus sporopollenin. New Phytol 192(2):553–560. doi:10.1111/j.1469-8137.2011.03815.x

Yu P (2011) Microprobing the molecular spatial distribution and structural architecture of feed-type sorghum seed tissue (*Sorghum Bicolor* L.) using the synchrotron radiation infrared microspectroscopy technique. J Synchrotron Radiat 18(Pt 5):790–801. doi:10.1107/S0909049511023727

Zimmermann B (2010) Characterization of pollen by vibrational spectroscopy. Appl Spectrosc 64(12):1364–1373

Zimmermann B, Kohler A (2014) Infrared spectroscopy of pollen identifies plant species and genus as well as environmental conditions. PLoS One 9(4):e95417. doi:10.1371/journal.pone.0095417

6.3 Paper III

analytical chemistry

Article

pubs.acs.org/ac

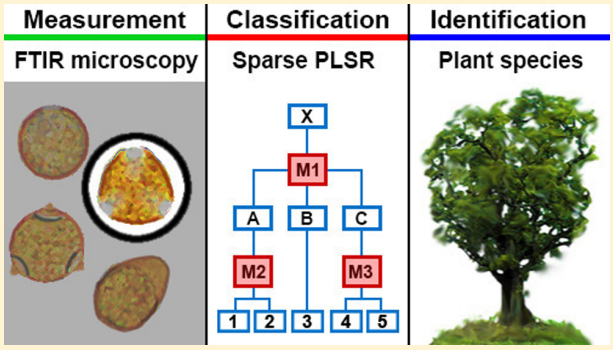
Analysis of Allergenic Pollen by FTIR Microspectroscopy

B. Zimmerman,* V. Tafintseva, M. Bağcıoğlu, M. Høegh Berdahl, and A. Kohler

Department of Mathematical Sciences and Technology, Faculty of Environmental Science and Technology, Norwegian University of Life Sciences, 1432 Ås, Norway

Supporting Information

ABSTRACT: Fourier transform infrared (FTIR) spectroscopy is a powerful tool for the identification and characterization of pollen and spores. However, interpretation and multivariate analysis of infrared microscopy spectra of single pollen grains are hampered by Mie-type scattering. In this paper, we introduce a novel sampling setup for infrared microspectroscopy of pollens preventing strong Mie-type scattering. Pollen samples were embedded in a soft paraffin layer between two sheets of polyethylene foils without any further sample pretreatment. Single-grain infrared spectra of 13 different pollen samples, belonging to 11 species, were obtained and analyzed by the new approach and classified by sparse partial least-squares regression (PLSR). For the classification, chemical and physical information were separated by extended multiplicative signal correction and used together to build a classification model. A training set of 260 spectra and an independent test set of 130 spectra were used. Robust sparse classification models allowing the biochemical interpretation of the classification were obtained by the sparse PLSR, because only a subset of variables was retained for the analysis. With accuracy values of 95% and 98%, for the independent test set and full cross-validation respectively, the method is outperforming the previously published studies on development of an automated pollen analysis. Since the method is compatible with standard air-samplers, it can be employed with minimal modification in regular aerobiology studies. When compared with optical microscopy, which is the benchmark method in pollen analysis, the infrared microspectroscopy method offers better taxonomic resolution, as well as faster, more economical, and bias-free measurement.



Aerollergenic pollens and spores trigger asthma, allergic rhinitis, and hypersensitivity pneumonitis in susceptible individuals. These respiratory diseases are affecting 10−20% of European population, causing multibillion costs of medical treatments and reduced work productivity.[1,2] Moreover, the prevalence of respiratory allergic diseases has increased markedly during the last 20 years.[3] It is expected that shifts in aeroallergen human exposures due to global climate change are going to increase the prevalence and severity of allergies.[4,5] The risks analysis of urban atmospheric pollution in Madrid has showed that the effect of allergenic pollen on the health of individuals is greater than the detected effect of traditionally analyzed parameters, such as chemical pollution.[6] Qualitative and quantitative analysis of pollen is the cornerstone of aeroallergen monitoring networks. Providing information on the occurrence of aeroallergens in the residential and working areas is vital for timely treatment of inhalant allergies through correlation of symptoms. Moreover, air quality forecasts are essential for avoidance and prevention that reduces allergic symptoms. Finally, palynology, the study of pollen, spores, and similar bioparticles, has great general interest in analysis of pollen grains, for example, in studies of environmental and climate change, biodiversity monitoring, and tracing vegetation history.

Current identification of pollen relies on optical (transmitted light) microscopy. Identification is based on qualitative descriptions of pollen's morphology, usually size, shape, and texture, as well on its substructures. The analysis is time-intensive, requiring extensive sample preparation and laborious identification by experienced specialists. The identification is limited to categories that can be reliably classified by multiple analysts, resulting in the rough taxonomic resolution. In addition, microscopy methods are prone to inconsistency and a degree of subjectivity because the performance of the individual analyst can arise through issues like fatigue and overfamiliarity with samples. For this reason, there is a trend in moving toward computer-based systems that deliver faster, more accurate, and more consistent semi- or fully automated identifications than any human operator.[7,8] However, even if traditional microscopies increase sample processing by automation, the morphological approach is limited since many species have indistinguishable morphology. For instance, grass family (Poaceae) is one of the largest plant families with over 10 000 species, but grass pollens are rarely identified at the

Received: August 20, 2015
Accepted: November 24, 2015
Published: November 24, 2015

DOI: 10.1021/acs.analchem.5b03208
Anal. Chem. 2016, 88, 803−811

subfamily level. Alternative analyses, such as genetic sequencing and biochemical methods, are even more time-consuming, expensive, and difficult to implement in routine analysis.

Infrared spectroscopy is a biophysical method that provides precise signatures of the overall biochemical composition of pollen. Infrared spectra of pollen contain specific signals of lipids, proteins, carbohydrates, and cell wall biopolymers, such as cellulose and sporopollenins. Therefore, Fourier transform infrared (FTIR) spectroscopy has emerged as a powerful tool in identification and characterization of pollen and spores.[9−22] Moreover, FTIR spectroscopy offers an operator-independent approach based on chemical fingerprinting via identifiable spectral features, with significant potential for automation of pollen identification.[7] Since samples can be measured in their native form, without any chemical pretreatment, FTIR spectroscopy achieves simple, economical and rapid analysis of pollen. In FTIR microspectroscopy (μFTIR), an infrared spectrometer is combined with a microscope, thus enabling focused measurement of an individual bioparticle. Unfortunately, bioparticles such as pollen are highly scattering particles in the infrared due to their size and morphology. Therefore, strong scattering signatures in μFTIR spectra of single pollen grains hamper biochemical interpretation of spectra.[13,22,23] The strong scattering by the measured pollen grains results in anomalous spectral features that significantly interfere with and distort the signals of chemical absorption.[23] Extracting information from recorded data by separating absorption and scattering contributions is extremely challenging because scattering and absorption are highly entangled. Moreover, the pollen shapes and structures differ considerably from spherical and homogeneous particles as assumed for existing correction models.[23,24] In general, scattering effects have been identified as a major obstacle for application of FTIR microspectroscopy in biological and biomedical science.

In the presented study, we have implemented an innovative approach for scatter-free FTIR microscopic measurement of single bioparticles, specifically single pollen grains. The aim of the study was to test the new approach for aeroallergenic pollen species by spectroscopy-based identification of plant taxa. The chemical characterization of pollen grains and extraction of morphological parameters from spectral data is discussed in details. Moreover, the complementarity of FTIR microspectroscopy and contemporary air-sampler systems has been taken into consideration. The presented approach would enable rapid implementation of FTIR microspectroscopy in aeroallergen monitoring networks, as well as in environmental and climate change studies.

EXPERIMENTAL SECTION

Samples. The pollen types used were collected from *Fagus sylvatica* (European beech), *Betula pendula* (Silver birch), *Carpinus betulus* (European hornbeam), *Quercus cerris* (Turkey oak) and *Quercus robur* (Penduculate oak) of the Fagales (beech) order, *Bromus erectus* (Erect brome) and *Carex pendula* (Pendulous sedge) of the Poales (grass) order, *Cryptomeria japonica* (Japanese cedar) *Cupressus sempervirens* (Mediterranean cypress), *Juniperus phoenicea* (Phoenicean juniper) and *Juniperus chinesis* (Chinese juniper) of the Pinales (pine) order. *Betula pendula* was represented with three samples, collected from three different individuals: one wild type and two "Youngii" cultivars. All other species were represented with a single sample. Each sample comprised pollen belonging to either: (1) approximately 10 inflorescences from the same turf (for Poales), (2) approximately 10 male catkins from the same individual plant (for Fagales), and (3) approximately 50 male strobili from the same individual plant (for Pinales). In total, 13 pollen samples were measured. Pollen samples were obtained through fieldwork at the Botanical Garden of the Faculty of Science, University of Zagreb, during 2012 pollination seasons. The pollen samples were collected directly from plants at flowering time, and stored at −15 °C.

Measurement. The pollen samples were embedded, without any sample pretreatment, in a matrix medium, consisting of a soft paraffin (Vaseline-type petroleum jelly) between two sheets of 40 μm thick polyethylene foils. The amount of paraffin was approximately 1 mg per 25 cm^2. A custom-made sample holder was made in order to stretch the foils and to minimize the appearance of air bubbles in the sample (Figure S-1, in the Supporting Information).

Infrared spectroscopy measurements were obtained on a VERTEX 70 FTIR spectrometer with a HYPERION 3000 IR microscope (Bruker Optik GmbH, Germany). The spectra were recorded in transmission mode with a spectral resolution of 4 cm^{-1} and digital spacing of 1.929 cm^{-1}. Thirty single grains were measured per pollen sample, and for each grain, one corresponding FTIR spectrum was obtained. The system is equipped with a globar mid-IR source and a liquid nitrogen-cooled mercury cadmium telluride (MCT) detector. The spectra were measured in the 4000−600 cm^{-1} spectral range, with 64 scans for both background and sample spectra, and using a 15× objective, with 25 × 25 μm aperture sizes. For the measurements in the matrix media, background (reference) spectra of air were recorded once, before starting each (sample) measurement using the sample-free setup. The quality of recorded spectra was measured by the performing Bruker OPUS quality test for measurement of microorganisms, by calculating the signal-to-noise ratio of Amide I signal (at 1700−1600 cm^{-1}).

The referent (matrix-free) single grain pollen spectra were measured by depositing pollen samples onto 3 mm ZnSe optical slides (windows) without any sample pretreatment. For the measurements on ZnSe optical windows, background (reference) spectra were recorded immediately before starting each measurement by measuring empty areas of ZnSe windows. The referent bulk pollen spectra were recorded by measuring potassium bromide (KBr) sample pellets. KBr pellets were prepared by mixing and grinding approximately 1 mg of a sample with approximately 100 mg of KBr using a pestle and mortar, and the created KBr matrix was then cold-pressed without degassing into a transparent disk. A sample-free setup was used to obtain the spectral background.

The Bruker system was controlled with OPUS 7.5 software (Bruker Optik GmbH, Germany).

Spectral Preprocessing. From each of the 13 pollen samples, 30 single grain spectra of 30 different pollen grains were obtained. In total, 360 μFTIR spectra were included in the data analysis. The spectra were preprocessed prior to calibration. The spectral region of 1900 cm^{-1} to 900 cm^{-1} was selected for data analysis, excluding the region of 1520 to 1290 cm^{-1}, which is the saturation region of the polyethylene foil and paraffin (i.e., the region with strong absorbance bands of matrix media). Two types of preprocessing procedures were used in the study: (1) preprocessing of second-derivative spectra and (2) preprocessing of nonderivated spectra. For the derivated data, all spectra were transformed to second-derivative spectra by the Savitzky−Golay algorithm using a

DOI: 10.1021/acs.analchem.5b03208
Anal. Chem. 2016, 88, 803−811

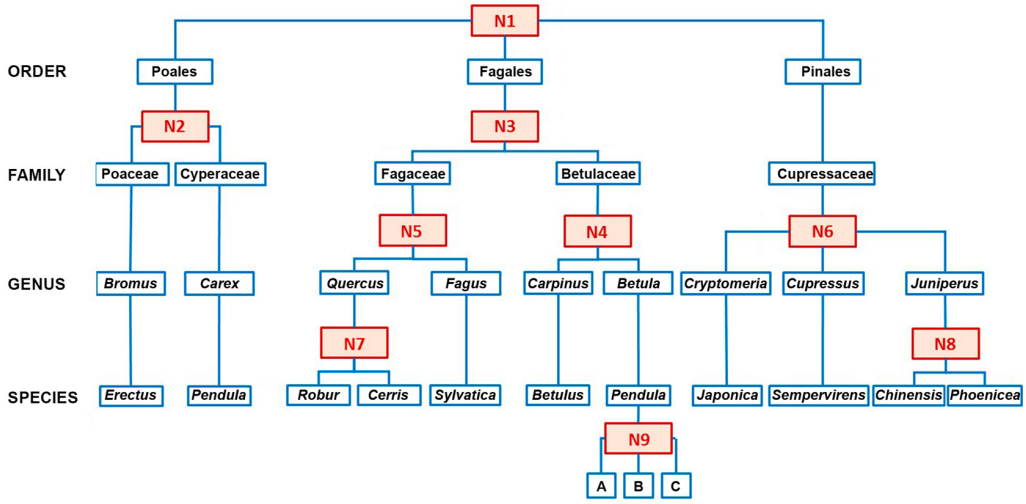


Figure 1. Phylogenetic tree for classification of pollen spectra, containing nine nodes (N1−N9).

polynomial of degree two and a window size of 31 points in total, followed by extended multiplicative signal correction (EMSC, an MSC model extended by a linear, quadratic and cubic component).[25−27] The selection of window size was based on suppressing noise while enhancing intensity of broad amide bands in the 1650−1520 cm^{-1} region, as explained in detail in our previous study.[26] The EMSC model used in the preprocessing is defined by equations S-1 and S-2, in the Supporting Information.

For the nonderivated data set, all spectra were preprocessed by EMSC. In the EMSC preprocessing of nonderivated data, 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 region 1900−1800 cm^{-1}, where the weighting vector was set to 10. The region 1900−1800 cm^{-1} is devoid of any chemical absorbance and thus this region should have the same baseline values in all pollen spectra when physical effects (such as light reflection) have been removed. Up-weighting of this region, by applying a weighting vector, is constraining the EMSC preprocessing and ensuring a stable baseline in all pollen spectra. For the data analysis, three different data sets were used: (A) the derivated and EMSC preprocessed spectral data set alone, (B) the EMSC parameters data set obtained from the nonderivated spectra, and (C) a combined data set where EMSC parameters obtained from the nonderivated spectra were concatenated with the derivated and EMSC preprocessed spectra. The spectral preprocessing was performed by algorithms developed in-house in the environment of Matlab, R2014a (The Mathworks Inc., Natick, United States).

Data Analysis and Classification. The data sets A and C were used to evaluate classification and biochemical similarities between pollen samples by using principal component analysis (PCA) and sparse partial least-squares regression (Sparse PLSR). Sparse PLSR was used for building a hierarchical classification scheme. The hierarchical classification scheme was built following the phylogenetic tree shown in Figure 1 allowing classification at various subcategories such as order, family, genus, species, and sample. In each node (N1−9 in Figure 1) a classification model was established for 2 or 3 classes, depending on the number of subclasses connected to each node. For establishing the classification models, matrices of indicator variables were set up for each node (**Y**-matrices). The indicator matrices contain one column for each subclass of the nodes with numbers ones and zeros indicating if a sample belongs to a respective subclass or not. Sparse PLSR was used to regress the spectral data sets (**X**-data sets) on the matrices of indicator variables (**Y**-matrices) at each node. In analogy to discriminant PLSR,[28] we denote the sparse PLSR model used for classification as discriminant Sparse PLSR. Sparse PLSR allows straightforward biochemical interpretation of spectral bands because only a limited subset of variables is selected in the modeling process.[29] Three classification models, using two different data sets (as defined in the Spectral Preprocessing section), were built on the basis of the tree shown in Figure 1: 1) A sparse model based on data set A, 2) a sparse model based on data set C, and 3) a mixed sparse model based on data set A at all levels except "order" level (N1) where data set C was used. Each spectral data set (390 samples in total) was divided into a training set used for calibration (20 spectra per sample, 260 spectra in total), and an independent test set (10 spectra per sample, 130 spectra in total). The training set was used to establish Sparse PLSR classification models employing Venetian Blinds Cross-Validation (CV) with three segments where samples were split into blocks in the following order: 123 123.... For the training set, block separation was done by full or leave-one-out CV due to more reliable and faster performance. For each principal component, Sparse PLSR algorithm penalizes loading weight vectors according to a parameter called degree of sparsity. This parameter defines how many variables will be neglected and has previously been predefined.[30] In this paper, we use an optimization procedure for sparsity to choose the optimal degree of sparsity. For each PLSR component, an optimal degree of sparsity was chosen among four in the range from 80 to 99% of variables with a step of 4.75. The degree of sparsity, which gave the minimum missclassification rate (MCR) was selected. The selected degrees of sparsity were then used in the final PLSR calibration. Finally, the optimal number of components (*AOpt*) was chosen. *AOpt* was found by evaluating the MCR as a function of the number of components A. *AOpt* is the one which does not have significantly higher MCR than the minimum MCR. A binomial distribution formed a basis for a binomial test, which

DOI: 10.1021/acs.analchem.5b03208
Anal. Chem. 2016, 88, 803−811

was used to evaluate the statistical significance (function binocdf in Matlab).

Data analyses were performed by algorithms developed in-house in the environment of Matlab, R2014a (The Mathworks Inc., Natick, United States).

RESULTS AND DISCUSSION

Scattering Spectral Artifacts and Experimental Setting. In Figure 2, spectra of single pollen grains measured on

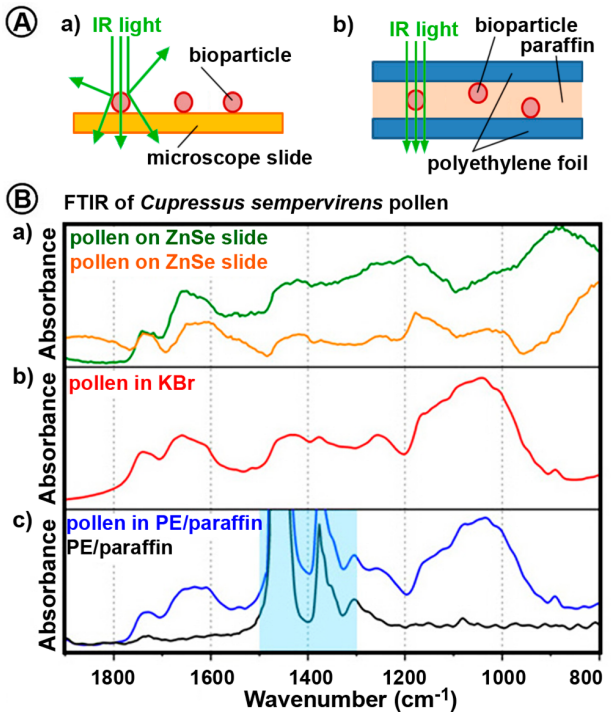


Figure 2. (A) Schematics of μFTIR measurements of bioparticles on: (a) microscope slide and b) within paraffin matrix; infrared light transmitted through and reflected on bioparticles is represented with arrows (specular reflection from microscope slide and polyethylene films is omitted for simplification). (B) Infrared spectra of the same *Cupressus sempervirens* pollen sample obtained with different experimental settings: (a) Spectra obtained by measurements of single pollen grains on microscope slides (unreproducible due to scatter effects); (b) Benchmark spectrum obtained by bulk measurement of a KBr sample pellet; (c) Scatter-free single grain spectrum (blue) obtained by measurement in paraffin and polyethylene (PE) embedding medium (spectrum of embedding media (black) is included for comparison). Blue shading designates the spectral region that was excluded from data analysis.

microscope slides are shown (Figure 2Ba). The spectra show substantial and complex anomalies when compared with the reference spectra of bulk samples (Figure 2Bb). Due to variations in morphological and optical properties of pollen grains within a sample, the corresponding pollen spectra have different scatter distortions. These spectral anomalies include wavelike fluctuating structures with narrower and broader signals, that can significantly interfere with and distort the signals of chemical absorption.[22,23] It is worth noting that scatter contributions can provide valuable information on morphology of pollen grain. However, due to the high morphological variation of pollen grains within the same sample, they hinder classification of pollen types by FTIR microspectroscopy of single pollen grains. In contrast, it is known that the biochemical composition as probed by FTIR spectroscopy is highly type-specific. Previous studies have shown that the biochemical information obtained by scatter-free absorbance spectra from KBr-pellet measurements of pollen bulk samples can separate pollen on different phylogenetic levels.[20] Thus, for the identification and chemical analysis of pollen, it is important to obtain pure (scatter-free) absorbance spectra that can serve as a fingerprint of the grain's biochemical components.

The scattering phenomena occurring in the μFTIR spectroscopy of pollen have been interpreted as Mie-type scattering.[23] Bioparticles, such as pollen, are highly scattering samples in infrared spectroscopy because their sizes are often of the same order as the mid-infrared wavelengths employed in FTIR spectroscopy (2.5−25 μm). This leads to strong Mie-type scattering.[27,31−34] It is very difficult to separate the Mie-type scattering from the pure absorbance spectra. The reason is that the real refractive index undergoes strong fluctuations due to chemical absorbance bands and is thus highly dispersive. The real refractive index spectrum can be obtained from the pure absorbance spectrum, or equivalently from the imaginary part of the refractive index, by the Kramers−Kronig transform and vice versa. Unfortunately, for a measured infrared spectrum with Mie-type scattering, both the pure absorbance spectrum and the real refractive index spectrum are unknown. Recently, an iterative algorithm based on extended multiplicative signal correction (EMSC) for the retrieval of the pure absorbance spectra from infrared spectra with Mie-type scattering has been developed. Although the developed algorithm is robust and has provided some success, it remains unclear, to what extent, in general, it retrieves the pure absorbance spectrum, i.e. whether the algorithm converges toward the underlying pure absorbance spectrum or toward the reference spectrum used in the iterative EMSC approach. Another weakness of the algorithm is that it does not take into account the high numerical aperture of infrared microscopes. Recently, a different iterative algorithm has been presented taking into account the high numerical aperture.[35] This algorithm has been only tested on a model system (PMMA spheres) and could not be successfully applied for the correction of pollen.[23] Finally, all previous developed models for correcting Mie-type scattering in FTIR spectroscopy of bioparticles assume a spherical scatterer. While the shape of pollen grains is in many cases close to spheres, it can deviate considerably.[14] In addition, the pollen structure is inhomogeneous, including a double layered grain wall and a multicellular grain interior made of cytoplasm, organelles, vacuoles, membranes, and nuclei.

In order to obtain scatter-free spectra of single pollen, we have developed an innovative experimental setting. Because the strength of the scatter-distortions depends on refractive index change at the surface of the bioparticle, scattering effects may be suppressed considerably by embedding samples into a medium with similar index of refraction as the bioparticle to be measured (Figure 2Ab). One suitable embedding matrix for μFTIR of bioparticles is a thin layer of soft paraffin. In our study, 13 different plant samples were embedded in a paraffin matrix between two thin sheets of polyethylene foil and measured by μFTIR. Thirty pollen grains per plant sample were measured, thus 360 spectra were obtained, each belonging to a different single grain. The obtained spectra are scatter-free (Figure 2Bc).

DOI: 10.1021/acs.analchem.5b03208
Anal. Chem. 2016, 88, 803−811

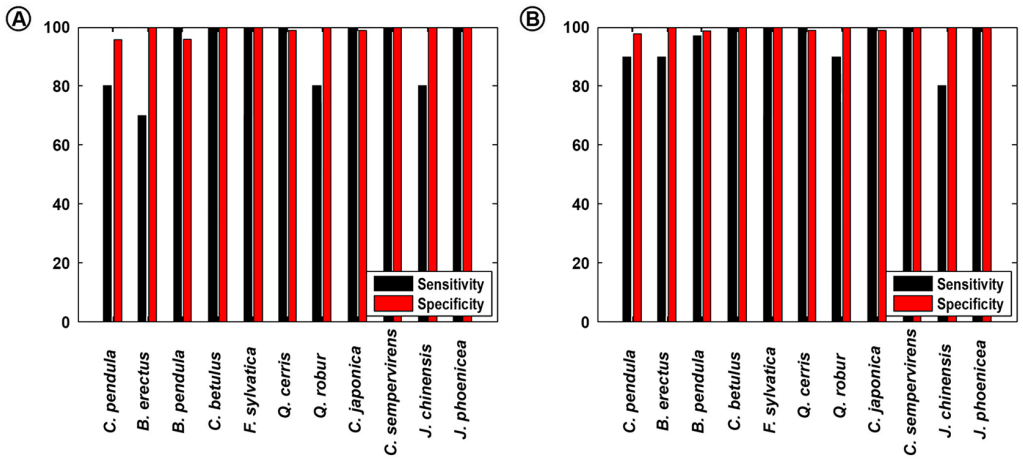


Figure 3. Identification results for Sparse PLSR analysis of pollen μFTIR spectra, showing bar plots of sensitivity and specificity for each species (based on 10 spectra per sample). The results are based on (A) a sparse PLSR classification model built on second-derivative spectral data (data set A) and on (B) a sparse PLSR classification model built on second-derivative spectral data (data set A) at all levels except "order" level (N1) where a sparse PLSR classification model built on second-derivative spectral data with included EMSC parameters (data set C) was used.

The embedding matrix contains several strong absorption signals in the spectral region from 1500 to 1350 cm^{-1}. These bands are associated with CH_2 bending (1470 cm^{-1}), CH_3 antisymmetric (1463 cm^{-1}) and symmetric (1377 cm^{-1}) deformations, and CH_2 wagging vibrations (1353 and 1305 cm^{-1}).[36] The principal matrix-associated bands are oversaturated in all spectra, and therefore, the surrounding region, from 1520 to 1290 cm^{-1}, was excluded from the analysis (Figure 2Bc). Becuase this region is devoid of principal absorption signals of pollen, the subsequent spectral data analysis was not affected by large information shortfall. Additional information on sample measurement, including Figure S-2, are provided in the Supporting Information.

Classification and Discrimination of Pollen. Pollen grains have unique vibrational spectra. PCA plots (Figure S-3, in the Supporting Information) show well-defined clustering at lower taxonomic levels, including species level. By using the first four principal components, spectra of samples belonging to distinct plant species can be separated quite well from each other. Moreover, EMSC parameter estimates of pollen μFTIR spectra have a range of taxon-specific values (Figures S-4 and S-5, in the Supporting Information). Although EMSC parameters, containing baseline, background, and other anomalous features, are often discarded in spectral analysis, they may contain information that is important for characterization of pollen.[21] EMSC parameter information is closely related to a characteristic morphology of a pollen grain, and pollen morphology varies substantially in shape, size, and wall pattern across plant taxa. As such, EMSC parameters offer an indirect physical measurement of a pollen sample that should be used for pollen characterization, alongside chemical absorption spectrum. Therefore, μFTIR spectra of pollen show large potential for classification analysis.

Multivariate methods such as PCA and PLSR, allow investigating to what extent infrared spectra can be used for classification and discrimination of pollen types. PCA and PLSR are based on latent variables (also called scores) and permit direct interpretation of spectral signatures on the biochemical level. This interpretation is based on assessing regression coefficients and loading vectors. The regression coefficients relate the FTIR variables (wavelengths and thus spectral bands) directly to the classes, while loading vectors relate the new latent variables to the FTIR variables. Evaluation of biochemical fingerprints is important, because a meaningful biological interpretation of the regression coefficients represents an additional validation of the discrimination and classification between different pollen on taxonomical level, for example differentiation of families, genera, species, or populations. However, interpretation of results can be challenging in situations where the number of variables is high, such is the case for FTIR data. In such cases, correlation coefficients and loading vectors are too complex for interpretation as a result of the multitude of spectral signatures revealed. Sparse PLSR combines variable selection and modeling, and thus enhances the interpretation of the classification analysis by discarding uninformative spectral variables.[30] The Sparse PLSR classification of the measured pollen spectra was based on the phylogenetic tree presented in Figure 1. As can be seen, the tree contained nine nodes (N1−N9) that split at different levels (order, family, genus, species, and sample) to different numbers of subclasses (two or three). Thus, the Sparse PLSR method was used to establish 9 calibration models, one for each node in the tree.

Among the 9 calibration models, the most complex model included *AOpt* with 4 and 6 components, for data set A and data set C, respectively, with average complexity of *AOpt* = 2 for both data sets. The advantage of sparse PLSR modeling lies in easier interpretation and often lower complexity of the models.

Sensitivity and specificity of discrimination analyses are presented in Figure 3. When considering only chemical spectral data (data set A), overall accuracy was 93.1%, with all except one taxa having sensitivity greater than or equal to 80% (Figure 3A). When EMSC parameters were included in the analysis (data set C), overall accuracy remained at 93.1%; however, all taxa had sensitivity equal or better than 80% (Figure S-6, in the Supporting Information). Specificity was high (more than 95%) in all cases. Although both data sets obtain the same overall accuracy, they differ substantially regarding accuracy at "order" level of classification (node N1 in Figure 1) where data set C performs much better than data set A. When EMSC parameters were included in the analysis (data set C), misclassification of

DOI: 10.1021/acs.analchem.5b03208
Anal. Chem. 2016, 88, 803−811

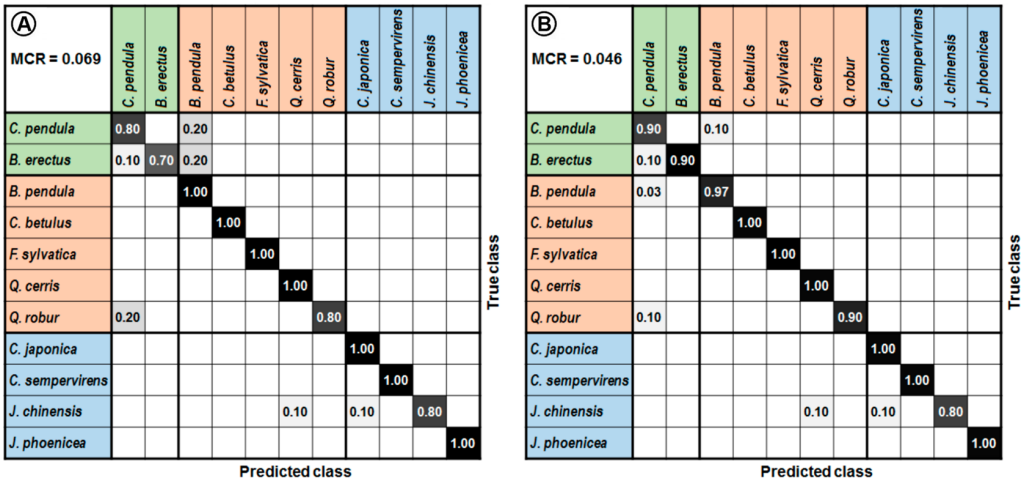


Figure 4. Confusion matrix for Sparse PLSR analysis of pollen μFTIR spectra; results based on (A) a sparse PLSR classification model built on second-derivative spectral data (data set A) and on (B) a sparse PLSR classification model built on second-derivative spectral data (data set A) at all levels except "order" level (N1) where a sparse PLSR classification model built on second-derivative spectral data with included EMSC parameters (data set C) was used. Color codes: green (Poales), red (Fagales), and blue (Pinales). (MCR = misclassification rate).

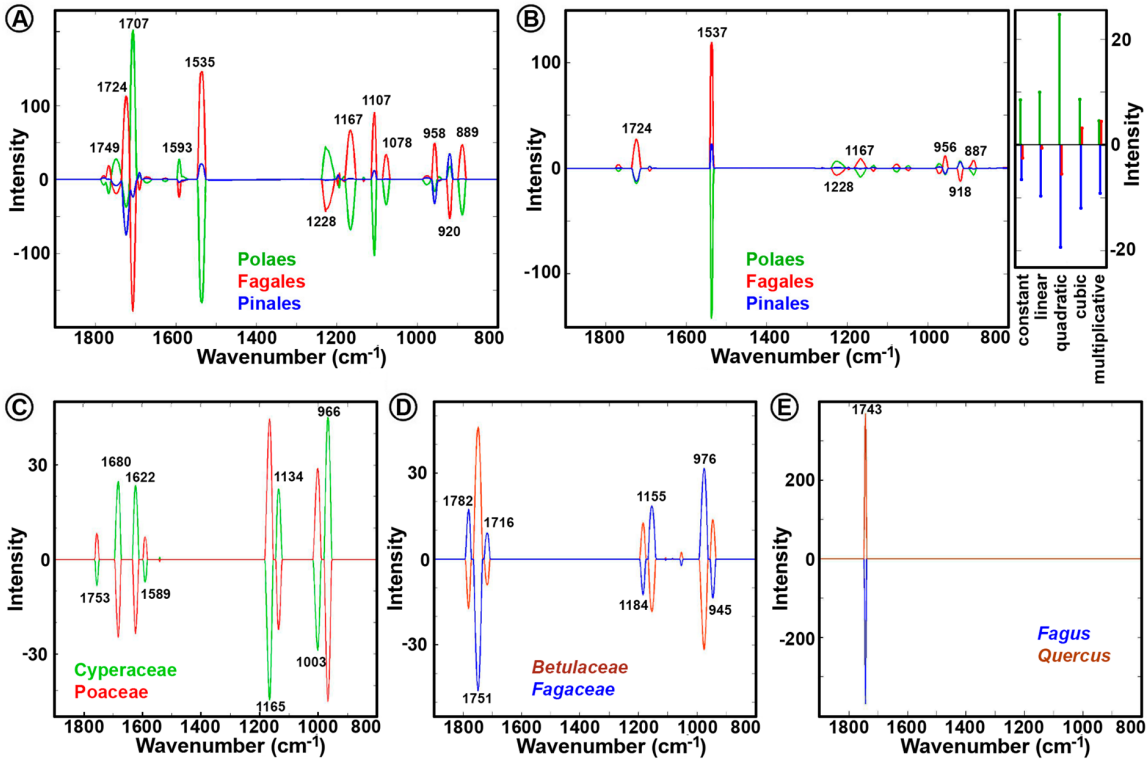


Figure 5. Plot of Sparse PLSR regression coefficients for second-derivative spectral data: classification of plant orders (node N1) by using (A) second-derivative spectral data (data set A) and (B) second-derivative spectral data with included EMSC parameters (data set C); classification by using second-derivative spectral data (data set A) of (C) Poales families (node N2), (D) Fagales families (node N3), and (E) Fagaceae genera (node N5).

plant orders was reduced by 43%. Therefore, we introduced a classification model combining a sparse PLSR classification model built on data set A and a sparse PLSR classification model built on data set C. We used a sparse PLSR classification model built on data set A at all levels except "order" level where a sparse PLSR classification model based on data set C was used. The model has resulted with an overall accuracy of 95.4% and specificity equal or more than 98% for all species (Figure 3B).

Figure 4 presents the corresponding confusion matrices for the analyzed data sets with independent validation. In general, inclusion of EMSC parameters in the classification model has resulted in adequate accuracy (i.e., 80% or higher) for all species and a low level of systematic misclassification. It should be noted that, when establishing an FTIR classification tree, the

DOI: 10.1021/acs.analchem.5b03208
Anal. Chem. 2016, 88, 803−811

actual pollen taxonomy might not be optimal for the establishment of all nodes, because spectral similarities of pollen might differ in some cases from established phylogeny.[20]

The presented model performance results are realistic and based on the best-case scenario in modeling when one can afford establishing calibration models and testing them on an independent test set. Cross-validated calibration results are more often reported in the literature on pollen classification.[13,37−40] Unfortunately, these results are usually overoptimistic and do not give a real estimate on the model performance. Notwithstanding, cross-validation is a good option if only a limited number of spectra is available. In order to compare our results with the ones presented in the literature we report full (leave-one-out) CV results which we run on the full data set (390 spectra). When considering only chemical spectral data (data set A), the accuracy on the species level was 96.9%. When a sparse PLSR classification model was used, based on data set A at all levels except "order" level where data set C was used, the accuracy on the species level increased to 97.7%.

It is important to notice that there was no misclassification at subgenus level (Figure 4). It is apparent that μFTIR enables classification of congeneric species as demonstrated by absence of misclassification between either two *Quercus* or two *Juniperus* species. Moreover, it is possible that the method could even enable classification at subspecies level. This has been tested by excluding one of the *Betula pendula* cultivar samples from the test set and using these 30 single grain spectra for an external validation. In that case, 77% of the kept away *B. pendula* cultivar grains were correctly identified as *B. pendula* cultivar, 17% were identified as *B. pendula* wild type, while 6% were not identified as *B. pendula*.

The plots of the regression coefficients for the Sparse PLSR model are displayed in Figure 5. It can be seen that, when considering only chemical spectral data (data set A), spectral differences between pollen orders (Poales, Fagales, and Pinales) are present in the whole spectral range (Figure 5A). The corresponding regression coefficients for a sparse PLSR classification model including EMSC parameters (data set C) are shown in Figure 5B. They are significantly less complex compared to the regression coefficients in Figure 5A, with only two main contributing bands that can be associated with lipids (C═O stretch at 1724 cm^{-1}) and proteins (Amide II band at 1537 cm^{-1}). Therefore, it is apparent that the significant reduction in misclassification at plant order level can be attributed to EMSC parameters. This improvement is expected because EMSC parameters are related to morphology of pollen, and pollen grains of the three plant orders have extremely different morphological features.[21] As mentioned previously, a morphological approach at lower taxonomic levels (family and below) is limited due to lack of distinctive morphological features, and thus, inclusion of EMSC parameters may even result in worse classification. Therefore, inclusion of EMSC parameters (data set C) in the whole analysis (i.e., in all nodes of classification tree) has not improved overall accuracy. However, a sparse PLSR classification model built on either data set A or C at different nodes, has resulted in improved overall accuracy.

The regression coefficients at lower taxonomic levels are less complex than at order level and emphasize the advantage of Sparse PLS in enhancing the interpretation of the classification results. Discrimination between grass (*Bromus erectus*) and sedge (*Carex pendula*) is based on stronger lipid-related absorption of sedge pollen at 1745 cm^{-1} (C═O stretch) and 1163 cm^{-1} (C−O stretch), and stronger protein-related absorption of grass pollen in the "protein" region (Figure 5C). Unfortunately, only two Poales species were included in this study, and thus, further studies with a larger pollen set are required. However, the result is in accordance with a previous study on 15 sedge (Cyperaceae) and 32 grass (Poaceae) pollens, which has shown that pollen of sedges have significantly increased lipid content when compared with the grasses.[20] Moreover, the previous study has shown that pollen of grass and sedge families have relatively low FTIR spectral variability compared to, for example cypress (Cupressaceae) and beech (Fagaceae) families.[20] Therefore, high accuracies for discrimination of grasses and sedges pollen are expected in general.

The regression coefficients for classification of Fagales pollen between birch (Betulaceae) and beech (Fagaceae) families show that the classification is based on "lipid" and "carbohydrate" regions. The regression coefficients for classification of Fagaceae pollen between oak (*Quercus*) and beech (*Fagus*) genera are even less complex and show that the principal difference is based on absorption in the "lipid" region at 1743 cm^{-1} (C ═O stretch) (Figures 5E). Again, this is in accordance with previous spectroscopic studies on pollen composition.[20]

Future Perspective of a μFTIR-Based Pollen Classifier System. In 1996, Stillman and Flenley addressed the need for an automated pollen classifier system.[41] The authors have compared the system with manual microscopy analysis and have established a set of goals for such system. Compared to existing technology and routines, an ideal automated system should allow (1) to count more pollen grains per sample, (2) to reduce analysis time per sample, (3) to reduce analyst bias and improve consistency, and (4) to determine pollen-type at higher levels, preferably at species level. A recently published review by Holt and Bennett on progress of automated pollen counters has extended these goals by adding that the system should require little modification of existing sample collection and preparation techniques, as well as impart no new limitations.[7] Finally, the stated quantitative criteria specify that the automated system should be able to identify at least 40 different taxa with an accuracy of 95%. In comparison, it has been estimated that accuracy based on the morphological recognition by a human operator, used in aerobiological monitoring networks, is approximately 80%.[13]

Of all the published attempts on an automated palynology system, only seven attempts have included classification of 10 taxa or more.[7] Most of the studies were performed on the morphology-based systems, by a combination of image processing and a learning-based classifier. The classification success rates of these systems, tested on a comparable range of taxa as in our study, were approximately 92%.[37−40] These classification accuracies are all lower than 98% accuracy obtained in the here presented μFTIR system. Unfortunately, all these studies were using cross-validation estimation of model performance, thus the reported accuracy is probably overestimated. There has been only one study, by Kaya et al., using an independent set of data for validation of model performance, reporting an accuracy of 88%.[42] This is lower the accuracy of 95% obtained here with μFTIR, when comparable independent estimation of accuracy was performed. However, it is difficult to make comparison between the two studies since in the study by Kaya et al., plant taxa were not specified, although it can be

DOI: 10.1021/acs.analchem.5b03208
Anal. Chem. 2016, 88, 803−811

assumed that only pollen of entomophilous (insect-pollinated) plants was included. Because aeroallergenic pollen is predominantly anemophilous (wind-pollinated), with quite different chemical and morphological properties than entomophilous pollen,[20,43] any further comparison of the results is irrelevant.

Up until now, there has been only one attempt with μFTIR classification of pollen, classifying 11 taxa at an accuracy of 84%.[13] The relatively low accuracy, compared to morphology-based systems, was due to relatively high spectral variability of pollen single grains within a sample. At that time, the spectral variability was attributed to chemical variability between the grains due to natural variation or due to pollen immaturity. However, a recent study has shown that the spectral variability is predominantly caused by Mie scattering artifacts.[23] This has been confirmed by the presented study, where reproducible and scatter-free spectra were obtained by conducting pollen measurements in embedding matrix. As expected, this has resulted in a significantly higher classification accuracy of 98%.

The presented embedding technique for μFTIR measurements of pollen is adaptable to routine aeroallergen analyses. It is interoperable with the Hirst-type volumetric pollen trap, which is the most prevalent type of pollen sampler for air-sampling. In our experimental setup, the pollen samples were embedded in soft paraffin, which is a standard adhesive substance in Hirst-type pollen traps.[44] The polyethylene foil, which we used as substrate for the embedding matrix, is comparable to the polyester foil, a commonly used holder in pollen traps. The amount of paraffin used in μFTIR measurements is comparable to the amount used in standard pollen monitoring measurements. Therefore, a future μFTIR system for pollen classification can be in full compliance with the requirements. It is fully in agreement with the existing pollen samplers and preparation techniques, and it can provide identification of pollen up to species level. Moreover, identification and classification is bias-free and thus provides higher consistency than human experts. The μFTIR methodology is devoid of labor-intensive sample preparation and data acquisition tasks, and the method does not require specific high-skilled expertise, as opposed to the traditional morphology-based pollen studies. In addition, μFTIR has a high potential for further automation of measurement processes.

Both embedding materials, polyethylene and paraffin, are transparent in both the visible and the infrared spectral region. This is an extremely valuable property because it enables not only chemical characterization of pollen by μFTIR microscopy but also morphological characterization by standard optical microscopy. It also implies that advancement in standard (optical microscopy-based) automated pollen classifier systems can be used in automation of the μFTIR pollen classifier system. For example, as a first measurement step both systems require reliable robotics and image processing to locate pollen on an adhesive sampling tape. After locating pollen grains, a system could acquire an image and an infrared spectrum in successive order to obtain morphological and chemical information on pollen. Therefore, coupling of both approaches into an automated FTIR and optical-microscopies-based system should enable unprecedented classification accuracy.

CONCLUSIONS

The presented study clearly demonstrates that identification of single pollen grains can be based on transmittance FTIR microspectroscopy by using an embedding matrix. The classification analysis shows an enormous potential of the μFTIR method, surpassing state-of-the art approaches by a wide margin. Nevertheless, there is still ample space for further improvement of the method.

First, optimization of the FTIR instrumental setting should be focused on decreasing spectral acquisition time. In this study, we have employed strict instrumental settings with a large safe margin to avoid any possible information shortfall regarding spectral quality. This has resulted in a relatively long spectral acquisition time of approximately 1 min per grain. However, it is safe to assume that this measurement time can be halved, or even quartered, without significant information shortfall. It should be mentioned once more that the μFTIR method saves significant time regarding sample preparation by avoiding any chemical pretreatment, as opposed to standard optical microscopy approach.

Second, embedding media may be further optimized taking into account both the infrared and the visible region of the electromagnetic spectrum. Although the optimization in the infrared is of importance to increase FTIR spectral quality, by decreasing signal-to-noise ratio, the visible region is of importance for aforementioned coupling of FTIR and optical microscopy.

Third, optimization of the classification model requires a 2-fold approach as well. Further measurements are required to evaluate and calibrate this method for a wide range of plant taxa. This includes assessment of pollen classification at subspecies level. Moreover, improvement of classification models require the buildup of a spectral library with a significantly wider range of species and populations than currently available. The results show that identification accuracy is taxon-specific, and thus, measurement of a wide range of species is a necessity for thorough assessment of methodology performance. Furthermore, for the classification of spectral data, we have used Sparse PLSR to establish robust calibration models.

The presented μFTIR method for analysis of pollen is already outperforming the previously published studies of comparable scope. Further improvements are anticipated in the near future, resulting in a better taxonomic resolution, as well as faster and more economical measurements. The study shows enormous potential for pollen and in general bioparticle analysis, such as plant and fungal spores, dinoflagellate cysts and diatom microalgae. Moreover, the experimental setting allows implementation of μFTIR spectroscopy in various fields of palynology, including aerobiology, palaeoecology, terrestrial and aquatic ecology, sedimentology, environmental archeology, forensics, and melissopalynology.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.5b03208.

The material includes details on sample measurement, spectral preprocessing, chemical and physical variability of pollen grains, and discrimination analysis (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: boris.zimmermann@nmbu.no. Tel.: +47 6723 1576. Fax: +47 6496 5001.

DOI: 10.1021/acs.analchem.5b03208
Anal. Chem. 2016, 88, 803−811

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The research was supported by the European Commission through the Seventh Framework Programme (FP7-PEOPLE-2012-IEF project no. 328289).

■ REFERENCES

(1) Cruz, A. A.; Popov, T.; Pawankar, R.; Annesi-Maesano, I.; Fokkens, W.; Kemp, J.; Ohta, K.; Price, D.; Bousquet, J. *Allergy* **2007**, *62*, 1−41.
(2) Wallace, D. V.; Dykewicz, M. S.; Bernstein, D. I.; Bernstein, I. L.; Blessing-Moore, J.; Cox, L.; Khan, D. A.; Lang, D. M.; Nicklas, R. A.; Oppenheimer, J.; Portnoy, J. M.; Randolph, C. C.; Schuller, D.; Spector, S. L.; Tilles, S. A.; May, K. R.; Miller, T. A.; Druce, H. M.; Baroody, F. M.; Bernstein, J. A.; Craig, T. J.; Georgitis, J. W.; Pawankar, R.; Rachelefsky, G. S.; Settipane, R. A.; Skoner, D. P.; Stoloff, S. W. *J. Allergy Clin. Immunol.* **2008**, *122*, S1−S84.
(3) Asher, M. I.; Montefort, S.; Bjorksten, B.; Lai, C. K. W.; Strachan, D. P.; Weiland, S. K.; Williams, H. *Lancet* **2006**, *368*, 733−743.
(4) Besancenot, J. P.; Thibaudon, M. *Rev. Mal. Respir.* **2012**, *29*, 1238−1253.
(5) Deak, A. J.; Makra, L.; Matyasovszky, I.; Csepe, Z.; Muladi, B. *Sci. Total Environ.* **2013**, *442*, 36−47.
(6) Diaz, J.; Linares, C.; Tobias, A. *Aerobiologia* **2007**, *23*, 231−238.
(7) Holt, K. A.; Bennett, K. D. *New Phytol.* **2014**, *203*, 735−742.
(8) Holt, K.; Allen, G.; Hodgson, R.; Marsland, S.; Flenley, J. *Rev. Palaeobot Palyno* **2011**, *167*, 175−183.
(9) Pappas, C. S.; Tarantilis, P. A.; Harizanis, P. C.; Polissiou, M. G. *Appl. Spectrosc.* **2003**, *57*, 23−27.
(10) Fischer, G.; Braun, S.; Thissen, R.; Dott, W. *J. Microbiol. Methods* **2006**, *64*, 63−77.
(11) Gottardini, E.; Rossi, S.; Cristofolini, F.; Benedetti, L. *Aerobiologia* **2007**, *23*, 211−219.
(12) Watson, J. S.; Sephton, M. A.; Sephton, S. V.; Self, S.; Fraser, W. T.; Lomax, B. H.; Gilmour, I.; Wellman, C. H.; Beerling, D. J. *Photoch Photobio Sci.* **2007**, *6*, 689−694.
(13) Dell'Anna, R.; Lazzeri, P.; Frisanco, M.; Monti, F.; Malvezzi Campeggi, F.; Gottardini, E.; Bersani, M. *Anal. Bioanal. Chem.* **2009**, *394*, 1443−1452.
(14) Zimmermann, B. *Appl. Spectrosc.* **2010**, *64*, 1364−1373.
(15) Fraser, W. T.; Scott, A. C.; Forbes, A. E. S.; Glasspool, I. J.; Plotnick, R. E.; Kenig, F.; Lomax, B. H. *New Phytol.* **2012**, *196*, 397−401.
(16) Mularczyk-Oliwa, M.; Bombalska, A.; Kaliszewski, M.; Wlodarski, M.; Kopczynski, K.; Kwasny, M.; Szpakowska, M.; Trafny, E. A. *Spectrochim. Acta, Part A* **2012**, *97*, 246−254.
(17) Pummer, B. G.; Bauer, H.; Bernardi, J.; Chazallon, B.; Facq, S.; Lendl, B.; Whitmore, K.; Grothe, H. *J. Raman Spectrosc.* **2013**, *44*, 1654−1658.
(18) Fraser, W. T.; Watson, J. S.; Sephton, M. A.; Lomax, B. H.; Harrington, G.; Gosling, W. D.; Self, S. *Rev. Palaeobot Palyno* **2014**, *201*, 41−46.
(19) Lahlali, R.; Jiang, Y. F.; Kumar, S.; Karunakaran, C.; Liu, X.; Borondics, F.; Hallin, E.; Bueckert, R. *Front. Plant Sci.* **2014**, *5*, DOI: 10.3389/fpls.2014.00747.
(20) Zimmermann, B.; Kohler, A. *PLoS One* **2014**, *9*, e95417.
(21) Zimmermann, B.; Tkalcec, Z.; Mesic, A.; Kohler, A. *PLoS One* **2015**, *10*, e0124240.
(22) Zimmermann, B.; Bagcioglu, M.; Sandt, C.; Kohler, A. *Planta* **2015**, *242*, 1237−1250.
(23) Lukacs, R.; Blumel, R.; Zimmermann, B.; Bagcoglu, M.; Kohler, A. *Analyst* **2015**, *140*, 3273.
(24) Bhargava, R. *Appl. Spectrosc.* **2012**, *66*, 1091−1120.
(25) Kohler, A.; Kirschner, C.; Oust, A.; Martens, H. *Appl. Spectrosc.* **2005**, *59*, 707−716.
(26) Zimmermann, B.; Kohler, A. *Appl. Spectrosc.* **2013**, *67*, 892−902.
(27) Kohler, A.; Sule-Suso, J.; Sockalingum, G. D.; Tobin, M.; Bahrami, F.; Yang, Y.; Pijanka, J.; Dumas, P.; Cotte, M.; van Pittius, D. G.; Parkes, G.; Martens, H. *Appl. Spectrosc.* **2008**, *62*, 259−266.
(28) Barker, M.; Rayens, W. *J. Chemom.* **2003**, *17*, 166−173.
(29) Le Cao, K. A.; Rossouw, D.; Robert-Granie, C.; Besse, P. *Stat Appl. Genet Mol.* **2008**, *7*, Article no. 35.
(30) Karaman, I.; Qannari, E.; Martens, H.; Hedemann, M. S.; Knudsen, K. E. B.; Kohler, A. *Chemom. Intell. Lab. Syst.* **2013**, *122*, 65−77.
(31) Bassan, P.; Byrne, H. J.; Lee, J.; Bonnier, F.; Clarke, C.; Dumas, P.; Gazi, E.; Brown, M. D.; Clarke, N. W.; Gardner, P. *Analyst* **2009**, *134*, 1171−1175.
(32) Bassan, P.; Kohler, A.; Martens, H.; Lee, J.; Byrne, H. J.; Dumas, P.; Gazi, E.; Brown, M.; Clarke, N.; Gardner, P. *Analyst* **2010**, *135*, 268−277.
(33) Bassan, P.; Kohler, A.; Martens, H.; Lee, J.; Jackson, E.; Lockyer, N.; Dumas, P.; Brown, M.; Clarke, N.; Gardner, P. *Journal of biophotonics* **2010**, *3*, 609−620.
(34) Mohlenhoff, B.; Romeo, M.; Diem, M.; Wood, B. R. *Biophys. J.* **2005**, *88*, 3635−3640.
(35) van Dijk, T.; Mayerich, D.; Carney, P. S.; Bhargava, R. *Appl. Spectrosc.* **2013**, *67*, 546−552.
(36) Krimm, S.; Liang, C. Y.; Sutherland, G. B. B. M. *J. Chem. Phys.* **1956**, *25*, 549−562.
(37) Ronneberger, O.; Schultz, E.; Burkhardt, H. *Aerobiologia* **2002**, *18*, 107−115.
(38) Treloar, W. J.; Taylor, G. E.; Flenley, J. R. *J. Quaternary Sci.* **2004**, *19*, 745−754.
(39) Dhawale, V. R.; Tidke, J. A.; Dudul, S. V. *Proceedings of the 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Mysore, India, August 22−25, 2013; pp 79−84.
(40) Ticay-Rivas, J. R.; del Pozo-Banos, M.; Travieso, C. M.; Arroyo-Hernandez, J.; Perez, S. T.; Alonso, J. B.; Mora-Mora, F. *Ifip Adv. Inf Comm Te* **2011**, *364*, 342−349.
(41) Stillman, E. C.; Flenley, J. R. *Quat. Sci. Rev.* **1996**, *15*, 1−5.
(42) Kaya, Y.; Erez, M. E.; Karabacak, O.; Kayci, L.; Fidan, M. *Grana* **2013**, *52*, 71−77.
(43) Tanaka, N.; Uehara, K.; Murata, A. *J. Plant Res.* **2004**, *117*, 265−276.
(44) Galan, C.; Smith, M.; Thibaudon, M.; Frenguelli, G.; Oteros, J.; Gehrig, R.; Berger, U.; Clot, B.; Brandao, R.; Grp, E. Q. W. *Aerobiologia* **2014**, *30*, 385−395.

DOI: 10.1021/acs.analchem.5b03208
Anal. Chem. 2016, 88, 803−811

Supporting Information

Analysis of allergenic pollen by FTIR microspectroscopy

B. Zimmerman*, V. Tafintseva, M. Bağcıoğlu, M. Høegh Berdahl and A. Kohler

Department of Mathematical Sciences and Technology, Faculty of Environmental Science and Technology, Norwegian University of Life Sciences, 1432 Ås, Norway

*Corresponding author:
Boris Zimmermann
Department of Mathematical Sciences and Technology
Faculty of Environmental Science and Technology
Norwegian University of Life Sciences
Drøbakveien 31, 1432 Ås, Norway.
Tel: +47 6723 1576
Faks: +47 6496 5001
E-mail: boris.zimmermann@nmbu.no

Table of Contents **Page**

Sample measurement

The pollen samples were embedded in a soft paraffin between two sheets of 40 µm thick polyethylene foils, and placed in a bespoke sample holder. A sample ready for µFTIR measurement is shown in Fig. S-1. As can be seen in Figs. S-1C and S-1D, µFTIR setting enables measurement of individual pollen grain. For the presented study, only physically separated pollen grains were measured, with particular attention to avoid grains with depth overlap (z-axis stacking of grains). Therefore, each of the 360 recorded spectra belonged to a different individual grain.

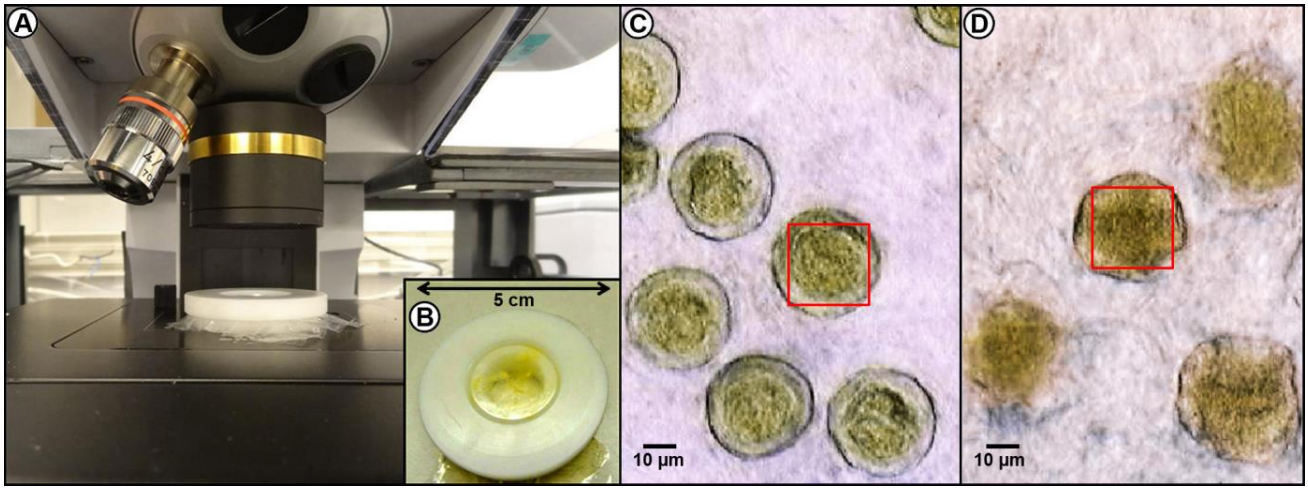


Fig. S-1. (**A**) Pollen sample on a moving bench and under objective of a HYPERION 3000 FTIR microscope, and (**B**) Sample holder with the prepared pollen sample in the polyethylene/paraffin embedding matrix. Optical microscope images taken during sample measurement with FTIR microscope: (**C**) *Cryptomeria japonica* grains, and (**D**) *Carpinus betulus* grains (out of focus grains are visible in lower left and upper right corners); 25×25 µm aperture of FTIR microscope is designated by the red square.

The measurement of pollen grains in the embedding medium is not as straightforward as the measurement of pollen grains on a microscope slide. Since bioparticles in the embedding medium are arranged within a three-dimensional matrix, it is improbable that they all have a common focal plane parallel with the foils (i.e. parallel with microscope stage). Usually the embedded bioparticles are arranged in different focal planes, and thus the focus of the FTIR microscope needs to be adjusted multiple times to measure all of the bioparticles. Fig. S-2 shows µFTIR spectra of pollen grains that were recorded using a single focal plane. Out of the ten presented spectra, half were out-of-focus. The out-of-focus spectra are easily distinguished by insignificant difference in absorbance values between the region were chemical absorbance of pollen is expected ('*Amide I*' region between 1700 and 1600 cm^{-1}) and the region were pollen grain does not have chemical absorbance ('*Baseline*' region between 2100 and 2000 cm^{-1}). Thus, the associated pollen grains were measured once more using a different focal plane in order to obtain high quality spectra. Quality of recorded spectra can be estimated by running the standard spectral

quality tests, as assessed in this study by calculating the signal-to-noise ratio of the Amide I signal as defined by OPUS/IR user manual.

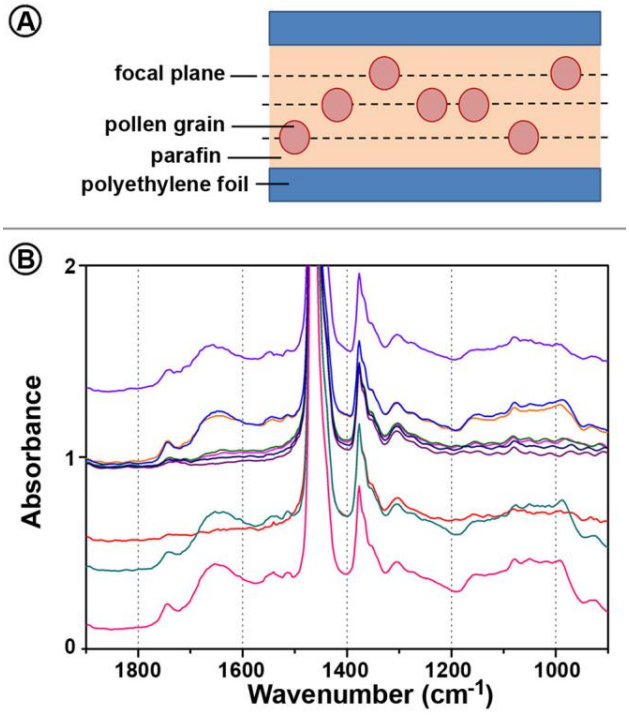


Fig. S-2. (**A**) Schematic drawing of µFTIR measurements of pollen within paraffin matrix, showing arrangement of pollen grains in three different focal plains. (**B**) µFTIR spectra obtained by measurement in paraffin and polyethylene embedding medium of ten individual *Betula pendula* pollen grains; half of the grains were in focus and half were out of focus during measurement

Spectral pre-processing: EMSC model

The EMSC model used in the pre-processing is defined by Eq. S-1, and the EMSC corrected spectra by Eq. S-2.

$$\mathbf{z}_i = b_i\mathbf{m} + a_i\mathbf{1} + d_i\tilde{\boldsymbol{\nu}} + e_i\tilde{\boldsymbol{\nu}}^2 + f_i\tilde{\boldsymbol{\nu}}^3 + \boldsymbol{\varepsilon}_i \quad \text{(S-1)}$$

$$\mathbf{z}_{i,\mathrm{Corr}} = (\mathbf{z}_i - a_i\mathbf{1} - d_i\tilde{\boldsymbol{\nu}} - e_i\tilde{\boldsymbol{\nu}}^2 - f_i\tilde{\boldsymbol{\nu}}^3) / b_i \quad \text{(S-2)}$$

where $\mathbf{z}_i$ is a 'measured' spectrum (where i = 1,…, N; number of spectra in a data set N is 100), $\mathbf{z}_{i,\mathrm{Corr}}$ is a corrected spectrum, b_i is a multiplicative parameter, $\mathbf{m}$ is a reference spectrum often presented by the mean spectrum for a given data set, $\mathbf{1}$=[1, 1, 1, . . ., 1], a_i, d_i, e_i and f_i are constant, linear, quadratic and cubic parameters respectively, $\boldsymbol{\varepsilon}_i$ is a residual term, $\tilde{\boldsymbol{\nu}}$ is spectral range (wavenumbers).

Chemical variability of pollen grains

Pollen grains have unique vibrational spectra. Although pollen grains of a given type were collected from the same maternal plant, each grain within the sample is one genetically unique organism. Pollen development and its chemical composition involves contributions from both the sporophytic (maternal plant) and gametophytic (pollen) genomes [1,2]. Therefore, each spectrum represents one biologically independent sample.

For principal component analysis (PCA), second derivative spectra were pre-pre-processed by EMSC as described in the section about spectral pre-processing (dataset A). PCA analysis (Fig. S-3) shows substantial differences in the biochemical composition between pollen of angiosperms (Fagales and Poales) and conifers (Pinales). The previous studies have shown that the cypress (Cupressaceae) family, to which all four measured conifer taxa belong, has pollen with distinct spectral features [3,4]. While the IR spectra of Cupressaceae pollen show dominant absorption due to vibrational bands in the 'carbohydrate region' (1200–900 cm^{-1}), pollen of angiosperms show strong IR absorption of amide I and amide II bands in the 'protein region' (1700–1520 cm^{-1}). Moreover, both pollen types differ in absorption of lipid related carbonyl stretch in the 'lipid region' (1750-1700 cm^{-1}), indicating differences in the chemical composition of pollen lipids. Cupressaceae pollens show strong absorption at approx. 1730 cm^{-1}, while angiosperms show strong absorption at approx. 1745 cm^{-1}.

Finally, PCA plots show well defined clustering at lower taxonomic levels, including species level. Of particular interest is a single clustering of all three *Betula pendula* samples. By using the first four principal components, spectra of samples belonging to distinct plant species can be separated quite well from each other. Therefore, µFTIR spectra of pollen show large potential for classification analysis.

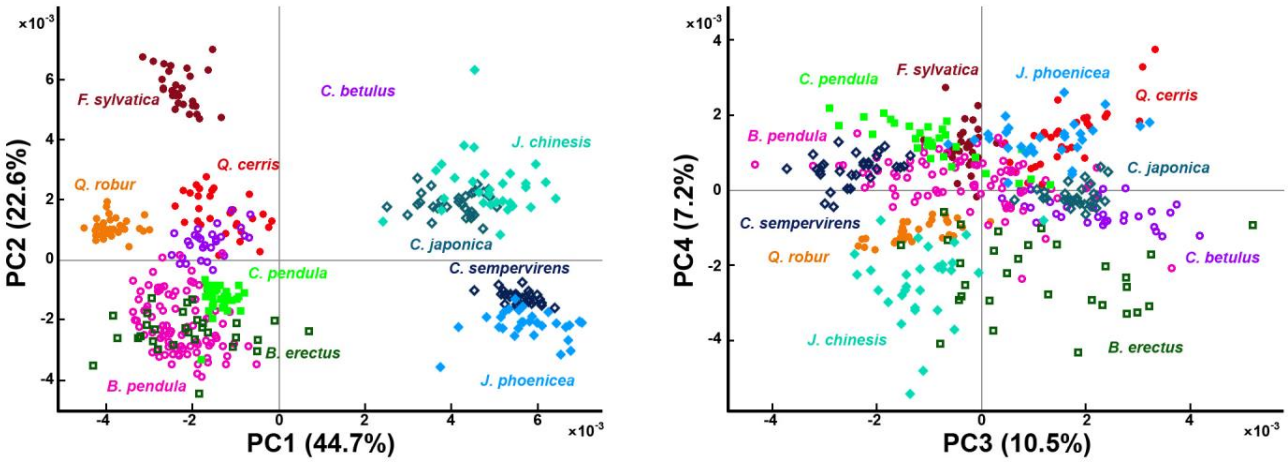


Fig. S-3. PCA score plots of a µFTIR spectral data set (30 spectra per sample; second derivative and EMSC corrected spectra – dataset A), with depiction of plant species. The percent explained variances for the first five PCs are 44.7, 22.6, 10.5, 7.2 and 3.8.

Physical variability of pollen grains

As stated above, a µFTIR spectrum of a pollen grain includes, along with absorption signals of its chemical constituents, physical effects such as scattering distortions. These effects depend on sample morphology and the refractive index. In addition, variations in sample thickness introduce strong multiplicative effects. Pre-processing methods, such as derivative transformation of recorded data and extended multiplicative signal correction (EMSC), are often applied in order to reduce multiplicative and scatter effects, and to enhance chemical interpretation of the spectra.

EMSC is a model-based pre-processing method that allows explicit parameterization of physical information. It enables the separation of physical and chemical information from the spectral data [5] . In order to consider physical information, EMSC parameters (dataset B) were obtained from the non-derivated spectra.

Although EMSC estimates, containing baseline, background and other anomalous features, are often discarded in spectral analysis, they may contain information that is important for sample characterization [5]. This is valid, in particular, for characterisation of pollen since physical information is closely related to a characteristic morphology of a bioparticle [4]. Pollen morphology varies substantially in shape, size, and wall pattern across plant taxa. The chemical composition of pollen is complex, due to chemical variation of pollen grain substructures, such as grain wall layers (exine and intine), different type of cells (tube and generative cells) and intercellular material (cytoplasm and nuclei). This chemical composition of pollen varies substantially across plant taxa as well [3]. Therefore, we expect that both chemical and physical information are of value for discrimination and classification of pollen.

Fig. S-4 shows the separation of physical and chemical information in µFTIR spectra of pollen by an EMSC model with constant, linear, quadratic, and cubic parameters. The EMSC separates the recorded spectra (Fig. S-4A) into a "corrected chemical spectrum" including chemical absorption bands (Fig. S-4B), and EMSC estimates that contain physical information (Fig. S-4C). The EMSC scatter estimates include the complete "baseline" estimate (wavelike pattern in Fig. S4-C which is composed of the constant, linear and quadratic term of the EMSC model in Eq. S-2), and the effective optical path length (multiplicative parameter *b* in Fig. S-4C).

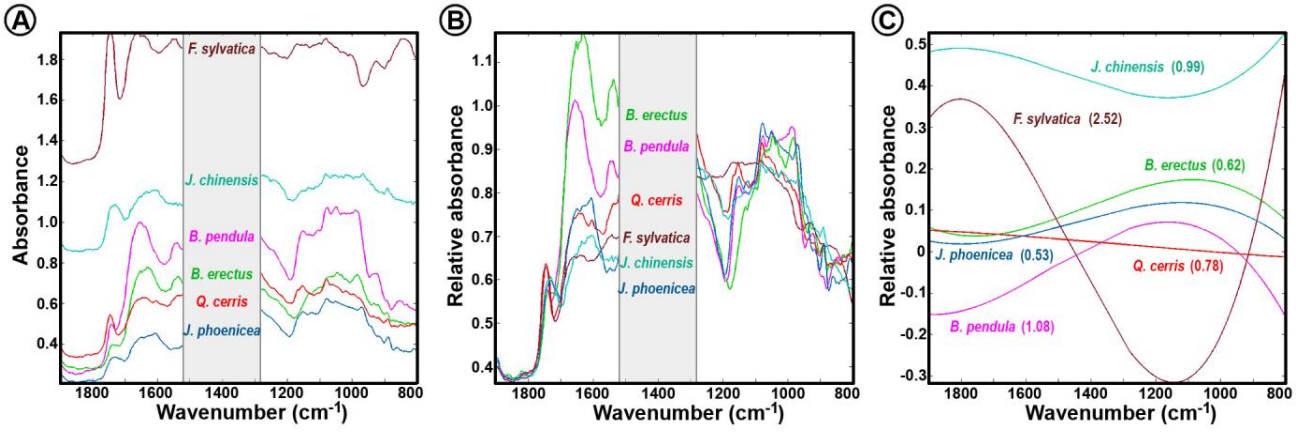


Fig. S-4. µFTIR spectra of representative pollen grains of six different pollen species: (**A**) recorded spectra, (**B**) spectra after EMSC pre-processing, and (**C**) EMSC estimates (values of multiplicative parameter *b* are in parenthesis).

The graphical representation of EMSC parameters (dataset B) clearly shows that EMSC estimates are not completely random but to some extent specific for each taxon (Fig. S-5). Similar taxon-specific results for EMSC estimates were obtained previously on bulk measurement of spore and pollen samples by single reflectance attenuated total reflectance (SR-ATR) FTIR [4]. It is expected, that a number of factors connected to pollen morphology of a measured pollen have impact on the resulting µFTIR spectrum, even when chemical absorption is disregarded.

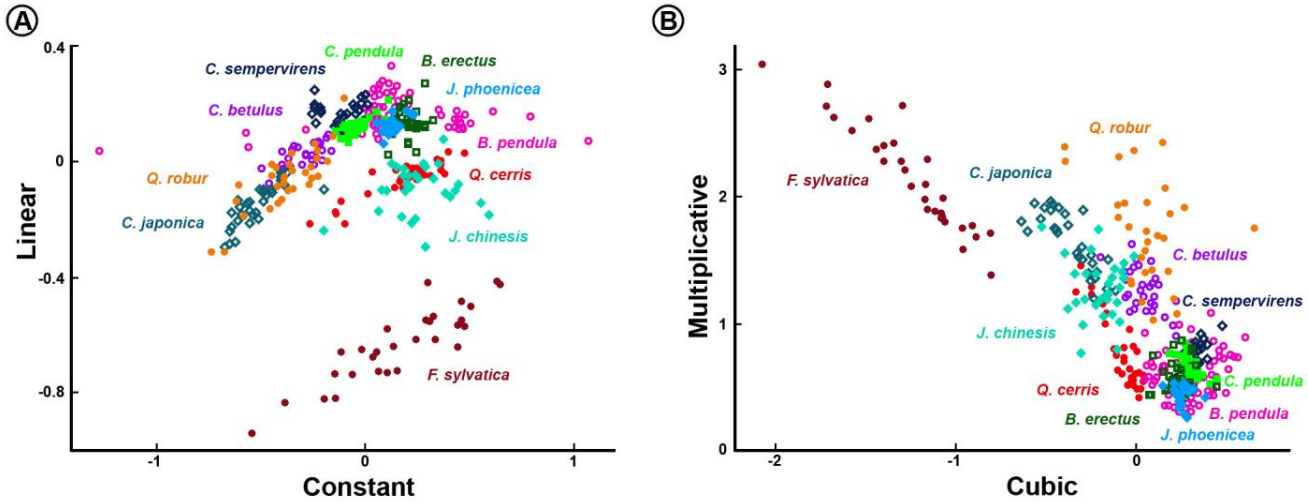


Fig. S-5. Plots of EMSC parameters (dataset B) of µFTIR spectra (30 spectra per sample; EMSC corrected non-derivated spectra), with depiction of plant species: (**A**) linear (*d*) and constant (*a*) parameters, and (**B**) multiplicative (*b*) and cubic (*f*) parameters.

One important factor is path length of infrared light through a pollen grain. Larger grains have longer optical path length, resulting in higher absorption values. The size of the measured grains varied from 15-25 µm in diameter for Pinales, to 25-35 µm for Fagales and Poales, with the exception of *Fagus sylvatica* that has significantly larger pollen grains with diameters of approx. 45 µm. It can be seen that *F. sylvatica* pollen has completely different spectral scatter estimates compared to the rest of the samples (Fig. S-5). Although both *F. sylvatica* and *Quercus cerris* have rather similar chemical composition (Fig. S-3), characterised by large relative content of triglyceride lipids [3], their recorded spectra differ considerably regarding total absorbance (Fig. S-4). The high total absorbance for *F. sylvatica* grains is due to the relatively large size of *F. sylvatica* grains. The high absorbance values of *F. sylvatica* spectra indicate that the major part of the infrared light is completely absorbed by *F. sylvatica* grains.

One additional important factor is the variability of optical properties of pollen due to variation in chemical composition of pollen grain substructures. For instance, the two measured *Juniperus* species, *J. chinensis* and *J. phoenicea*, share similar morphology of pollen grains. However, their chemical composition is quite different, since *J. chinensis* has considerably larger relative content of triglyceride lipids than *J. phoenicea* [6]. Triglyceride lipids are stored

predominantly within the pollen grain, in the form of intracellular spherical droplets, termed lipid bodies [1]. Triglycerides exhibit strong absorption of infrared light, and therefore the difference in the relative content of triglyceride will lead to a significant difference related to multiplicative effect between the two *Juniper* species. Moreover, it can be seen that the two *Juniper* species show large differences in the spectral baseline (Fig. S-4). The baseline differences are most probably related to: 1) differences in optical densities due to differences in the constant part of the refractive index, and 2) differences in scattering properties due to Rayleigh scattering on lipid bodies and other organelles.

In general, EMSC estimates of pollen µFTIR spectra have a range of taxon-specific values. As such, they offer an indirect physical measurement of a pollen sample that should be used for pollen characterization, alongside chemical absorption spectrum.

Discrimination analysis

Sensitivity and specificity of discrimination analysis for spectral data with included EMSC parameters (dataset C) is presented in Figure S-6.

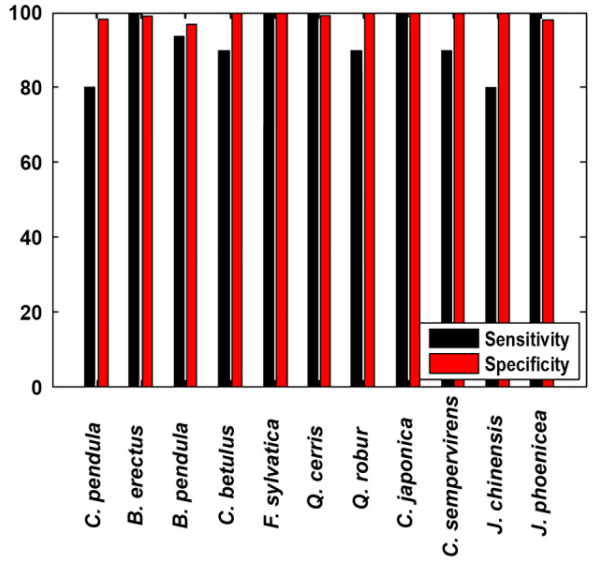


Fig. S-6. Identification result for Sparse PLSR analysis of pollen µFTIR spectra, showing bar plots of sensitivity and specificity for each species (based on 10 spectra per sample). The result is based on second derivative spectral data with included EMSC parameters (dataset C).

Supporting Information References:

(1) Piffanelli, P.; Ross, J. H. E.; Murphy, D. J. *Sex Plant Reprod* **1998**, *11*, 65-80.
(2) Zhang, M.; Fan, J. L.; Taylor, D. C.; Ohlrogge, J. B. *Plant Cell* **2009**, *21*, 3885-3901.
(3) Zimmermann, B.; Kohler, A. *Plos One* **2014**, *9*, e95417.
(4) Zimmermann, B.; Tkalcec, Z.; Mesic, A.; Kohler, A. *Plos One* **2015**, *10*, e0124240.
(5) Kohler, A.; Kirschner, C.; Oust, A.; Martens, H. *Applied spectroscopy* **2005**, *59*, 707-716.
(6) Zimmermann, B. *Applied spectroscopy* **2010**, *64*, 1364-1373.

6.4 Paper IV

Monitoring of plant-environment interactions by high throughput FTIR spectroscopy of pollen

Murat Bağcıoğlu[1], Achim Kohler[1,2], Stephan Seifert[3,4], Janina Kneipp[3,4], Boris Zimmermann[1*]

[1]Department of Mathematical Sciences and Technology, Faculty of Environmental Science and Technology, Norwegian University of Life Sciences, 1432, Ås, Norway
[2]Nofima AS, 1430, Ås, Norway
[3]Humboldt-Universität zu Berlin, Department of Chemistry, 12489 Berlin, Germany
[4]BAM Federal Institute for Materials Research and Testing, 12489 Berlin, Germany
*** Corresponding author**
E-mail: boris.zimmermann@nmbu.no. Tel.: +47 6723 1576. Fax: +47 6496 5001.

ABSTRACT

Fourier transform infrared (FTIR) spectroscopy enables chemical analysis of pollen samples for plant phenotyping to study plant-environment interactions, such as influence of climate change or pathogens. However, current approach, such as microspectroscopy and Attenuated Total Reflection spectroscopy, does not allow for high-throughput protocols. The paper at hand suggests a new spectroscopic method for high-throughput characterization of pollen. In this approach, samples are measured as thin films of pollen fragments using a Bruker FTIR spectrometer with an eXTension (HTS-XT) unit employing 384-well plates. Critical steps in the sample preparation and measurement, such as variabilities between technical replicates, between microplates and between spectrometers, were studied. Measurement variations due to sample preparation, microplate holders and instrumentation were low, and thus allowed differentiation of samples with respect to phylogeny and biogeography. Biochemical composition of 31 species of Fagales (beech order), collected during three different pollination seasons (2012-2014) at three different locations (Germany, Croatia and Norway) were studied. The spectral variability for a range of Fagales species (*Fagus*, *Quercus*, *Betula*, *Corylus*, *Alnus* and *Ostrya*) showed high species-specific differences in pollen's chemical composition due to either location or pollination season. Significant inter-annual, intra-seasonal and locational differences in the pollen spectra indicate that pollen chemical composition has high phenotypic plasticity and is influenced by local climate conditions. The variations in composition are connected to lipids, proteins, carbohydrates and sporopollenins that play crucial roles in cold and desiccation tolerance, protection against UV radiation, and as material and energy reserves.

INTRODUCTION

The vast majority of land plants are sporophyte-dominant seed plants that have male gametophytes with microscopic sizes and short lifespans. However, pollen is crucial in reproduction and survival of seed plants since it enables production of genetically diverse offspring, and thus it is exposed to a diversity of selection pressures (Delph *et al.* 1997, Williams and Mazer 2016). Climate effects, including global climate change, can have a strong detrimental effect on reproductive structures of plants, including pollen. For example, temperature stress can have a damaging effect on development, transport and germination of pollen (Bokszczanin *et al.* 2013, Chakrabarti *et al.* 2011, Endo *et al.* 2009). Pollen development has been pointed to as the most heat-sensitive process in plant sexual reproduction (Bokszczanin, Fragkostefanakis and Thermotolerance 2013). An increase in sterility of pollen results in lower seed yield, and thus significantly reduced population resilience as well as higher economic damages to agricultural crops.

It has been hypothesized that phenotypic plasticity will play an important role in the persistence of plants in increasingly variable climatic conditions (Anderson *et al.* 2012, Nicotra *et al.* 2010). Regarding pollen, it is critical to estimate plasticity of the most important traits such as longevity, viability, hydration and germination speed, as well as pollen tube growth rate and length. The majority of these traits should have correlation with chemical composition of pollen grains, for example with nutrient reserves in form of triglycerides and polysaccharides that play crucial role in pollen tube growth (Rodriguez-Garcia *et al.* 2003, Speranza *et al.* 1997). Therefore, to monitor climate-effects on plant communities it is imperative to improve phenotyping of pollen. Standard pollen analysis by optical microscopy can provide only characterization based on morphological analysis, while chemical information and information on pollen quality cannot be obtained. Infrared spectroscopy is one of the most widely used methods for chemical characterization of samples, and in the last decade, rapid progression in infrared studies of plant material was conducted.

Development and implementation of Fourier transform infrared (FTIR) spectroscopy of pollen is a novel approach with the potential to greatly expand palynological and botanical research. Infrared spectra of pollen are precise fingerprints of the overall biochemical composition of a pollen grain, and contain specific signals of lipids, proteins, carbohydrates, cell wall

biopolymers, and water. The spectra can be obtained by using diverse techniques, such as single- (Bağcıoğlu *et al.* 2015, Gottardini *et al.* 2007, Jiang *et al.* 2015, Lahlali *et al.* 2014, Zimmermann 2010, Zimmermann and Kohler 2014, Zimmermann *et al.* 2015b) and multi-reflection (Mularczyk-Oliwa *et al.* 2012) attenuated total reflection (ATR) FTIR, diffuse reflectance FTIR (DRIFT) (Mularczyk-Oliwa, Bombalska, Kaliszewski, Wlodarski, Kopczynski, Kwasny, Szpakowska and Trafny 2012, Pappas *et al.* 2003), transmission FTIR measurements between windows (Crowe *et al.* 1989, Sowa *et al.* 1991) and in pellets (Bağcıoğlu, Zimmermann and Kohler 2015, Mularczyk-Oliwa, Bombalska, Kaliszewski, Wlodarski, Kopczynski, Kwasny, Szpakowska and Trafny 2012, Pappas, Tarantilis, Harizanis and Polissiou 2003, Zimmermann 2010, Zimmermann and Kohler 2014), FTIR microspectroscopy measurements of single pollen grains (Bağcıoğlu, Zimmermann and Kohler 2015, Gottardini, Rossi, Cristofolini and Benedetti 2007, Pappas, Tarantilis, Harizanis and Polissiou 2003, Pummer *et al.* 2013, Zimmermann *et al.* 2015a, Zimmermann and Kohler 2014, Zimmermann *et al.* 2016) as well as multi grains (Bağcıoğlu, Zimmermann and Kohler 2015, Dell'Anna *et al.* 2009, Wolkers and Hoekstra 1995, Wolkers and Hoekstra 1997), and photoacoustic FTIR spectroscopy (Parodi *et al.* 2013). Recent studies have shown that FTIR spectroscopy achieves simple, economical, and rapid identification and classification of pollen by spectral fingerprinting, as well as characterization and biochemical interpretation with respect to environmental stress (Jiang, Lahlali, Karunakaran, Kumar, Davis and Bueckert 2015, Lahlali, Jiang, Kumar, Karunakaran, Liu, Borondics, Hallin and Bueckert 2014, Zimmermann and Kohler 2014). The identification and classification of samples is mostly based on infrared spectral databases, thus offering an unbiased and operator-independent approach with huge potential for automatization (Dell'Anna, Lazzeri, Frisanco, Monti, Malvezzi Campeggi, Gottardini and Bersani 2009, Pappas, Tarantilis, Harizanis and Polissiou 2003, Zimmermann 2010, Zimmermann, Tafintseva, Bagcioglu, Hoegh Berdahl and Kohler 2016). Our recent study on aeroallergen pollen by FTIR microspectroscopy has shown that the FTIR has great potential for an automated pollen analysis (Zimmermann, Tafintseva, Bagcioglu, Hoegh Berdahl and Kohler 2016). Moreover, since the method is compatible with standard air-samplers, it can be employed in the assessment and control of bioaerosols, including allergy forecasts.

Standard methodologies, including microscopies and sequencing, are overly complex, expensive and time-consuming. Therefore, they are slow in providing the massive data needed for

92 determining and monitoring effects of abiotic and biotic stress on plant communities. Since FTIR
93 spectroscopy supports high-throughput measurements, the technique is ideal for rapid screening of
94 plant populations under different conditions. A high throughput screening (HTS) FTIR system has
95 been successfully applied in a variety of studies in natural and biomedical sciences for the
96 investigation of microorganisms and tissues (Dean *et al.* 2010, Kohler *et al.* 2015, Mignolet and
97 Goormaghtigh 2015, Ollesch *et al.* 2013, Sirikwanpong *et al.* 2010). An extension of this approach
98 on studies of pollen would have a great potential, particularly for comprehensive monitoring of
99 terrestrial ecosystems. An HTS FTIR methodology has never been used in pollen analysis. Our
100 previous study, on comparative analysis of FTIR techniques, has shown that spectra of ground
101 pollens, obtained by either ATR or in pellets, is optimal for gathering information on their chemical
102 composition (Bağcıoğlu, Zimmermann and Kohler 2015). Thus, a high-throughput FTIR approach
103 where samples are prepared by pulverization of pollen into water suspension and by subsequent
104 drying as films, should offer an analogous result. In the paper at hand, we explore the usage and
105 development of a new high-throughput FTIR-based protocol for the biochemical characterization
106 of pollen. The FTIR study was conducted on a substantial set of samples, consisting of six different
107 genera of Fagales (*Fagus*, *Quercus*, *Betula*, *Corylus*, *Alnus* and *Ostrya*), collected during the three
108 different pollination seasons (2012-2014) at three different locations (Germany, Croatia and
109 Norway). The potential of high-throughput FTIR spectroscopy of pollen as a tool for studying plant
110 communities and plant-environment interactions is elucidated.

111 **RESULTS AND DISCUSSION**

112 **Effects of sample preparation**

113 The high throughput FTIR unit employs silicon microtiter plates as sample holders for thin
114 films of pollen samples, created by drying of pollen suspensions. Following the preparation of
115 films, the fully automated transmission measurement of the microplate was performed. The
116 microtiter plate has 384 positions (wells) for sample suspensions, thus 127 samples can be
117 measured in a single run, assuming triplicate measurements per sample and an empty position for
118 background measurement.

119 Recording of an optimal spectrum critically depends on the preparation of the dry sample
120 film. Our previous study has demonstrated that pulverization and homogenization of pollen by
121 grinding obtains more comprehensive information on grain chemistry than measurement of intact

pollen.(Bağcıoğlu, Zimmermann and Kohler 2015) Moreover, intact pollen grains can create saturation effects when measured in transmission mode (Zimmermann, Bagcioglu, Sandt and Kohler 2015a). The saturation effect happens when the central part of the grain is opaque for a large part of the infrared spectrum of the light, and the small fraction of light that reaches the detector has passed through the periphery areas of pollen grain (i.e. grain wall). The resulting spectrum has high baseline contribution and oversaturation of signals, while the spectral information is predominantly obtained from the grain wall. Lastly, intact pollen grains can cause wavelength-dependent scattering effects since the sizes of pollen grains are at the same order as the wavelength range employed in the infrared spectroscopy measurements.

The covered Fagales pollen have grain sizes within 25-50 μm range, and the first step of sample preparation was grain breakdown by sonication treatment. Pollen grains have a double-layered grain wall composed predominantly of natural polymers: cellulose and sporopollenin. The grain wall has exceptional mechanical and chemical resilience and thus offers excellent protection from ambient effects. In particular, the outer layer of pollen grain wall (exine) offers vital protection to physical stress due to remarkable properties of sporopollenins, the highly resistant and complex biopolymers. Therefore, relatively strong disruptive and prolonged force needs to be applied in order to obtain complete breakdown of pollen samples. As can be seen in Figures 1 and 2, incomplete fragmentation of pollen is obtained after short sonication times, resulting in low quality spectra. On the other hand, prolonged sonication treatment results in extensive fragmentation of sporopollenous exine and homogenization of intracellular material (Figure 1B).

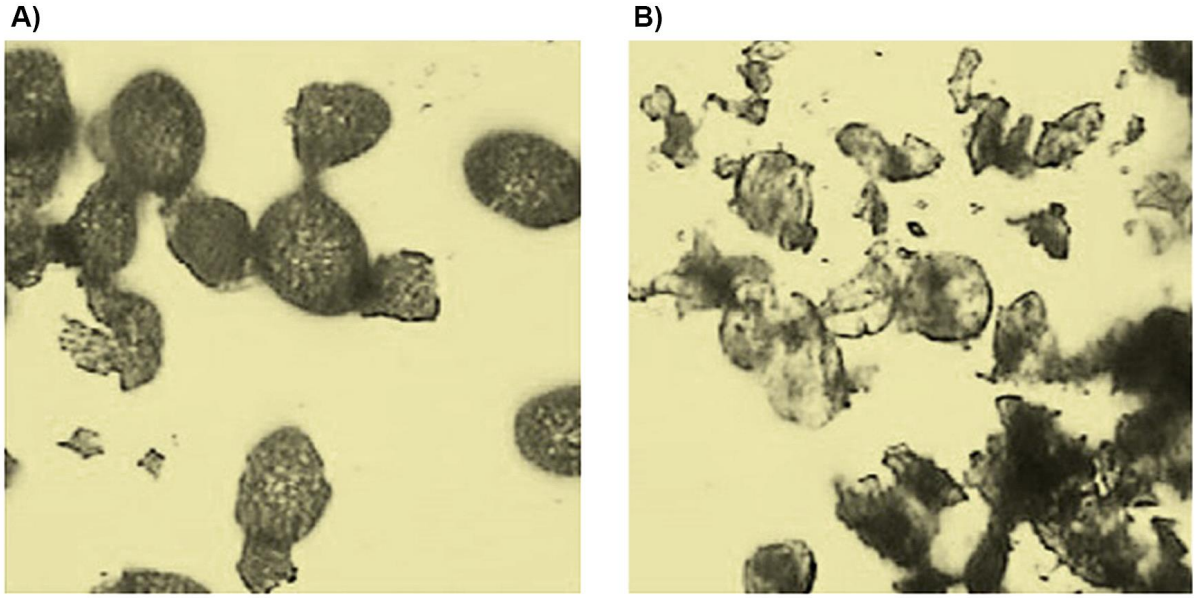


Figure 1. Optical microscopic images of *Quercus robur* (Pedunculate oak) pollen grains with respect to different sonication treatment: A) 1 min, and B) 2 min.

Figure 2 shows the effects of the sonication treatment on the quality of the FTIR spectra. In general, a high quality FTIR spectrum has a flat baseline positioned around zero absorbance and high signal-to-noise ratio (Figure 2). Since a sample film is acquired from 5-10 µL of a liquid sample, the exact quantity and concentration of pollen suspension should be adjusted for obtaining high quality spectra. Ideally, the pollen IR spectra should display an absorbance signal of the amide I band within 0.4-1.0 absorbance range (Kohler, Bocker, Shapaval, Forsmark, Andersson, Warringer, Martens, Omholt and Blomberg 2015). Amide I is used as a reference signal since it is usually the strongest absorption band in the FTIR spectra of biological cells. In the principal study, the pollen suspension was prepared by mixing approx. 1 mg of pollen sample with 500 µL of water, which is the minimal liquid amount needed for sonication. Following the sonication, the pollen suspension was centrifuged, concentrated by removing 400 µL of supernatant, and using 8 µL of the remaining suspension for the preparation of sample film, resulting in high quality FTIR spectra (Figure 2). The red spectrum in Figure 2 shows characteristic signals for pollen chemical constituents: carbohydrate at 1200-900 cm^{-1} (C-O-C, C-C and C-O stretching vibrations), sporopollenin at 1605, 1515, 1170 and 835 cm^{-1} (all vibrations are related to phenyl ring vibrations), lipid signals at 1745 (C=O stretch in esters), 1705 (C=O stretch in carboxylic acids), 1470 (CH_2 deformation), and 1170 cm^{-1} (C-O-C stretch), and protein signals at 1660-1630 (amide

I:C = O stretch) and 1555-1530 (amide II: NH deformation and C–N stretch) (Gottardini, Rossi, Cristofolini and Benedetti 2007, Zimmermann 2010, Zimmermann, Tkalcec, Mesic and Kohler 2015b). In addition, the spectrum shows signals at 3050 cm^{-1} (=C-H stretch) and 2960-2830 cm^{-1} (C-H stretch in $–CH_3$ and $–CH_2–$) that are predominantly associated with lipids, and in a smaller degree with carbohydrates and proteins as well.

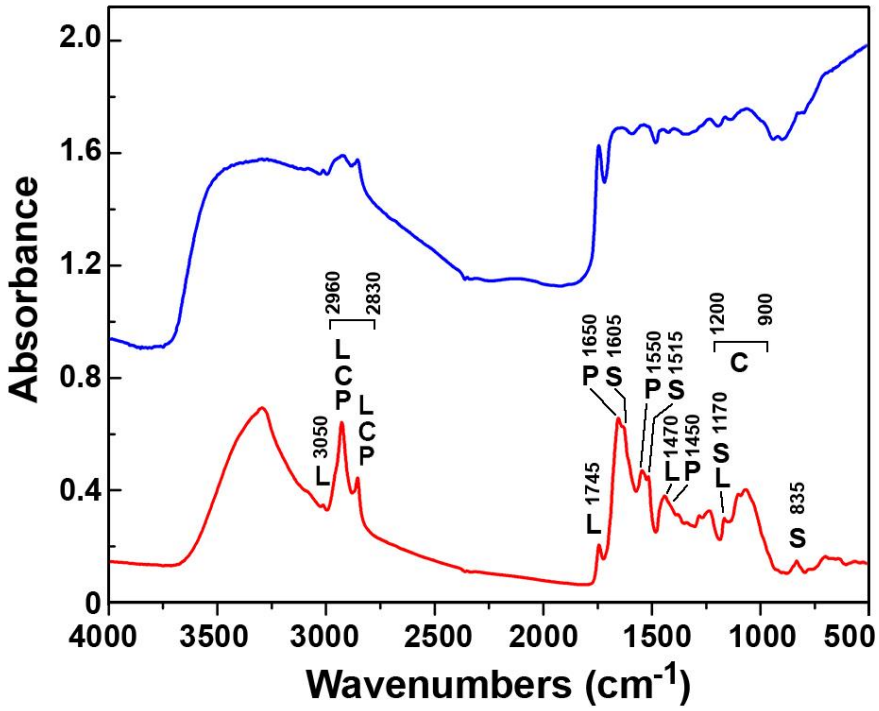


Figure 2. FTIR spectra of *Alnus glutinosa* (Black alder) pollen showing effect of sample preparation. Spectra obtained after: 1 min of sonication (top, blue) and 2 min of sonication (down, red). The marked bands are associated with molecular vibrations of (P) proteins, (L) lipids, (C) carbohydrates and (S) sporopollenins.

Measurement variability in high throughput FTIR

Dry films can have variations in morphology, including thickness and texture. This leads to noticeable baseline and absorbance variations in their FTIR spectra. Although these variations can be significant, they are predominantly additive and multiplicative effects, and thus can be readily suppressed by spectral pre-processing. In this study, we have converted spectra into second derivatives followed by EMSC, resulting in relatively small variation between replicate measurements. Conversion into second derivatives enhances chemical spectral features and

suppresses baseline variations, while EMSC enables correction of wavelength-dependent scattering effects and thus provides better separation of physical (i.e. scattering) and chemical information (Zimmermann and Kohler 2013).

In order to investigate remaining measurement variability (after preprocessing) and to compare it to chemical variability between the samples, a measurement variability test was performed. It was expected that sonication pre-treatment and dry film preparation are the critical steps in high throughput measurement of pollen. Substantial variations in any of these two steps could considerably influence spectral reproducibility. Therefore, the measurement variability was examined regarding variability between sonication preparations and between technical replicates. The variability between sonication preparations refers to independently prepared suspensions of the same pollen sample by using the same sonication procedure. Technical replicates refer to preparation of dry films, i.e. to repeated FTIR measurements using the same pollen suspension that was applied to different sample positions on the silicon microplate. In addition, variability between microplates and between spectrometers was estimated as well. The variability between microplates was estimated by measurement of the same sample suspension on different silicon microtiter well plates, while the instrument variability was estimated by measuring the same sample plates with two FTIR spectrometers.

The Pearson Correlation Coefficient (PCC), expressed as 1-PCC, was used to estimate the spectral variability of pollen samples (Table 2). The PCC measures correlation between variables, where a value of 1 indicates high positive correlation. Therefore, small variability is indicated by small 1–PCC values. The PCC was calculated for two infrared regions: 3200-2800 and 1800-1500 cm^{-1}.

As can be seen in Table 2 and Figure 3, the measurement variability is relatively small. The biggest source of variation are technical replicates, i.e. imperfections in preparation of dry films on a microplate.

The variations between measurement microplates, as well as between sample preparations by sonication, were smaller then for the variation caused by technical replicates. The difference between instruments was small, with slightly better precision for newer and more advanced instrument #2 (VERTEX 70 spectrometer). In general, the measurements reproducibility for a

given pollen sample is high compared to variability between different pollen samples. Therefore, the new high throughput FTIR approach can be considered as robust.

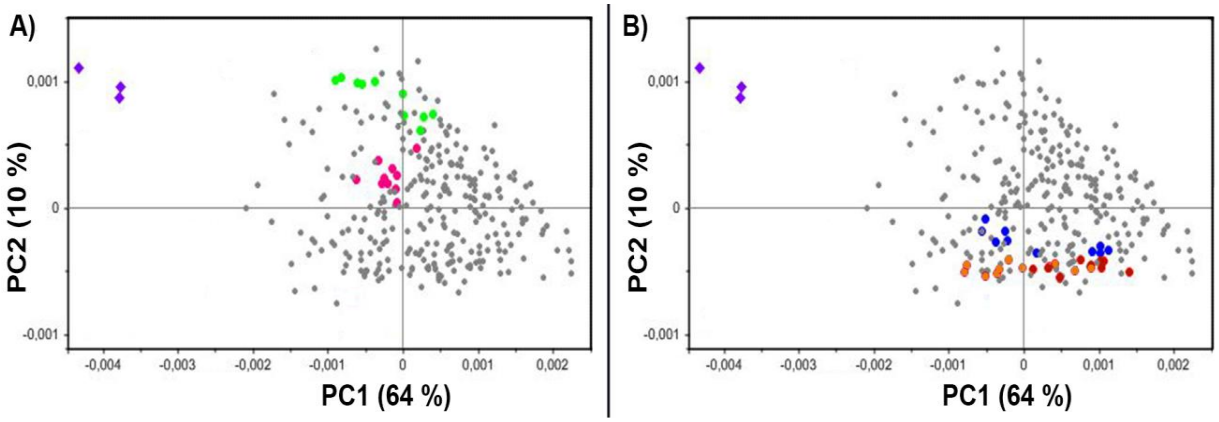


Figure 3. (A) Variability between the two independently prepared pollen suspensions of the same sample (green and magenta); each sample includes 10 technical replicates. (B) Variability between the two microplates (orange and dark red) and between the two instruments (dark red and blue); each measurement includes 10 technical replicates. In addition, the PCA score plots include 300 measurements of a *sonication variability test plate* and a *microplate variability test plate* in two instruments (dots). The control sample (purple diamonds) is pollen sample from different *Alnus incana* tree, belonging to the same population and collected on the same day as the sample for the variability study.

Table 1. Variability for different technical replicates, sonications, microtiter plates, instruments, taxa, and pollination seasons. # designates number of spectra used in the variability tests.

Type of variability	3200-2800 cm^{-1} (1-*PCC**)x10^{-4}	1800-1500 cm^{-1} (1-*PCC**)x10^{-4}	Number of spectra
Technical replicates; instrument #1 [1]	0.70±0.40	3.09±2.47	100
Sonications; instrument #1	0.37	1.23	10
Microplates; instrument #1	0.31	1.75	5
Technical replicates; instrument #2 [1]	0.43±0.25	2.28±1.63	100
Sonications; instrument #2	0.31	1.09	10
Microplates; instrument #2	0.08	0.58	5
Instruments [1]	0.63±0.10	0.40±0.04	100
Within Fagales	20.47	82.72	146
Within Fagaceae	9.59	62.21	41
Within Betulaceae	19.09	66.04	106
Within *Alnus*	16.88	44.40	37
Within *Corylus*	16.42	62.18	23
Within *Betula*	18.13	54.45	43
Within *Quercus*	10.55	44.20	28
Within *Alnus hirsuta*	8.05	19.54	3
Within *Alnus viridis*	9.87	30.14	4
Within *Alnus incana*	12.63	30.25	15
Intraseasonal of *Alnus incana*, Norway 2014	12.78	16.75	9
Within *Alnus glutinosa*	16.73	44.17	15
Intraseasonal of *Alnus glutinosa*, Norway 2014	8.40	35.68	10
Within *Corylus avellana*	15.64	69.08	12
Intraseasonal of *Corylus avellana*, Germany 2014	8.89	61.62	6
Intraseasonal of *Corylus avellana*, Norway 2013	3.93	10.02	4
Within *Betula pendula*	18.69	44.85	19
Interannual of *Betula pendula*, Norway	18.29	44.46	9
Intraseasonal of *Betula pendula*, Norway 2014 [4]	15.59	19.50	4
Intraseasonal of *Betula pendula*, Norway 2013 [5]	11.14	25.09	3
Within *Quercus robur*	10.50	29.29	28
Interannual of *Quercus robur*, Norway	8.90	25.04	11
Intraseasonal of *Quercus robur*, Norway 2013	7.14	6.56	4
Intraseasonal of *Quercus robur*, Norway 2014	6.03	22.67	7
Within *Fagus sylvatica*	0.99	3.91	13
Intraseasonal of *Fagus sylvatica*, Norway 2014	0.84	4.39	6
Intraseasonal of *Fagus sylvatica*, Croatia 2013	1.05	2.91	6

[1]Based on 10 samples with 10 technical replicates each (± one standard deviation)

[4] Based on 4 samples from the same population, collected in Norway 2014

[5] Based on 3 samples from the same population, collected in Norway 2013

Biochemical fingerprints of pollen samples related to taxonomy

The studied set of pollen samples covers related species of Fagales. The set includes six genera of Betulaceae (birch family) and Fagaceae (beech family), with 31 species in total. Moreover, the set contains samples from three distinct pollination seasons from 2012 to 2014, as well as from three distinct locations in the Northern hemisphere: Croatia, Germany and Norway. Therefore, the set offers a broad phylogenetic diversity with good representation at different taxa levels, as well as temporal and spatial diversity with good representation of climate conditions (Table 1).

The data show an expected increase in spectral variability regarding taxonomy, going from species and genera to families and order (Table 2). In addition to the calculation of PCC values, principal component analysis (PCA) was performed to examine main variation patterns in the FTIR spectra (Figure 4). The variability within the different phylogenetic levels is approximately 1 to 2 orders of magnitude higher than on the technical replicates level, and even variability within species can be 20-30 times higher than the technical replicates. One exception is the high consistency of *Fagus sylvatica* pollen composition. *Fagus sylvatica* pollen samples were collected in Norway and Croatia, and they show extremely small intra- and inter-locational variability (Table 2 and Figure 4).

Spectral variability for a range of Fagales species shows high differences in pollen's chemical composition across taxa, in accordance with the published studies (Gottardini, Rossi, Cristofolini and Benedetti 2007, Pappas, Tarantilis, Harizanis and Polissiou 2003, Zimmermann and Kohler 2014, Zimmermann, Tafintseva, Bagcioglu, Hoegh Berdahl and Kohler 2016). It is important to notice that FTIR spectra present relative chemical composition of pollen grains, and the difference between *Fagus* and *Quercus* pollen composition is a good example. *Fagus* pollen shows higher relative amount of lipids and lower relative amount of proteins than *Quercus* pollen, as indicated by lipid related signals at 1745, 1463 and 1180 cm^{-1} and protein related signals at 1655 and 1545 cm^{-1}. This result is consistent with standard measurements of pollen composition, that have shown pollen protein content of 30.4% and 17.4 % for *Quercus robur* and *Fagus sylvatica* respectively (Roulston *et al.* 2000). On the other hand, one needs to take into consideration that a typical pollen grain of *F. sylvatica* has approx. double the size of *Q. robur* grain (radii of approx. 22 and 17 μm respectively). Therefore, regarding absolute chemical content, both species have similar amount of proteins per individual grain, while *F. sylvatica* has significantly higher amount

of lipids per grain compared to *Q. robur*. However, our previous FTIR studies have shown that pollen lipids between various species of *Quercus* can vary tenfold, with some, such as *Q. cerris*, *Q. libani* and *Q. frainetto*, having approx. three times higher amounts of lipids that *Q. robur* (Zimmermann and Kohler 2014).

Chemical composition of pollen and phenotypic plasticity

The variability shows high differences in pollen's chemical composition due to either location or pollination season. For example, pollen samples of *Quercus robur*, collected during the two consecutive seasons at the same location in Norway, show big variations in stored triglyceride nutrients (Figure 4 and Figure S1 in the Supporting Information). Higher triglyceride relative content (compared to sporopollenin) is present in the pollen grains from 2014 pollination season, as indicated by changes in lipid associated signals at 1745, 1470 and 1170 cm^{-1}. Regarding location, *Quercus robur* pollen samples from Germany show even lower amounts of triglycerides. In addition, German samples show slight differences in the sporopollenin composition, as indicated by changes in signals at 1602, 1515, 1171 and 833 cm^{-1}, as well as differences in lipids as indicated by signals at 1744 and 1706 cm^{-1} associated with triglycerides and free fatty acids respectively (Figure S1 in the Supporting Information).

The differences in sporopollenin composition are of interest due to intricate correlation between sporopollenin chemistry and environmental conditions. Sporopollenin is the most complex and robust plant extracellular matrix, and is composed of polyhydroxylated unbranched aliphatic and phenolic constituents (Kim and Douglas 2013). The composition of sporopollenin is not uniform, but rather a group of related polymers with significant taxon-specific variations in chemical composition (Dominguez *et al.* 1999). The main components of sporopollenin are oxygenated phenylpropanoid building blocks, particularly p-coumaric and ferulic acids. Production of these phenolic acids in plants is induced by solar ultraviolet (UV-B) via the phenylpropanoid pathway, the same pathway responsible for synthesis of lignin and suberin biopolymers. Studies have shown that concentrations and ratios of phenylpropanoids in sporopollenin depend on the magnitude of UV-B radiation to which plants had been exposed, and could serve as a proxy for estimating variations in UV-B radiation in palaeoecology (Fraser *et al.* 2011, Rozema *et al.* 2009, Willis *et al.* 2011). Moreover, pollen grains are covered with pollenkit, a sticky pollen coat composed mainly of lipids and plant pigments, and, alike sporopollenin, composition of pollenkit is taxon-specific (Pacini and Hesse 2005). Both sporopollenin and pollenkit are developed under

control of sporophytic genome, as opposed to development of intracellular materials and intine that are under control of gametophytic genome (Blackmore *et al.* 2007). Alongside sporopollenin, pollenkit is also responsible for protection against UV radiation. Therefore, slight differences in the sporopollenin and lipid composition for *Quercus robur* samples collected in Germany and Norway could be result of variations in exposure to UV-B. In general, detailed FTIR measurements of sporopollenin and pollenkit compositions could offer valuable information on plant-environment interaction and photoprotection mechanisms in plants.

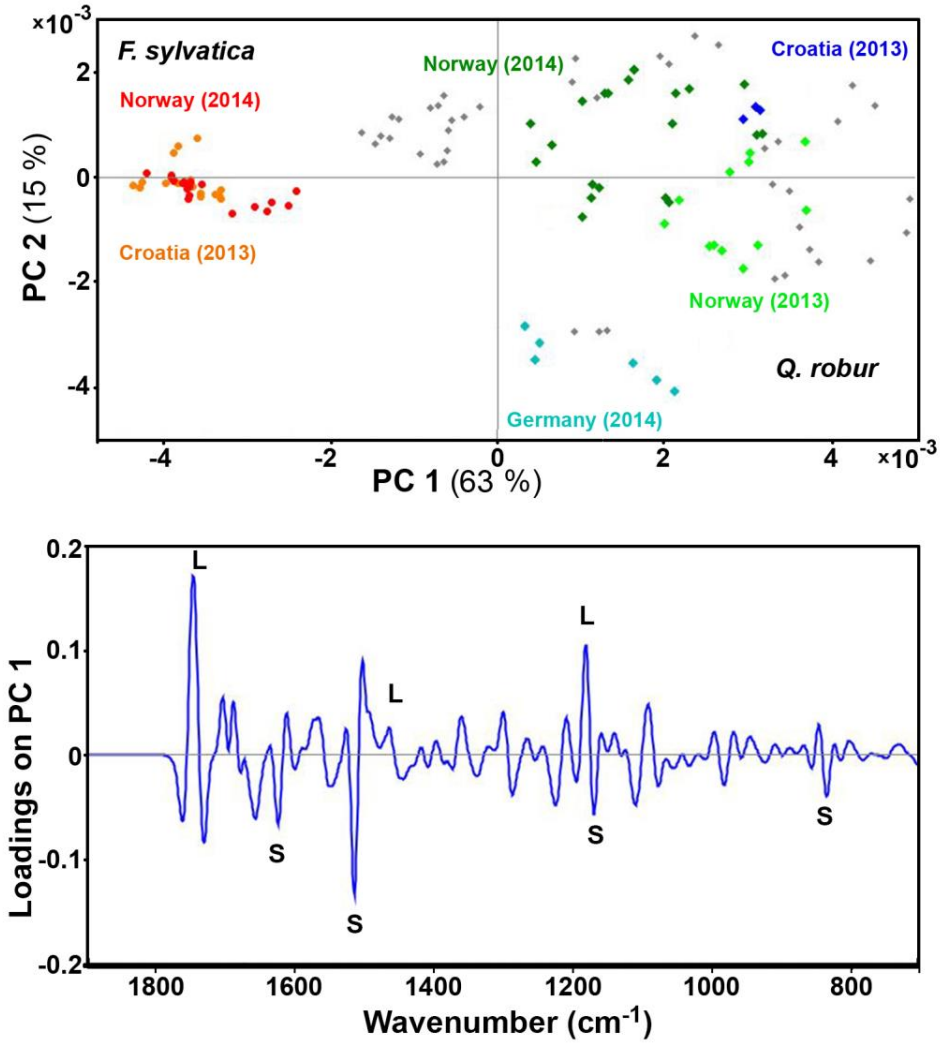


Figure 4. PCA plot of IR spectral data set of Fagaceae pollen (three spectra per sample), with depiction of plant genera (dot=*Fagus*, diamond=*Quercus*), locations and seasons. The percent variances for the first five PCs are 63%, 15%, 9%, 4% and 2%. (B) Loadings plot on the first principal component of the PCA; the marked signals are associated with vibrational bands of lipids (L) and sporopollenins (S).

Results for spectral variability for *Betula pendula* pollen samples are similar to the results obtained for *Quercus robur* samples, in regards that separations based on seasons and locations are present as well (Figure 5 and Figure S2 in the Supporting Information). In general, the spectra from German samples indicate a lower amount of carbohydrates compared to those of the Norwegian samples. Carbohydrates have several important roles in pollen, including energy reserves and cold tolerance, and all of them would be beneficiary in relatively cold Norwegian environment (Speranza, Calzoni and Pacini 1997). Moreover, inter-annual differences in the Norwegian pollen samples are based on higher lipid and protein relative content (compared to sporopollenin) for the 2014 pollination season. The climate data for Ås (Table S1 in the Supporting Information) shows big difference between the seasons, with 2-4 °C higher average temperature and 50% more precipitation for February-May period in 2014 compared to 2013. Study on pollen of Petunia hybrida, has shown that protein composition was dependent on the day temperature during the plant growth (Herpen 1986). Similar to our result, the total amount of protein was higher for the plant grown at high temperature. In addition, in case of Petunia hybrid, pollen tube length was positively correlated with growth temperature (Herpen 1986). It has been stipulated that pollen developed at the high temperature has more reserve material and thus makes longer pollen tubes than pollen developed at the low temperature. However, the increase of lipids in pollen with the increase of temperature is opposite to the results of study on Pinaceae pollen. The study on several *Abies*, *Picea* and *Pinus* species has shown negative correlation between pollen lipid content and ambient temperature (Zimmermann and Kohler 2014). Considering that Fagaceae and Pinaceae are far related taxa, such difference in plant-environment interaction is not entirely unexpected. In general, pollen triglycerides are valuable energy and material reserve, and higher amounts of lipids in pollen grain should be advantageous for the grain germination. Nevertheless, triglyceride lipids are not the only nutrient reserve in pollen. For instance, pollen starch reserves can be interconverted to lower molecular weight saccharides as well as lipids, thus creating a complex balance of pollen nutrients (Pacini and Juniper 1984).

The relatively large intra-seasonal variations for *Alnus glutinosa*, *Betula pendula* and *Quercus robur* indicate that differences in pollen composition can be substantial, even for the same species living at the same location. In particular, this is prominent for *Corylus avellana* pollen samples collected in Germany during winter 2014. The samples were collected from six plants,

with large differences in pollination phenology spanning more than five weeks. As can be seen in Figure 6, the composition of German *Corylus avellana* pollen samples changes dramatically during the course of the season. The early pollinating German plants (mid-January – early February) have the lowest lipid-to-protein ratio, followed by huge increase in the ratio in mid-pollinating plants (mid-February). Late pollinating German plants (late February) have lower lipid-to-protein ratio, and their pollen composition is quite similar to the Norwegian plants.

The high variation in pollen composition could be explained by *Corylus avellana* pollination strategy. *Corylus avellana* starts production of pollen during summer, and in late fall enters a dormancy period (Frenguelli *et al.* 1994). Dormant male catkins have high cold tolerance and can survive severe freezing and chilling conditions. Dormancy is finished with the rise of mean daily temperature, and the beginning of pollination is usually in late winter or early spring. The warming periods can vary substantially from year to year, and thus *Corylus avellana*, as well as other early pollinating species such as *Alnus incana*, often have large inter-annual variations in pollination phenology. Due to different adaptations to cold and warm weather, it can be expected that early pollinating species have larger variations in pollen composition between populations than late pollinators such as *Fagus sylvatica* and *Quercus robur*. Germany has several wild populations of *Corylus avellana* (Wanjiku and Bohne 2015), and it is highly probable that the plants growing in the Berlin botanical garden originate from different wild populations or cultivars. The Norwegian plants belong to the same cultivar, *Corylus avellana* 'Aurea', thus explaining more uniform phenology and pollen chemical composition.

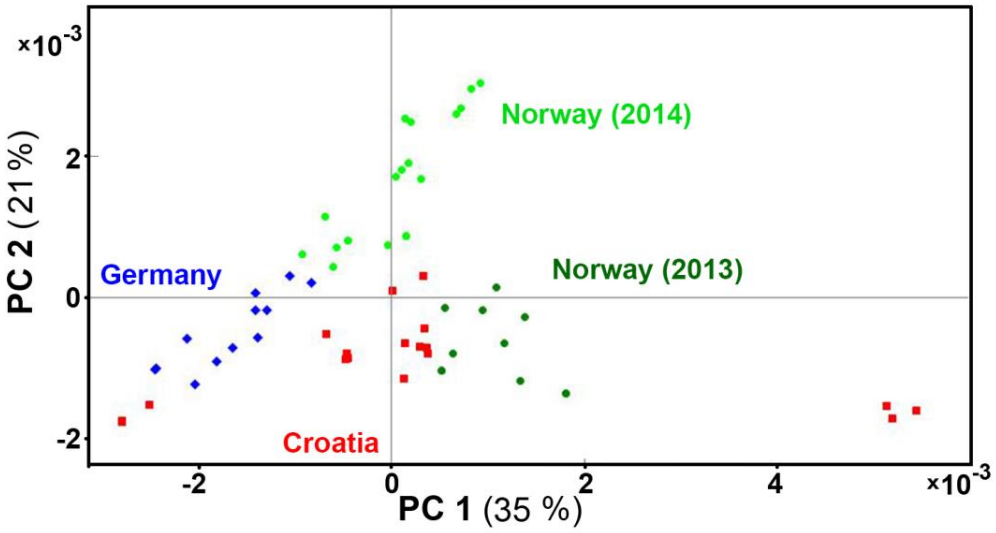


Figure 5. PCA plot of IR spectral data set of *Betula pendula* pollen (three spectra per sample), with depiction of plant locations and seasons. The percent variances for the first five PCs are 35%, 21%, 14%, 9% and 9%.

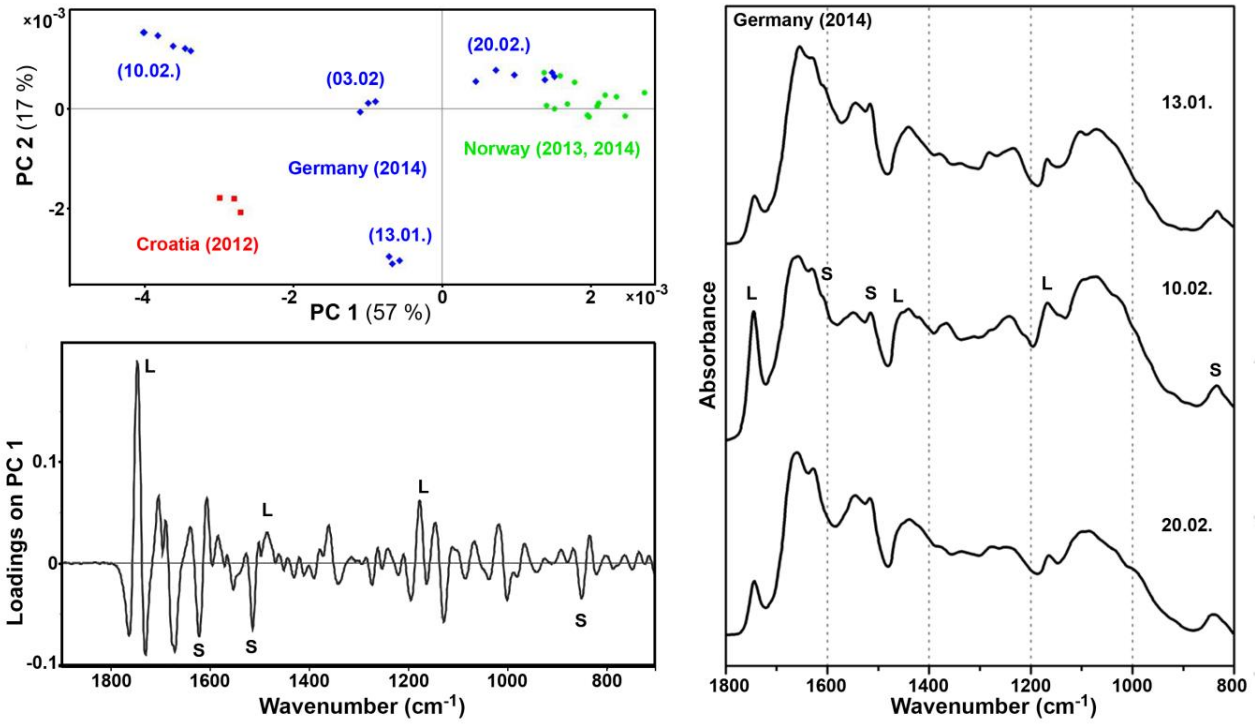


Figure 6. (A) PCA plot of IR spectral data set of *Corylus avellana* pollen (three spectra per sample), with depiction of plant locations and seasons. The percent variances for the first five PCs are 57%, 17%, 13%, 5% and 3%. (B) Loadings plot on the first principal component of the PCA. (C) *Corylus avellana* pollen spectra of samples collected at different dates in Germany 2014; the marked signals are associated with vibrational bands of lipids (L) and sporopollenins (S).

CONCLUSIONS

The high throughput FTIR approach discussed here enables rapid screening of large numbers of pollen samples. High quality spectra can be obtained, provided that sonication achieves sufficient fragmentation of pollen grains. Measurement variations due to sample preparation, microplate holders and instrumentation are low, and allow differentiation of samples with respect to phylogeny and biogeography. Specifically, the high quality of the data enables an observation of significant inter-annual, intra-seasonal and locational differences in the pollen spectra, which provide evidence that pollen chemical composition has high phenotypic plasticity and is influenced by local climate conditions. The changes are predominantly connected to changes in lipid triglyceride nutrients that play a crucial role in pollen germination and pollen tube growth.(Rodriguez-Garcia, M'rani-Alaoui and Fernandez 2003, Zhang *et al.* 2009) Regarding inter-annual variations, both *Betula pendula* and *Quercus robur* pollen samples from Norway had a higher relative content of triglycerides in 2014 samples compared to the samples from 2013 pollination season. Large inter-seasonal variations in *Corylus avellana* pollen collected in Germany in 2014 are probably associated with population-specific adaptations to quick and early pollination.

The results of the study, conducted on 146 Fagales samples, demonstrate the value of high throughput FTIR approach for the systematic collection of data on ecosystems. The novel FTIR approach offers fast, reliable and economical screening of large number of samples by semi-automated methodology. The high-throughput approach could provide crucial understanding on plant-climate interactions with respect to biochemical variation within genera, species and populations.

EXPERIMENTAL PROCEDURES

Samples

The studied pollen belongs to 31 different Fagales pollen species, covering 146 pollen samples in total (**Table 2**). Each sample comprised pollen belonging to approximately 10 male flowers (catkins) from the same individual plant. The pollen samples were collected directly from plants at flowering time during 2012, 2013 and 2014 pollination seasons at locations in Croatia, Germany and Norway. Pollen samples from Croatia were collected in Zagreb at the Botanical Garden of the Faculty of Science, the University of Zagreb. Pollen samples from Germany were collected in Berlin at the Berlin-Dahlem Botanical Garden, the Free University of Berlin. Pollen

samples from Norway were collected in Ås at campus area of Norwegian University of Life Sciences. After the sampling, the samples were kept at room temperature for 24 hours, and subsequently stored at −15 °C until measurements.

Table 2. List of the pollen species. Sampling locations: **N** – Norway, **G** – Germany, **C** – Croatia.

Family	Genus	Species	Name	N		G	C			#
				2013	2014	2014	2012	2013	2014	
Fagaceae	***Quercus***	*Q. cerris*	Turkish oak		1		1			2
		Q. cocciena	Scarlet oak		1					1
		Q. shumardii	Shumard oak				1			1
		Q. palutris	Spanish oak		1					1
		Q. robur	Pedunculate oak	4	7	2		1		14
		Q. libani	Lebanon oak				1	1	1	3
		Q. ilex	Evergreen oak					1		1
		Q. petreae	Irish oak		1		1	1		3
		Q. faginea	Lebanon oak					1		1
		Q. frainetto	Italian oak					1		1
	Fagus	*F. sylvatica*	European beech		6		1	6		13
Betulaceae	***Betula***	*B. pendula*	Silver birch	3	6	4	1	4	1	19
		B. alleghaniensis	Yellow Birch	1	2					3
		B. divaricata	Middendorff's birch		1					1
		B. ermanii	Erman’s birch	2	2			1		5
		B. lenta	Black birch		2					2
		B. papyrifera	Paper birch	1	2					3
		B. pubescens	European white birch		2					2
		B. raddeana	Radde’s birch	1						1
		B. utilis	Himalayan birch	1	4					5
		B. fruticosa	Shrubby birch	1	1					2
	Alnus	*A. glutinosa*	Black alder	1	10	3	1			15
		A. hirsuta	Manchurian alder	1	2					3
		A. incana	Grey alder	4	9	1	1			15
		A. viridis	Green alder	1	3					4
	Corylus	*C. americana*	American hazel	1	2					3
		C. avellana	Common hazel	4	1	6	1			12
		C. chinensis	Chinese hazel				1			1
		C. maxima	Filbert hazel	2		3				5
		C. cornuta	Beaked hazel	1	1	1				3
	Ostrya	*O. virginiana*	American hophornbeam		1					1

Sample preparation

In order to produce homogeneous suspensions of pollen samples that could be transferred to high-throughput FTIR plates and measured in transmission, the pollen samples were sonicated. Approximately 1 mg of a pollen sample was transferred into a 1.5 ml microcentrifuge tube

containing 500µl of distilled H_2O. The sample was sonicated in an ice bath, by a 2mm probe coupled to a Q55 Sonicator ultrasonic processor (QSonica, LLC, USA) under 100 % power. The sonication period was two minutes in total, with 30s intermission after the first minute of sonication. Following the sonication, the sample was centrifuged with 13000 rpm for 10 min, and the suspension was concentrated by removing 400 µl of the supernatant. Out of the remaining suspension, three aliquots, each containing 8 µl, were transferred onto an IR-transparent silicon 384-well microtiter plate (Bruker Optik GmbH, Germany). The microtiter plate was dried for 1 hour at r.t. to create adequate films for FTIR measurement.

From each pollen sample, only one sample solution was prepared. In order to test variability between different sonications and dry film preparations, *Alnus incana* (Norway, 2014) was measured on different microtiter plates using different instruments. For the sonication variability test, ten independent pollen suspensions were prepared by using the described procedure with the following modifications: microcentrifuge tubes were filled with the double amount of pollen and water, and, following centrifugation, 800 µl of supernatant was removed. Ten aliquots (each containing 8 µl) from each of the ten prepared suspensions were transferred onto one silicon 384-well microtiter plate, designated as *sonication variability test plate*, thus creating 100 sample films. For technical replicates variability, ten spectra for each suspension were used. For sonication variability, average spectra for each suspension were used. Following the preparation of the *sonication variability test plate*, the remaining suspensions were combined into one stock suspension. For microplate variability test, five different microtiter well plates were prepared by using ten aliquots (each containing 8 µl) of the stock suspension. The five plates, designated as *microplate variability test plates*, contained 50 sample films in total. For microplate variability, average spectra for each microtiter plate were used. For instrument variability test, five *microplate variability test plates* and the *sonication variability test plate* were measured using two different HTS-XT FTIR spectrometer systems. For instrument variability, average spectra for each suspension from the *sonication variability test plate* were used. For variability between different pollen samples (taxa, pollination seasons, and locations) average spectra of three technical replicates were used. Schematic diagram of variability test study is in the Supporting Information (Fig. S3).

Measurement

Infrared spectroscopy measurements were obtained using a HTS-XT extension unit coupled to a TENSOR 27 spectrometer (both Bruker Optik GmbH, Germany) (referred to as instrument #1). For instrument variability test, an additional spectrophotometer system, comprising HTS-XT extension unit coupled to a VERTEX 70 spectrometer (both Bruker Optik GmbH, Germany), was used (referred to as instrument #2). Both systems are 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 4000−500 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 was carried out using the OPUS/LAB software (Bruker Optik GmbH, Germany).

Spectral pre-processing and data analysis

The whole spectral range was used in the variability study, while the spectral region of 1900 to 700 cm^{-1} was selected for PCA since this spectral region contains bands distinctive for pollen grains (Bağcıoğlu, Zimmermann and Kohler 2015, Zimmermann and Kohler 2014). All spectra were transformed to second derivative form by the Savitzky-Golay algorithm using a polynomial of degree two and a window size of 21 points in total. The window size was chosen based on noise level and digital spacing (Zimmermann and Kohler 2013). Following the transformation to derivative form, the spectra were processed by extended multiplicative signal correction (EMSC), an MSC model extended by a linear, quadratic and cubic component (Kohler *et al.* 2005, Kohler *et al.* 2008, Zimmermann and Kohler 2013).

Biochemical similarities between pollen samples were estimated by using principal component analysis (PCA) and variability test based on Pearson correlation coefficients. All pre-processing methods and data analyses were performed using in-house developed routines written in MATLAB 2014a. 8.3.0.532 (The MathWorks, Natick, USA).

Acknowledgements

The research was supported by the Norwegian Research Council (IS-DAAD project No. 233941, DAAD No. 57068987), and by the European Commission through the Seventh Framework Programme (FP7-PEOPLE-2012-IEF project No. 328289). J.K. acknowledges funding by ERC Starting grant 259432 MULTUBIOPHOT.

Author Contributions

Conceived the research idea: AK, BZ, JK. Designed the experiments: AK, BZ, MB. Performed the experiments: BZ, MB, SS. Analyzed the data: AK, BZ, MB. Wrote the paper: BZ, MB. Discussed and revised the paper: AK, JK, SS.

Supporting Information

Contains additional information on PCAs, climate data for Ås (Norway), and schematic diagram of variability test study.

References

Anderson, J.T., Inouye, D.W., McKinney, A.M., Colautti, R.I. and Mitchell-Olds, T. (2012) Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. *P Roy Soc B-Biol Sci*, **279**, 3843-3852.

Bağcıoğlu, M., Zimmermann, B. and Kohler, A. (2015) A Multiscale Vibrational Spectroscopic Approach for Identification and Biochemical Characterization of Pollen. *PLoS ONE*, **10**, e0137899.

Blackmore, S., Wortley, A.H., Skvarla, J.J. and Rowley, J.R. (2007) Pollen wall development in flowering plants. *New Phytologist*, **174**, 483-498.

Bokszczanin, K.L., Fragkostefanakis, S. and Thermotolerance, S.P. (2013) Perspectives on deciphering mechanisms underlying plant heat stress response and thermotolerance. *Frontiers in plant science*, **4**.

Chakrabarti, B., Singh, S.D., Nagarajan, S. and Aggarwal, P.K. (2011) Impact of temperature on phenology and pollen sterility of wheat varieties. *Aust J Crop Sci*, **5**, 1039-1043.

Crowe, J.H., Hoekstra, F.A. and Crowe, L.M. (1989) Membrane Phase-Transitions Are Responsible for Imbibitional Damage in Dry Pollen. *P Natl Acad Sci USA*, **86**, 520-523.

Dean, A.P., Sigee, D.C., Estrada, B. and Pittman, J.K. (2010) Using FTIR spectroscopy for rapid determination of lipid accumulation in response to nitrogen limitation in freshwater microalgae. *Bioresource Technol*, **101**, 4499-4507.

Dell'Anna, R., Lazzeri, P., Frisanco, M., Monti, F., Malvezzi Campeggi, F., Gottardini, E. and Bersani, M. (2009) Pollen discrimination and classification by Fourier transform infrared (FT-IR) microspectroscopy and machine learning. *Analytical and bioanalytical chemistry*, **394**, 1443-1452.

Delph, L.F., Johannsson, M.H. and Stephenson, A.G. (1997) How environmental factors affect pollen performance: Ecological and evolutionary perspectives. *Ecology*, **78**, 1632-1639.

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

Endo, M., Tsuchiya, T., Hamada, K., Kawamura, S., Yano, K., Ohshima, M., Higashitani, A., Watanabe, M. and Kawagishi-Kobayashi, M. (2009) High Temperatures Cause Male Sterility in Rice Plants with Transcriptional Alterations During Pollen Development. *Plant Cell Physiol*, **50**, 1911-1922.

Fraser, W.T., Sephton, M.A., Watson, J.S., Self, S., Lomax, B.H., James, D.I., Wellman, C.H., Callaghan, T.V. and Beerling, D.J. (2011) UV-B absorbing pigments in spores: biochemical responses to shade in a high-latitude birch forest and implications for sporopollenin-based proxies of past environmental change. *Polar Res*, **30**.

Frenguelli, G., Ferranti, F., Fomaciari, M. and Romano, B. (1994) Dormancy and Chilling Requirement for Catkins in Hazel (Corylus Avellana L., Corylaceae). *Giornale botanico italiano*, **128**, 198-198.

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.

Herpen, M.M.A. (1986) Biochemical Alterations in the Sexual Partners Resulting from Environmental Conditions before Pollination Regulate Processes after Pollination. In *Biotechnology and Ecology of Pollen: Proceedings of the International Conference on the Biotechnology and Ecology of Pollen, 9–11 July, 1985, University of Massachusetts, Amherst, MA, USA* (Mulcahy, D.L., Mulcahy, G.B. and Ottaviano, E. eds). New York, NY: Springer New York, pp. 131-133.

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.

Kim, S.S. and Douglas, C.J. (2013) Sporopollenin monomer biosynthesis in arabidopsis. *J Plant Biol*, **56**, 1-6.

Kohler, A., Bocker, U., Shapaval, V., Forsmark, A., Andersson, M., Warringer, J., Martens, H., Omholt, S.W. and Blomberg, A. (2015) High-throughput biochemical fingerprinting of Saccharomyces cerevisiae by Fourier transform infrared spectroscopy. *Plos One*, **10**, e0118052.

Kohler, A., Kirschner, C., Oust, A. and Martens, H. (2005) Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform infrared microscopy images of cryo-sections of beef loin. *Applied spectroscopy*, **59**, 707-716.

Kohler, A., Sule-Suso, J., Sockalingum, G.D., Tobin, M., Bahrami, F., Yang, Y., Pijanka, J., Dumas, P., Cotte, M., van Pittius, D.G., Parkes, G. and Martens, H. (2008) Estimating and correcting Mie scattering in synchrotron-based microscopic Fourier transform infrared spectra by extended multiplicative signal correction. *Applied spectroscopy*, **62**, 259-266.

Lahlali, R., Jiang, Y.F., Kumar, S., Karunakaran, C., Liu, X., Borondics, F., Hallin, E. and Bueckert, R. (2014) ATR-FTIR spectroscopy reveals involvement of lipids and proteins of intact pea pollen grains to heat stress tolerance. *Frontiers in plant science*, **5**.

Mignolet, A. and Goormaghtigh, E. (2015) High throughput absorbance spectra of cancerous cells: a microscopic investigation of spectral artifacts. *Analyst*, **140**, 2393-2401.

Mularczyk-Oliwa, M., Bombalska, A., Kaliszewski, M., Wlodarski, M., Kopczynski, K., Kwasny, M., Szpakowska, M. and Trafny, E.A. (2012) Comparison of fluorescence spectroscopy and FTIR in differentiation of plant pollens. *Spectrochim Acta A*, **97**, 246-254.

Nicotra, A.B., Atkin, O.K., Bonser, S.P., Davidson, A.M., Finnegan, E.J., Mathesius, U., Poot, P., Purugganan, M.D., Richards, C.L., Valladares, F. and van Kleunen, M. (2010) Plant phenotypic plasticity in a changing climate. *Trends in plant science*, **15**, 684-692.

Ollesch, J., Drees, S.L., Heise, H.M., Behrens, T., Bruning, T. and Gerwert, K. (2013) FTIR spectroscopy of biofluids revisited: an automated approach to spectral biomarker identification. *Analyst*, **138**, 4092-4102.

Pacini, E. and Hesse, M. (2005) Pollenkitt - its composition, forms and functions. *Flora*, **200**, 399-415.

Pacini, E. and Juniper, B. (1984) The Ultrastructure of Pollen Grain Development in Lycopersicum-Peruvianum. *Caryologia*, **37**, 21-50.

Pappas, C.S., Tarantilis, P.A., Harizanis, P.C. and Polissiou, M.G. (2003) New method for pollen identification by FT-IR spectroscopy. *Applied Spectroscopy*, **57**, 23-27.

Parodi, G., Dickerson, P. and Cloud, J. (2013) Pollen Identification by Fourier Transform Infrared Photoacoustic Spectroscopy. *Applied spectroscopy*, **67**, 342-348.

Pummer, B.G., Bauer, H., Bernardi, J., Chazallon, B., Facq, S., Lendl, B., Whitmore, K. and Grothe, H. (2013) Chemistry and morphology of dried-up pollen suspension residues. *J Raman Spectrosc*, **44**, 1654-1658.

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.

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.

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. *Photoch Photobio Sci*, **8**, 1233-1243.

Sirikwanpong, S., Dahlan, W., Ngamukote, S., Sangsuthum, S., Adisakwattana, S., Nopponpunth, V. and Himathongkam, T. (2010) The Alterations of Erythrocyte Phospholipids in Type 2 Diabetes Observed after Oral High-Fat Meal Loading: The FTIR Spectroscopic and Mass Spectrometric Studies. *J Clin Biochem Nutr*, **47**, 111-120.

Sowa, S., Connor, K.F. and Towill, L.E. (1991) Temperature-Changes in Lipid and Protein-Structure Measured by Fourier-Transform Infrared Spectrophotometry in Intact Pollen Grains. *Plant Sci*, **78**, 1-9.

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.

Wanjiku, J. and Bohne, H. (2015) Early frost reactions of different populations of hazelnut (Corylus avellana L.). *Eur J Hortic Sci*, **80**, 162-169.

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.

Willis, K.J., Feurdean, A., Birks, H.J., Bjune, A.E., Breman, E., Broekman, R., Grytnes, J.A., New, M., Singarayer, J.S. and Rozema, J. (2011) Quantification of UV-B flux through

time using UV-B-absorbing compounds contained in fossil Pinus sporopollenin. *The New phytologist*, **192**, 553-560.

Wolkers, W.F. and Hoekstra, F.A. (1995) Aging of Dry Desiccation-Tolerant Pollen Does Not Affect Protein Secondary Structure. *Plant physiology*, **109**, 907-915.

Wolkers, W.F. and Hoekstra, F.A. (1997) Heat stability of proteins in desiccation-tolerant cattail (Typha latifolia L) pollen - A Fourier transform infrared spectroscopic study. *Comp Biochem Phys A*, **117**, 349-355.

Zhang, M., Fan, J.L., Taylor, D.C. and Ohlrogge, J.B. (2009) DGAT1 and PDAT1 Acyltransferases Have Overlapping Functions in Arabidopsis Triacylglycerol Biosynthesis and Are Essential for Normal Pollen and Seed Development. *Plant Cell*, **21**, 3885-3901.

Zimmermann, B. (2010) Characterization of Pollen by Vibrational Spectroscopy. *Appl Spectrosc*, **64**, 1364-1373.

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.

Zimmermann, B. and Kohler, A. (2013) Optimizing Savitzky-Golay Parameters for Improving Spectral Resolution and Quantification in Infrared Spectroscopy. *Applied Spectroscopy*, **67**, 892-902.

Zimmermann, B. and Kohler, A. (2014) Infrared spectroscopy of pollen identifies plant species and genus as well as environmental conditions. *Plos One*, **9**, e95417.

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.

Zimmermann, B., Tkalcec, Z., Mesic, A. and Kohler, A. (2015b) Characterizing aeroallergens by infrared spectroscopy of fungal spores and pollen. *Plos One*, **10**, e0124240.

Supporting Information

Monitoring of plant-environment interactions by high throughput FTIR spectroscopy of pollen

Murat Bağcıoğlu[1], Achim Kohler[1,2], Stephan Seifert[3,4], Janina Kneipp[3,4], Boris Zimmermann[1*]
[1]Department of Mathematical Sciences and Technology, Faculty of Environmental Science and Technology, Norwegian University of Life Sciences, 1432, Ås, Norway
[2]Nofima AS, 1430, Ås, Norway
[3]Humboldt-Universität zu Berlin, Department of Chemistry, 12489 Berlin, Germany
[4]BAM Federal Institute for Materials Research and Testing, 12489 Berlin, Germany

*Corresponding author:
Boris Zimmermann
Department of Mathematical Sciences and Technology
Faculty of Environmental Science and Technology
Norwegian University of Life Sciences
Drøbakveien 31, 1432 Ås, Norway.
Tel: +47 6723 1576
Faks: +47 6496 5001
E-mail: boris.zimmermann@nmbu.no

Table of Contents **Page**

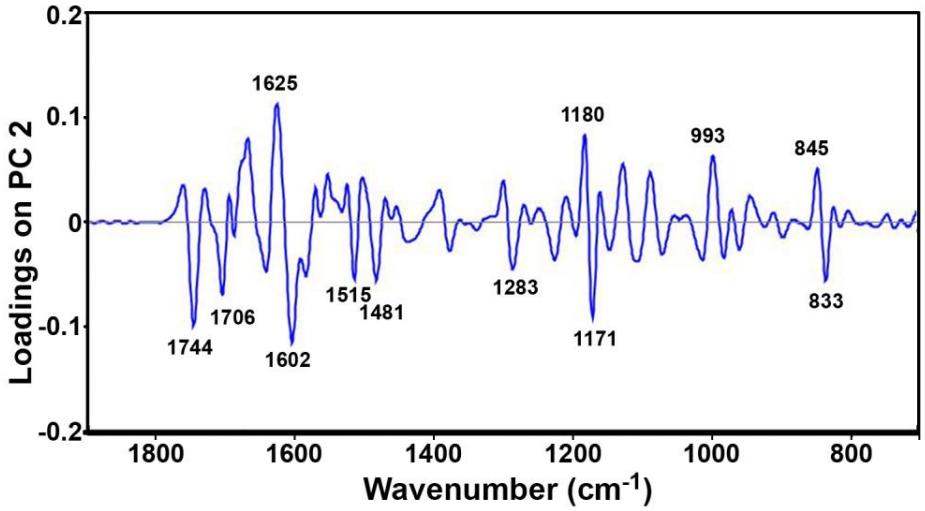


Fig. S1. PCA of IR spectral data set of Fagaceae pollen (see Figure 4); Loadings plot on the second principal component of the PCA.

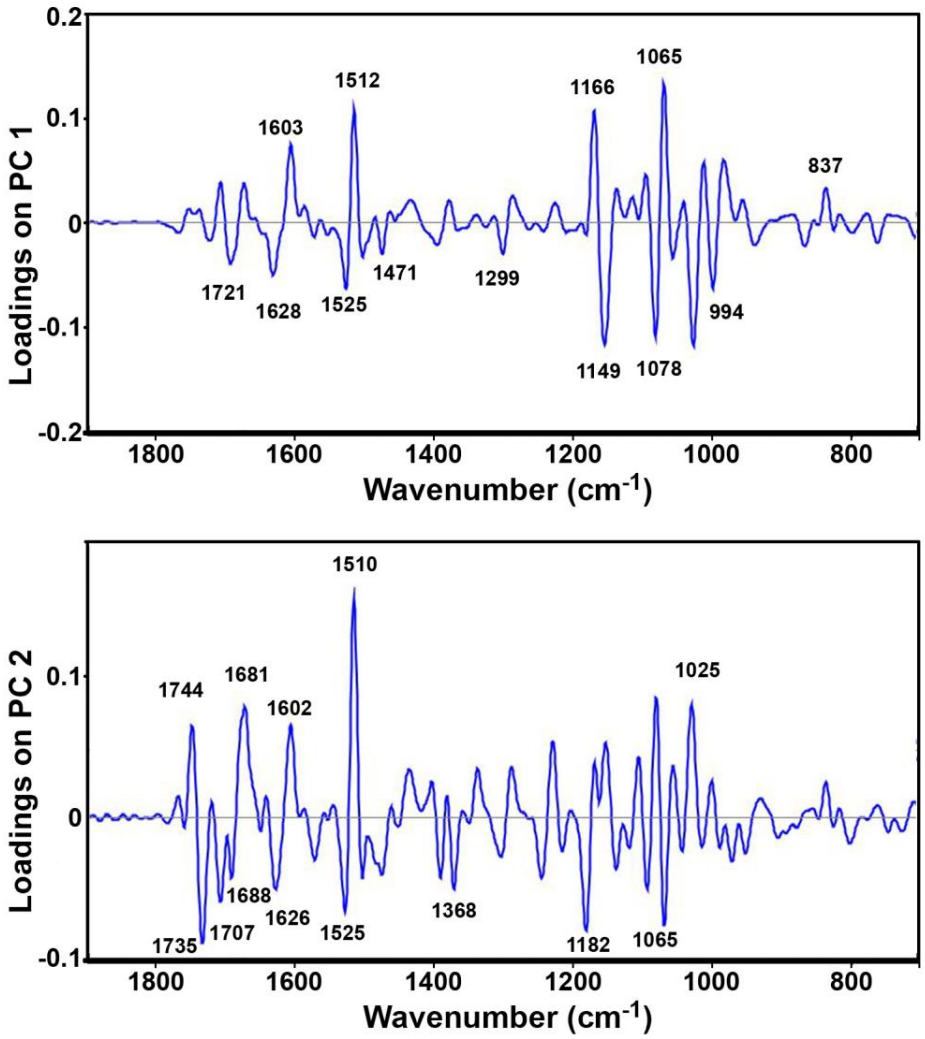


Fig. S2. PCA of IR spectral data set of *Betula pendula* pollen (see Figure 5); Loadings plot on the first two principal components of the PCA.

Table S1. Climate data for Ås, Norway, 2013-14

Month	Temperature, ºC (average)		Precipitation, mm (cumulative)	
	2013 *	**2014** **	**2013** *	**2014** **
January	-5,2	-2,6	44	89,7
February	-4,4	1,7	20,6	160,1
March	-3,6	3,8	3,6	47,3
April	3,6	6,8	60,2	64,5
May	12,0	11,1	125,3	44,1

Detailed weather information:

*Thue-Hansen V. and Grimenes A.A.
Meteorologiske data for Ås 2014.
Norwegian University of Life Sciences 2015.
ISBN 978-82-7636-028-8
(https://www.nmbu.no/download/file/fid/10111)

**Thue-Hansen V. and Grimenes A.A.
Meteorologiske data for Ås 2013.
Norwegian University of Life Sciences 2014.
ISBN 978-82-7636-027-1
(https://www.nmbu.no/download/file/fid/4763)

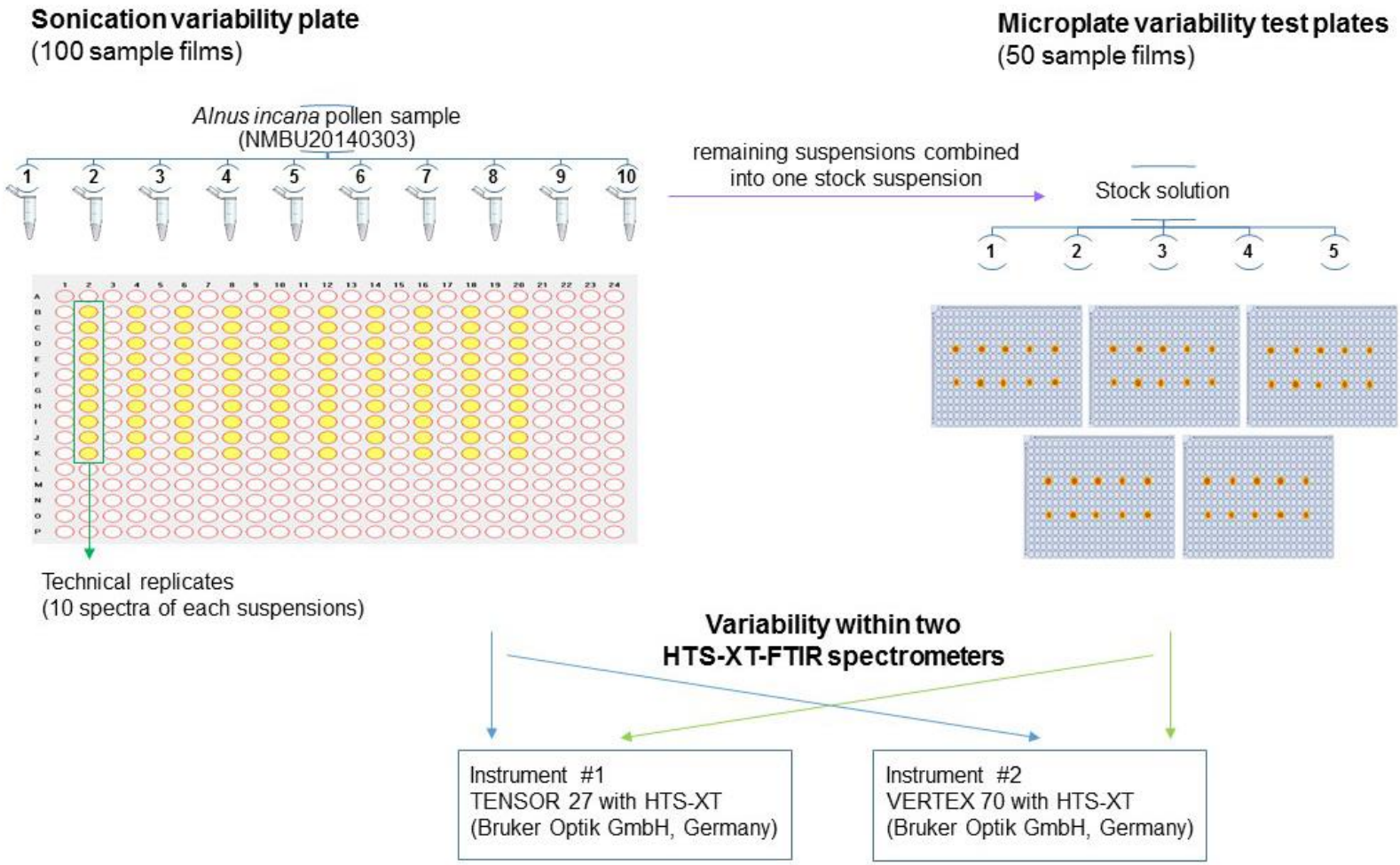


Fig. S3. Schematic diagram of variability test study.

6.5 Paper V

A greenhouse study for assessing environmental effects on pollen composition by FTIR spectroscopy

Boris Zimmermann[1], Murat Bağcıoğlu[1], Valeria Tafinstseva[1], Achim Kohler[1,2], Mikael Ohlson[3], Siri Fjellheim[4]

[1]Department of Mathematical Sciences and Technology, Faculty of Environmental Science and Technology, Norwegian University of Life Sciences, Ås, Norway
[2]Nofima AS, Ås, Norway
[3]Department of Ecology and Natural Resource Management, Faculty of Environmental Science and Technology, Norwegian University of Life Sciences, Ås, Norway
[4]Department of Plant Sciences, Faculty of Environmental Science and Technology, Norwegian University of Life Sciences, Ås, Norway

*** Corresponding author**
E-mail: siri.fjellheim@nmbu.no; phone: +4767232801

ABSTRACT

Pollen quantity and quality are two significant factors regarding male reproductive success in plants. The quality component is related to genetic or environmental basis. The male reproductive function is reflected in pollen quality due to a trade-off between pollen size, nutritional value and metabolic cost. Pollen quality can be altered for any condition through phenotypic plasticity to maximize reproductive output under different environments. Pollen analysis have been widely used in climate and paleoclimate studies of environmental change and vegetation history by standard optical microscopy in order to provide information on pollen morphology and germination. Nevertheless, grass pollen is rarely identified at the subfamily or lover taxonomic levels. Fourier transform infrared (FTIR) spectroscopy, a biophysical technique that is providing a precise fingerprint of the overall biochemical composition of pollen grain, has recently emerged for pollen analysis. In this study, pollen samples belonging to three different species of Poales grown in controlled environments were measured by FTIR in order to study plasticity and genetic effects. Specifically, environmental and nutrients effects as well as phylogenetic relationships among the pollen were examined. The role of pollen variability in light of adaptation and success of male reproduction were discussed. It was shown that FTIR spectroscopy is not only a valuable tool for classification of pollen species, but also for shedding new light on the plant male function. As a consequence, the study demonstrated that pollen composition is species-specific and related to life-history strategies. Further, pollination season is species-specific and reflects biological function. Finally, environmental conditions influence pollen composition and reflects adaptation to local climate.

INTRODUCTION

Pollen and ovules are the two key players in one of the most fundamental of all processes in plant biology, i.e fertilization. While the female function is well described, the biology of the male function is less investigated. The difference in weight given to male and female reproductive output in research can be ascribed to the relative ease with which the female reproductive output is measured compared to that of pollen. However, given its fundamental role in reproduction, we expect that the male function is under strong selective pressure to maximize reproductive success. Here we use novel spectroscopic techniques to investigate adaptations in the male reproductive function.

Most studies on pollen structure and morphology have focused on taxonomy and phylogeny and little has been done on establishing the relationship between varying structure and composition and the evolution and adaptation of plant male function. Pollen quantity and quality are two key determinants of male reproductive success that both can limit fertilization through quantity in terms of pollen availability, and through quality in terms of pollen viability, germination capacity and tube development rate (Arceo-Gomez *et al.*, 2016; Williams & Mazer, 2016). We may expect trade-offs between quantity and distribution on one side and viability, germination, growth and fertilization success on the other side. The quality component can have both a genetic and an environmental basis (Burd, 2008; Rosenheim *et al.*, 2014), but to what extent pollen quality is genetically fixed or environmentally induced remains unknown (or perhaps better with "poorly understood"). Several studies have, however, investigated the role of environmental factors on various aspects of pollen performance, but these studies are largely focused on post anther pollen conditions rather than conditions during pollen production and maturation (Williams & Mazer, 2016). Consequently, the link between the drivers of pollen quality during its production and pollen performance is largely unexplored.

The cost of male reproductive function is likely reflected in pollen quality due to a trade-off between pollen size, nutritional value and metabolic cost. Two main drivers of pollen quality evolution are widely accepted: one is pollination mode (by wind, water or animals), and the other is the length the pollen tube needs to grow from the stigma to the ovule i.e. that pollen of species with long distances need to be rich in energy and nutrients (Torres; Williams & Rouse; Lau & Stephenson, 1993). Consequently, pollen quality varies between species and families (Arceo-Gomez *et al.*, 2016; Williams & Mazer, 2016). Insect pollinated species, for example, frequently

use pollen as reward for pollinators and have in general higher protein and/or oil content (Jones & Little, 1983). Accordingly, Lidforss has showed that plants with high pollen protein content attract more insects (Lidforss, 1899). In comparison, wind- or self-pollinated species such as grasses have high starch contents (Speranza *et al.*, 1997; Wang *et al.*, 2004). The evolutionary rationale is that the metabolic cost of producing proteins and oils are higher than carbohydrates, and plants with non-rewarding pollen (e.g. wind-pollinators, self-pollinators, and nectar rewarders) will thus be characterized by pollen that are more rich in carbohydrates.

If different pollination strategies are utilized within plant families, there may be within-family mixtures of species that have pollen that differ in their nutrient composition, but the strictly anemophilous grass family Poaceae has exclusively starch rich pollen (Baker & Baker, 1979). Moreover, anemophilous species generally have small pollen as they need to produce vast amounts of pollen (Roulston *et al.*, 2000). There is likely an evolutionary trade-off between size and number as it has been suggested that larger pollen give an advantage in competitive situations (Lau & Stephenson, 1993; Lau & Stephenson, 1994). Furthermore, grasses have only one ovule (Andersen & Bertelsen, 1972; Salih *et al.*, 1997; Kellogg, 2001; Mander *et al.*, 2013) hence strong male gametophytic competition may be a significant determinant of pollen quality within the Poaceae family.

Given its fundamental importance, pollen quality may be genetically fixed for high performance independent of environmental conditions. Alternatively, it may be plastic to maximize reproductive output under different environmental conditions. One of the most studied and influential factors limiting pollen growth and development is temperature (Hedhly *et al.*, 2004; Rang *et al.*, 2011). Already in 1893, Molisch found that the content of starch in pollen grains of *Anthirrinum majus* changed throughout the season (Molisch, 1893)and Sears and Metcalf (Sears & Metcalf, 1926) identified less starch in pollen grains of *Pelargonium* matured at 25°C than at 15°C. Germination and pollen tube growth under different temperature regimes have been much investigated. However, the role of pollen chemical composition – i.e. quality – is rarely (if ever) investigated, and the studies were focused only on the conditions during germination and pollen tube growth (Alcaraz *et al.*, 2011; Pham *et al.*, 2015). It has been shown that inter-annual variation in lipid content for *Pinus*, *Picea* and *Abies* is strongly correlated with temperature prior to pollination onset (Van Herpen & Linskens, 1981; Van Herpen, 1986; Lau & Stephenson, 1994; Zimmermann & Kohler, 2014). A few other studies also provide examples of how environmental

conditions during pollen maturation influence their chemical composition (Stanley & Linskens, 1974). On the other hand, Baker and Baker (Baker & Baker, 1979) report that pollen size and nature of the pollen reserves are only little influenced by environmental factors.

Pollen analyses have been widely used in climate and paleoclimate studies of environmental change and vegetation history, but also in biodiversity and community monitoring. Standard pollen analysis is predominantly obtained by optical microscopy in order to provide information on pollen morphology and germination. (Koti *et al.*, 2004; Kakani *et al.*, 2005; Sivaguru *et al.*, 2012) However, microscopy identification is limited to reliably classified categories, such as size, shape and texture, and often results with low taxonomical resolution due to lack of distinctive morphological features. For instance, grass pollen is rarely identified at the subfamily or lover taxonomic levels. Alternative analyses, such as bio-imaging, proteomics and transcriptomics, can provide more detailed pollen analyses, particularly regarding pollen biochemistry.(Holmes-Davis *et al.*, 2005; Matsushima *et al.*, 2008; Suwabe *et al.*, 2008) Unfortunately, all these methods are time-consuming and require complex sample preparation and laborious analysis by experienced specialists.

Recently, Fourier transform infrared spectroscopy (FTIR) has emerged as a significant breakthrough in pollen analysis.(Pappas *et al.*, 2003; Gottardini *et al.*, 2007; Dell'Anna *et al.*, 2009; Zimmermann, 2010; Lahlali *et al.*, 2014; Zimmermann & Kohler, 2014; Bağcıoğlu *et al.*, 2015; Jiang *et al.*, 2015; Zimmermann *et al.*, 2015a; Zimmermann *et al.*, 2015b; Zimmerman *et al.*, 2016) FTIR is a biophysical technique that is providing a precise fingerprint of the overall biochemical composition of pollen grain.(Zimmermann & Kohler, 2014; Bağcıoğlu *et al.*, 2015; Zimmermann *et al.*, 2015a; Zimmermann *et al.*, 2015b) Moreover, FTIR can be combined with a microscope for focused measurement, thus providing single grain identification, as well as chemical imaging of pollen grain ultrastructure.(Zimmermann *et al.*, 2015a; Zimmerman *et al.*, 2016) In general, FTIR enables fast and economical measurement of samples without any chemical pre-treatment. The studies have shown that FTIR spectroscopy is a powerful method for pollen phenotyping since it provides precise identification and biochemical characterization with respect to phylogeny and environmental stress.(Lahlali *et al.*, 2014; Zimmermann & Kohler, 2014; Jiang *et al.*, 2015; Zimmermann *et al.*, 2015b)

In this paper we use FTIR methodology to study variability of pollen quality in the grass family. We grew closely related species under controlled climate to determine the degree and

nature of variation in pollen quality within and between species, and investigated to what extent pollen quality is plastic in response to differing temperatures and nutrient levels. We discuss the role of pollen variability in light of adaptation and success of male reproduction. We show that the use of FTIR paves way for more detailed study of the chemical composition of pollen and its role in plant adaptation. In combination with rigorous testing in controlled environments, FTIR can contribute new knowledge about plant male reproductive biology.

MATERIALS AND METHODS

Sample set

Seed of three populations from each of the species *Poa alpina, Anthoxanthum odoratum* and *Festuca ovina* was acquired from the Nordic Gene Bank. The population were chosen to cover geographic and climatic variation. Seeds were sown spring 2013 in Tjerbo Gartnerjord (Tjerbo, Rakkestad, Norway), and 15 individuals per species grew outside at Norwegian University of Life Sciences, Ås, Norway over summer before we divided each individual into four clones. During this period, plants were fertilized upon need with water containing 4% Yara Kristalon Indigo (Yara, Skøyen, Norway) and 3% YaraLiva Calcium Nitrate (Yara, Skøyen, Norway) adjusted to an electron conductivity of 1.5. To induce flowering we vernalized the plants 12 weeks at 4°C under short day length (8 hours). After vernalization the four clones were subjected to four different treatments: high (20°C) and low temperature (14°C) in combination with high and low levels of nutrients (+/- NU). Plants from all treatments were grown under log day inductive conditions of 20 hours. The +NU plants were watered twice a week with nutrient containing water. The –NU plants received only water, thus no extra fertilizer apart from what was in the soil at the start of the flowering period was added. Plants were randomized between and within treatments. Within each treatment, plants were rotated three times a week. In total, 520 individuals were included in the flowering experiment, after removing misclassified individuals and some individuals with mould infection (see Table S1 in the Supporting information for details). 460 individuals produced inflorescence, and, out of those, 435 produced sufficient amount of pollen for spectroscopic analysis (at least 0.5 mg per individual).

Spectroscopic analysis

380 pollen samples, each belonging to a different individual plant, were covered by the *main FTIR study*. The main study covered 140 samples of *Anthoxanthum odoratum* (44, 48 and 48 for populations France, Greece and Finland, respectively), 96 samples of *Festuca ovina* (48, 32 and 16 for populations Sweden, Finland and Italy, respectively) and 144 samples of *Poa alpina* (48 samples for each of the populations: Sweden, Italy and Norway). Samples were uniformly distributed, to cover each of the four growth conditions with a quarter of samples (95 samples in total per growth condition). In addition to the main FTIR study, *FTIR timeline study* was conducted by covering 9 additional pollen samples of the same individual plant (*Festuca ovina*, population Italy, 20 °C with added nutrients), collected at different times (spanning 11 days in total).

For FTIR measurements, homogeneous suspensions of pollen samples were prepared. Approximately 1mg of a pollen sample was transferred into 1.5 ml microcentrifuge tube containing 500 µl of distilled H_2O. The sample was sonicated in ice bath, by a 2mm probe coupled to a Q55 Sonicator ultrasonic processor (QSonica, LLC, USA) under 100 % power. The sonication period was two minutes in total, with 30 s intermission after the first minute of sonication to minimize the increase of temperature. Following the sonication, the sample was centrifuged with 13000 rpm for 10 min, and the suspension was concentrated by removing 400 µl of supernatant. Out of the remaining suspension, three aliquots, each containing 8 µl, were transferred onto IR-light-transparent silicon 384-well microtiter plate (Bruker Optik GmbH, Germany). The microtiter plate was dried at room temperature for 1 hour at r.t. to create adequate films for FTIR measurement.

Infrared spectroscopy measurements were obtained using a HTS-XT extension unit coupled to a TENSOR 27 spectrometer (both Bruker Optik GmbH, Germany). The systems 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 was recorded before each sample well measurement. The spectra were measured in the 4000−500 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 was carried out using the OPUS/LAB software (Bruker Optik GmbH, Germany).

Spectral pre-processing and data analysis

For the main FTIR study the FTIR spectral data set consisted of 1140 spectra, belonging to 380 pollen samples measured in three technical replicates. The entire spectral region of 4000 to 500 cm^{-1} was used in the data analysis. The spectra were subjected to preprocessing by second derivative Savitzky-Golay algorithm(Savitzky & Golay, 1964) using a polynomial of degree two and a windows size of 7 points followed by extended multiplicative signal correction (EMSC) with linear and quadratic components. The final step of preprocessing was averaging of the technical replicates resulting in 380 average spectra in total. For the FTIR timeline study the FTIR spectral data set consisted of 75 spectra, belonging to 16 pollen samples of Italian *Festuca ovina*, population (from the main study) plus additional 9 samples of the same plant collected at different time points. The data was pre-processed in the same way as for the main study, but without the averaging step.

Classification and biochemical similarities between pollen samples were evaluated by using principal component analysis (PCA), partial least-squares regression (PLSR), and variability tests based on Pearson correlation coefficients and Fisher discriminant analysis. PLSR was used for building discrete and hierarchical classification schemes.

For the hierarchical classification, the data set was split into a training set, comprising 218 spectra, and a validation set, comprising 162 spectra. The calibration model was established using Sparse partial least squares regression (SPLSR) with 3-block cross-validation (CV) of randomly split samples for the model evaluation. SPLSR is applied on the entire spectrum and penalizes those variables in loading weight vectors which are below a certain threshold. The threshold λ is defined by the desired number of variables to be neglected called degree of sparsity(Karaman *et al.*, 2013). For each principal component the degree of sparsity is optimized in a range between 90 and 99 percent of variables. The degree of sparsity resulting in minimum misclassification rate (MCR) was chosen as optimal. At the end, the optimal number of principal components is defined in a course of CV as the one which gives not significantly higher MCR than the minimum MCR (significance was evaluated by a binomial test; function binocdf in Matlab). Success rate (SR) is expressed in percentage and is defined as (1-MCR) ×100. A hierarchical classification tree with three levels (species, population, growth conditions) was established with one model for each node. Thus, one model for the species, 9 models for the populations and 36 models for the clones, resulting in 46 models in total. The CV results were of satisfactory quality with only few models

of poor performance. However, an independent validation is the best way to evaluate the classification model performance.

To evaluate the ability to detect an influence of environmental conditions on the plants and pollen, classification models were established for nine different populations. First, each population was split into two groups of different temperatures. Second, each population was split into two groups of different nutrition conditions. The average number of samples in each population was 42, which is not enough to perform an external validation. The statistical significance of the obtained models was evaluated by permutation test.(Lindgren *et al.*, 1996; Szymanska *et al.*, 2012) In each step of a permutation test, the labels of the samples are permuted randomly and a new classification model is established. A distribution of the null hypothesis, H_0, is obtained by performing N=1000 permutations. A p-value is calculated by comparing permuted MCR, MCR_p, to the original unpermuted MCR: $p = (1 + \#(MCR_p \leq MCR))/N$.

To evaluate the models' performance, cross-validation, with 5-block and 6-block Venetian blinds was performed for temperature and nutrient classification respectively. To evaluate the ability to differentiate different nutrient and temperature conditions nine other models were established in the course of cross-validation with 4-block Venetian blinds.

Fisher discriminant analysis was performed on the PCA scores of spectral data for each population separately. The between and within class scatter matrices were calculated as follows: $S_b = \sum_i n_i\,(\mu - \bar{x}_i)(\mu - \bar{x}_i)^T$ and $S_w = \sum_i \sum_j (\bar{x}_i - x_{ij})(\bar{x}_i - x_{ij})^{\mathrm{T}}$, where μ is a global mean, $\bar{x}_i$ is a group mean of group *i*, n_i is a number of samples in group *i*. The ratio, Fisher clustering coefficient (FCC), S_b/S_w was used to evaluate the genotype- based clustering of spectral data in the PC subspace.

The variability between samples for the *FTIR timeline study* was estimated by calculating Pearson Correlation Coefficient (PCC)(Lee Rodgers & Nicewander, 1988), and was expressed as $1–PCC\times10^{-4}$ for the spectral region 1700–1600 cm^{-1}.

All pre-processing methods and data analyses were performed using in-house developed routines written in MATLAB 2015a (The MathWorks, Natick, USA).

RESULTS

Growth and flowering of sampled plants

Of the 520 individuals that were included in the flowering experiment, 460 individuals developed inflorescence, and out of those 435 created sizeable amount of pollen (≥0.5 mg) required for the FTIR measurement. The species have shown different flowering phenologies: While *Poa alpina* individuals were pollinating within 3-4 days, *Anthoxanthum odoratum* individuals were pollinating within several weeks, and with almost continuous production of new inflorescence. 761 samples of pollen were collected in total (Table S1 and S2 in the Supporting information).

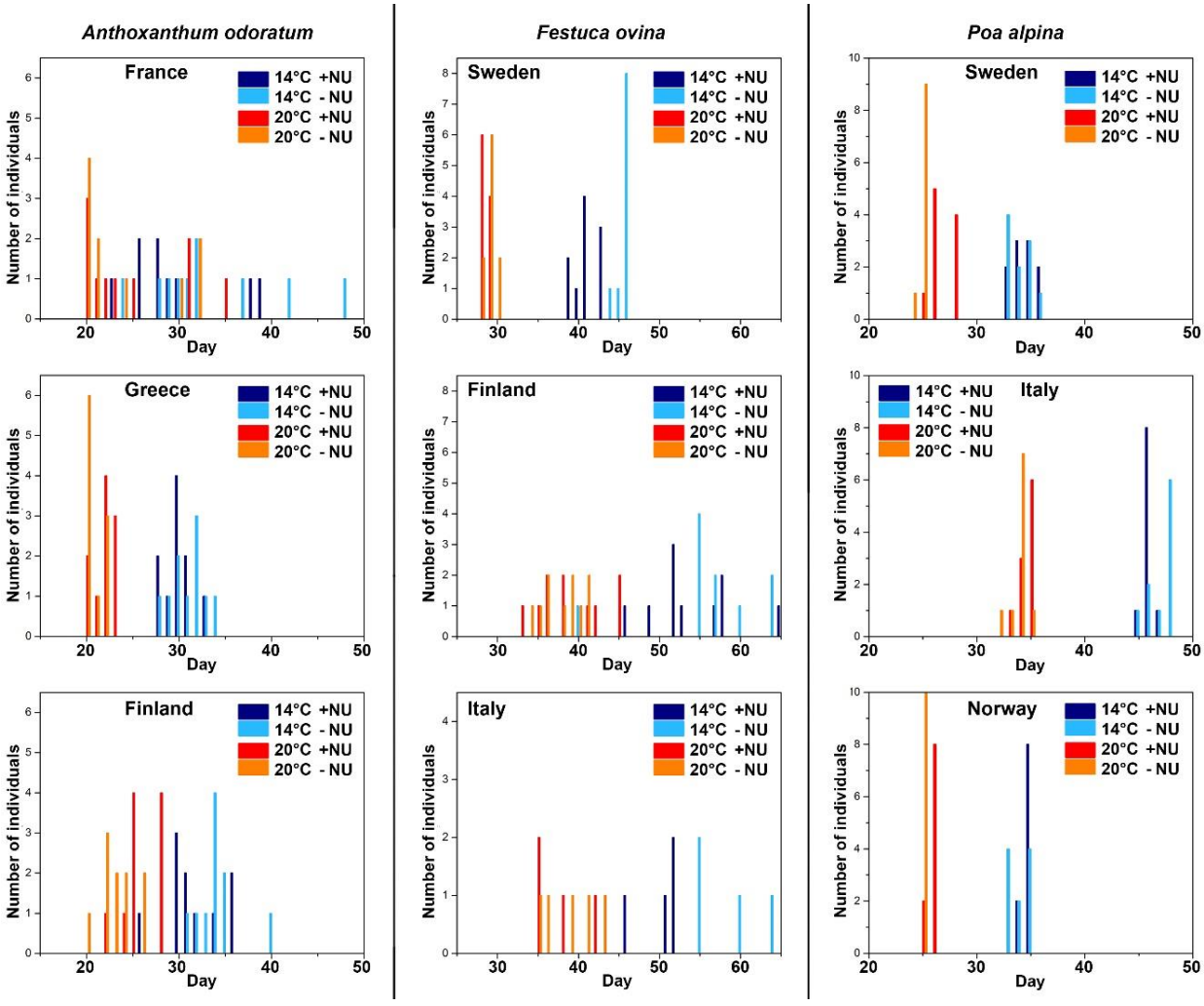


Figure 1. Pollen phenology for grass species and populations, designating the start of pollination for first ten individuals per population; days are counted from the start of four different treatments.

As can be seen in Figure 1, higher temperature shifts start of pollination period by approx. 10 days. Nutrient regime does not have big impact, except for Festuca ovina in cold treatment, where low nutrient regime delays pollination season for approx. 5 days. *Poa alpina* individuals have extremely synchronized pollination start. This is of importance since total pollination period for *Poa alpina* is approx. 3-4 days. Without good synchronization there would be small chance for cross pollination and gene flow. *Anthoxanthum odoratum* individuals have variable pollination start. This has not big impact since the individuals were pollinating within several weeks, and with almost continuous production of new inflorescence. *Festuca ovina* populations show both type of behavior. Swedish population had synchronized pollination start with short total pollination period (3-4 days) while Finish and Italian populations had variable pollination start with long pollination period (10-15 days).

FTIR measurement of pollen

The spectral variability of pollen samples was estimated by Pearson Correlation Coefficient (PCC) (Table 1). Variability was estimated by measurement of samples belonging to Italian population of *Festuca ovina*. Technical replicates refer to repeated FTIR measurements using the same pollen suspension, which was applied to different sample positions on the silicon well plate. The variability of the pollen composition for one individual during the pollination season was estimated based on spectral variability of samples belonging to the same plant (*Festuca ovina*, population Italy, 20 °C with added nutrients), collected during 11 consecutive days, from the start to the end of pollination. As can be seen from the Table 1, the variability in pollen composition for this individual was slightly higher than variability within FTIR measurement (i.e. technical replicates), and considerably lower than variability for the whole population in the corresponding growth condition. The spectral variability for the whole population in all four growth conditions is two-and-a-half times higher than variability within samples belonging to one individual plant. The results indicate that pollen composition of one individual grass plant is quite invariant during the intraseasonal pollination.

Table 1. Variability within technical replicates and biological samples of *Festuca ovina* (Italy).

Type of variability	(1-PCC)×10^{-4}
Technical replicates*	15±7
Within *F. ovina* individual*	34
Within *F. ovina* (Italy), 20 °C, N+	56
Within *F. ovina* (Italy)	85

* Based on measurements of samples from the *FTIR timeline study* (see Materials and methods).

Pollen composition and phylogeny

The representative FTIR spectra of pollen of the three species are shown in Figure 2. The spectra is dominated by signals associated with proteins, lipids and carbohydrates.(Gottardini *et al.*, 2007; Zimmermann *et al.*, 2015b) The characteristic protein signals are at approx. 1650 cm^{-1} (amide I:C = O stretch) and approx. 1540 cm^{-1} (amide II: NH deformation and C–N stretch), carbohydrate at 1200-900 cm^{-1} (C-O-C, C-C and C-O stretching vibrations), and lipid signals at 1745 (C=O stretch), 1470 (CH_2 deformation), and 1170 cm^{-1} (C-C stretch). The major carbohydrate signals (1200-1000 cm^{-1}) can be associated with cellulose and amylose (starch). In addition, the spectra show signals related to CH stretching at 3050 cm^{-1} (=C-H stretch) and 2960-2830 cm-1 (C-H stretch in $–CH_3$ and $–CH_2–$), that will always be observed due to the presence of alkyl and alkenyl groups in cellular components like lipids, proteins and carbohydrates. Finally, the spectra show signals associated with sporopollenins, a grain wall biopolymers based on phenylpropanoid acids.(Zimmermann, 2010; Bağcıoğlu *et al.*, 2015) The major sporopollenins signals (1668, 1658, 1625, 1512, 850, 835, and 815 cm^{-1}) can be associated with ferulic and sinapic building blocks, although the presence of p-coumaric building bloc cannot be excluded.

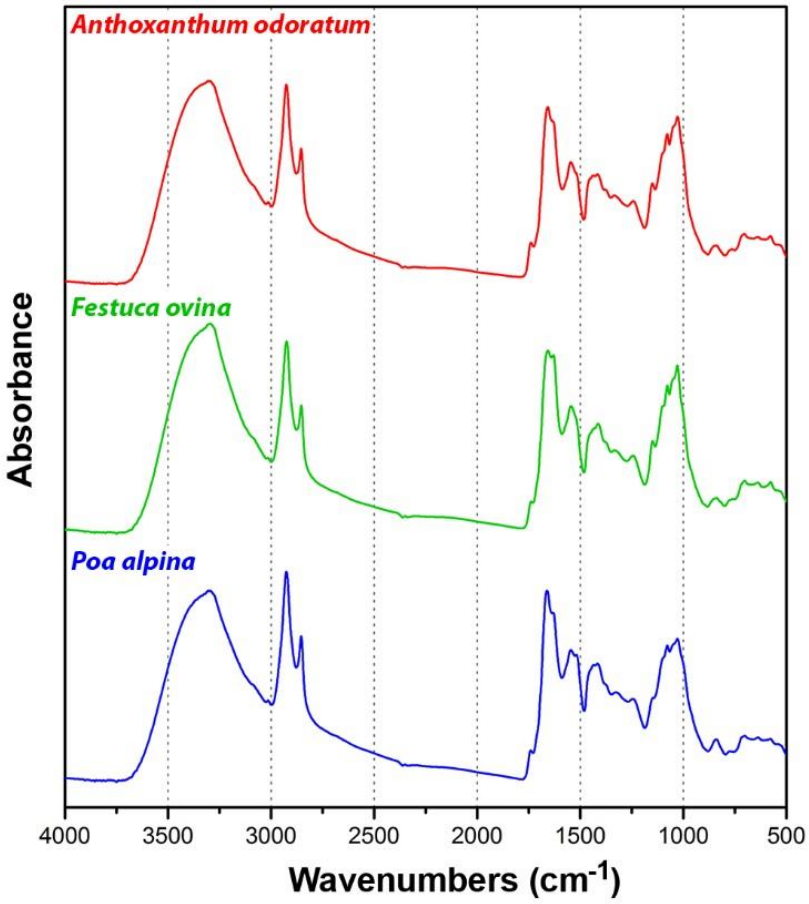


Figure 2. Representative spectra of pollen; all spectra belong to pollen samples collected from plants that were growing at 20°C in combination with high levels of nutrient.

Hierarchical classification of spectral data was conducted on three levels by SPLSR. Success rates (SR) for identification were: 1st level (species) SR=99.4%, 2nd level (population) SR=77.0%, and 3rd level (growth conditions) SR=43.5%. Confusion matrices for all three levels are presented in Figures S1-S3 in the Supporting information.

The success rates for identification of samples obtained by SPLSR hierarchical classification of pollen FTIR spectral data can be seen in Table 2. The misclassification at species level occurred only for Finish population of *Festuca ovina*. The plants from this population often produced quite small amount of pollen, thus resulting with suboptimal sample quantities for FTIR measurement. The corresponding FTIR spectra had significantly lower absorbance values (approx. half of the average value for *Anthoxanthum* and *Poa* samples). Therefore, this misclassification could be attributed to technical reasons. In general, all three species have distinguishing FTIR spectra that can be used for extremely accurate identification.

Table 2. Success rates for identification of samples obtained by SPLSR hierarchical classification of pollen FTIR spectral data; results are based on validation set.

Samples	Identification of species		Identification of populations
A. odoratum, France	100 %	100 %	75 %
A. odoratum, Greece		100 %	75 %
A. odoratum, Finland		100 %	20 %
F. ovina, Sweden	98 %	100 %	85 %
F. ovina, Finland		94 %	59 %
F. ovina, Italy		100 %	100 %
P. alpina, Sweden	100 %	100 %	90 %
P. alpina, Italy		100 %	100 %
P. alpina, Norway		100 %	100 %

The regression coefficients for classification of species show that the classification is based on several species-specific chemical fingerprints (Figure S4 in the Supporting information). *P. alpina* has higher protein-to-carbohydrate ratio than the other two species. This feature is clearly seen in Figure 1 by comparing 'protein' spectral regions (1700-1500 cm^{-1}) and 'carbohydrate' spectral regions (1200-900 cm^{-1}) of the three species. Differentiation between *A. odoratum* and the other two species is based on signals at 1076 and 1052 cm^{-1} that can be associated with higher concentration of amylose. In addition, *Anthoxanthum* has higher lipid-to-carbohydrate ratio than *F. ovina*. Differentiation between *P. alpina* and the other two species is based on the signal at 1164 cm^{-1} that is probably associated with higher quantity of arabinoxylan. Moreover, ratios of sporopollenins-related signals at 1512, 1625, 1658 and 1668 cm^{-1} for *A. odoratum* and *F. ovina* indicate differences in sporopollenin composition. Finally, all three species show differences in the amount and chemical composition of lipids, as indicated by signals at 2923, 1745 and 1712 cm^{-1}.

Chemical differences between populations are species specific (The regression coefficients for classification of populations are presented in Figures S5 in the Supporting information). While differences between populations of *P. alpina* are significant, enabling accurate identification of all three populations, the differences between either *A. odoratum* or *F. ovina* are less prominent (Table 2). Regarding *P. alpina*, the Norwegian population has higher ratio of carbohydrates-to-lipids, compared to other two populations. These carbohydrate signals, in 1200-900 cm^{-1} region, can be associated with amylose. In addition, differentiation of *P. alpina* populations is based on extensive differences in amide I and amide II protein bands.

Pollen composition and growth conditions

Table 3 shows the P-values for environmental effect on pollen FTIR spectral data. Eight populations, including all populations of *Poa alpina*, show statistically significant ($P < 0.05$) correlation between pollen composition and temperature conditions during plants growth. In addition, six populations show statistically significant correlation between pollen composition and nutrient conditions during plants growth, while in the remaining three cases the correlation is high ($P<0.1$), but not statistically significant. Italian population of *Festuca ovina* is the only population that has not shown high correlation between pollen composition and growth conditions. As stated previously, this could be due to lower spectral quality and smaller number of samples compared to an average population. To evaluate the influence of all four growth conditions on pollen composition, nine other models were established. The models were of successful quality, but on average worse than the separate temperature and nutrition classification models.

The chemical interpretation of spectral differences regarding growth conditions should be done with precaution, since the differences are small, population specific and highly complex (The regression coefficients for classification based on growth conditions are presented in Figures S9-S11 in the Supporting information). Nevertheless, some general conclusions can be made. Regarding nutrient conditions, the main spectral differences are associated with changes in sporopollenin (1626 and 1510) and lipid signals (1740-1710, 1473 and 1150 cm^{-1}). Moreover, more nutrients results with higher protein-to-carbohydrate-ratio, as indicated by changes in protein signals (at approx. 1660 and 1530 cm^{-1}) and starch signals (at approx. 1070 and 1050 cm^{-1}). These differences are consistent for both *A. odoratum* and *P. alpina*. Regarding temperature conditions, the differences are species specific. Carbohydrate signals at 1083 and 1023 cm^{-1}, probably related to amylose, are higher for *A. odoratum* samples grown in cold conditions than for warmer conditions. In case of *P. alpina*, the protein-related amide signals are higher for samples grown in warm conditions than for colder conditions.

Table 3. P-values and success rates (SR) for environmental effect on pollen FTIR spectral data; results are based on SPLSR calibration set.

Samples	Temperature		Nutrients		Growth conditions*	
	P-value	SR (%)	P-value	SR (%)	P-value	SR (%)
A. odoratum, France	**0.001**	90.9	**0.002**	75.0	**0.001**	79.5
A. odoratum, Greece	**0.001**	79.2	**0.004**	75.0	**0.001**	66.7
A. odoratum, Finland	**0.001**	77.1	0.055	62.5	**0.001**	66.7
F. ovina, Sweden	**0.001**	77.1	**0.002**	72.9	**0.001**	58.3
F. ovina, Finland	**0.009**	75.0	0.074	65.6	**0.005**	43.8
F. ovina, Italy	0.23	62.5	0.068	75.0	0.021	43.8
P. alpina, Sweden	**0.001**	95.7	**0.013**	70.2	**0.001**	76.6
P. alpina, Italy	**0.001**	97.9	**0.001**	91.7	**0.001**	85.4
P. alpina, Norway	**0.001**	89.6	**0.001**	85.4	**0.001**	79.2

*For SPLSR models for separation of all four environmental conditions with 10 principal components (detailed results are in the Supporting information).

Phenotypic plasticity and rigidity of pollen composition

The Table 3 shows that pollen compositions of the majority of populations are quite plastic, particularly regarding temperature conditions. However, it is questionable if the remaining cases, for which there is no significant correlation with growth conditions, are showing opposite behaviour, i.e. phenotypic rigidity. To test this, we have performed clustering analysis based on genotypes (i.e. genet with four remets growing under different conditions), to estimate phenotypic rigidity based on Fisher clustering coefficient (FCC). The high FCC values indicate clustering based on genotype (i.e. high phenotypic rigidity), while low values indicate either high plasticity or random variability.

The Fisher clustering coefficient (FCC) for estimating clustering based on genotype can be seen in Table 4. Finish population of *Anthoxanthum odoratum* has the highest FCC value, strongly implying high phenotypic rigidity (Figure S12 in the Supporting information). The FCC values for Finish population of *Festuca ovina* are relatively high, thus potentially indicating high phenotypic rigidity for this population as well. However, this could be due to, previously mentioned, lower spectral quality of data for this population, so random variability cannot be excluded. The lowest FCC values were obtained for *Poa alpina* populations, thus validating notion that this species has high plasticity of pollen chemical composition (Figure S12 in the Supporting information).

Table 4. Fisher clustering coefficient (FCC) for estimating clustering based on genotype.

Samples	FCC
A. odoratum, France	0.26
A. odoratum, Greece	1.09
A. odoratum, Finland	10.29
F. ovina, Sweden	0.60
F. ovina, Finland	1.29
F. ovina, Italy	0.16
P. alpina, Sweden	0.08
P. alpina, Italy	0.04
P. alpina, Norway	0.01

DISCUSSION

The evolutionary and ecological role of variations in the chemical composition of pollen is largely unexplored, even though it is recognized that pollen chemical composition influences dispersal, germination and growth (Williams & Mazer, 2016), and thus most likely has adaptive significance. The lack of research on the role of pollen composition for pollen function is mostly due to methodological constrains. In this paper we show that FTIR spectroscopy is not only a valuable tool for classification of species, but also as a tool to shed new light on the plant male function. Three main conclusions can be drawn from our results: I) pollen composition is species-specific and related to life-history strategies; II) pollination season is species-specific and reflects the biological niche of the species; and III) environmental conditions influence pollen composition and reflects adaptation to local climate.

FTIR reveal intra-familial differences that can be used for species differentiation

The recent introduction of FTIR for analyses of pollen spectra show that the technique is capable of revealing phylogenetic patterns at high taxonomic levels (Zimmermann & Kohler, 2014; Bağcıoğlu *et al.*, 2015; Zimmerman *et al.*, 2016). To explore the versatility of this tool in classification and phylogeny we asked if closely related species on a sub-family level could be separated based on their spectra. Our results show that all three grass species have highly specific FTIR profiles and the potential to use such profiles for identification is promising. As expected, spectra were dominated by proteins, lipids, carbohydrates and sporopollenin (Figure 1). We further asked if populations within a species could be separated based on spectra. Few attempts have been

made to analyze intraspecific variation of pollen spectra by FTIR, but Buta et al. (Buta *et al.*, 2015) who successfully separated different cultivars of *Saintpaulia*. In our study we detected within species variations in pollen composition, hence our results corroborate (Buta *et al.*, 2015) Interestingly, the amount of within species variation in pollen composition was species-specific. In particular, *P. alpina* had highly specific profiles for the three populations whereas classification of populations in *A. odoratum* and *F. ovina* was less precise. Importantly, this should not be interpreted as a limitation of the method, but as a reflection of the biological strategy of the different species. We will discuss this below.

Growth conditions influence pollen quality

Most studies on pollen structure and morphology have focused on taxonomy and phylogeny and little has been done on establishing the relationship between varying structure and composition and the evolution and adaptation of plant male function. We may expect trade-offs between quantity and distribution on one side and viability, germination, growth and fertilization success on the other side. Several studies have investigated the role of environmental factors on various aspects of pollen performance, however, they are largely focused on post anther pollen conditions rather than conditions during pollen production and maturation (Williams & Mazer, 2016). Consequently, the link between pollen composition and pollen performance is largely unexplored.

In *P. alpina*, pollen matured at high temperature had a higher level of proteins, whereas colder temperature gave higher levels of lipids and carbohydrates. Similar results were found for pollen of *Petunia hybrida* sired at lower temperatures (19.5 ^{o}C vs 25.5 ^{o}C), which showed a reduced level of protein and low-molecular weight carbohydrates (Van Herpen & Linskens, 1981; Van Herpen, 1986), whereas amount of lipids were the same. High levels of proteins under warmer conditions may increase the fitness of the plant as it facilitates pollen tube growth (Van Herpen & Linskens, 1981). Furthermore, stigma tissue is more short-lived in warmer climate (Lora *et al.*, 2009) hence fast germination and growth facilitated by proteins may be favorable. In colder climate, wide dispersion could be favorable as the stigma tissue is more long-lived. A higher carbohydrate contents would make the pollen lighter and more easily distributed. This show the trade-off between high value nutrients during favorable conditions and lipids and carbohydrates during less favorable conditions.

More nutrients, including nitrogen and phosphorus, gave higher protein-to-carbohydrate-ratio in both *A. odoratum* and *P. alpina*. A relationship between nutrition (phosphorous) and fitness was established in a study of *Cucurbita pepo* where plants in nutrient rich soils showed higher pollen production per flower and larger pollen grain size (Lau & Stephenson, 1994). This gave a fitness advantage as pollen from high phosphorous treatment sired significantly more mature seeds than pollen from nutrient poor soils. It is reasonable to conclude that a similar fitness advantage can be found in *A. odoratum* and *P. alpina* grown under high nutrient conditions in our study. Although we have not measured pollen grain size, it can be assumed that certain increase is pollen grain size is present for grass species. Given that, the parent plant will need to invest more nutrients per pollen grain in order to maintain functionality of bigger pollen grains. This will require higher amount of carbohydrates for grain wall (intine), in form of cellulose and arabinoxylan, and starch in cytoplasm. This will also require higher amount of proteins. Although proteins are present in the grain wall, this amount is minor compared to proteins in cytoplasm. Since increase of pollen size leads to higher increase of volume of cytoplasm than increase of pollen grain wall (due to square-cube law), bigger pollen grains will have higher protein-to-carbohydrate-ratio than smaller grains.

Furthermore, we also demonstrate that changes in lipids and sporopollenin are associated with different nutrient regimes of parent plants. However, the nature of these changes are more diffuse and we might only speculate that it is adaptive. Lipids play an important role in pollen germination (Dickinson *et al.*) and are essential for the pollen tube to penetrate the stigma, possibly through directing pollen-tube growth by controlling flow of water to pollen (Wolters-Arts *et al.*, 1998). Sporopollenin is a complex biopolymer of variable chemical composition that provides strength and resilience to outer layer (exine) of pollen grain wall. Alike lignin, chemical composition of sporopollenin is not uniform and it can show large variations in chemical composition (Dominguez *et al.*, 1999). Both sporopollenin and surface grain-wall lipids are developed under control of sporophytic genome, as opposed to intracellular materials that are developed under control of gametophytic genome (Piffanelli *et al.*; Blackmore *et al.*, 2007). In both cases, the results could indicate that biosynthetic pathways involved in the formation of lipids and sporopollenin are altered towards production of chemically somewhat different products due to different nutrient treatments of sporophytes.

Finally, decrease of amylose related signals with temperature in *A. odoratum* could indicate acclimation to warmer conditions. The decreased concentration of amylose could indicate that pollen starch is partly hydrolyzed into low molecular weight carbohydrates, such as sucrose, glucose and fructose. These soluble carbohydrates are related to pollen longevity through regulation of water content (Hesse, 2006). The pollen of grasses is not dehydrated upon release like most plant species and loose water quickly (Bassani *et al.*, 1994). Study on maize pollen has shown that sucrose levels are strongly correlated to sensitivity to desiccation, and that sucrose is a key factor in preserving membranes in dry pollen (Hoekstra *et al.*, 1989). The same study has shown that starch hydrolysis increases during slow drying, thus significantly increasing resistance of pollen to desiccation.

Species-specific responses indicate adaptation to different niches

The three species show species-specific responses that are explainable in the light of different life histories, distribution patterns and niches. *Festuca ovina* and *A. odoratum* are widespread Eurasian species that grow in a wide range of environments, from lowland to alpine regions. *P. alpina* has a restricted distribution and an alpine and sub-arctic distribution. Corresponding to this, *A. odoratum* and *F. ovina* are more of generalists (Fjellheim *et al.*, 2015) than *P. alpina*, although all species are plastic in response to different environments. *S*hort growing season is characteristic for all parts of the distribution area of *P. alpina* and accordingly, the pollination season of *P. alpina* was short and concluded in 3-4 days. *Festuca ovina* and *A. odoratum* mostly had prolonged pollination continuing over several weeks and *A. odoratum* continuously produced new inflorescences.

Poa alpina has a higher protein-to-carbohydrate ratio than *A. odoratum* and *F. ovina* reflecting the advantage of high quality pollen in a short growing season (Fjellheim *et al.*, 2015). *P. alpina* also had the highest level of plasticity of the species which may be a specialization to maximize reproductive output in the short growing season often with unpredictable and unstable climatic conditions. However, to establish if plasticity is adaptive and enhance fitness, further research is needed.

As opposed to *P. alpina*, *A. odoratum* and *F. ovina* produce more pollen over longer periods of time (Figure 2), invest less in each pollen grain (i.e. higher carbohydrate-to-pollen ratio)

and pollen composition is more invariant between populations and environments. Taken together this reflects more of a generalist strategy with a trade-off between quality and quantity giving more rather than high-quality pollen. *P. alpina* invest more in quality in terms of higher protein-to-carbohydrate ratio and plastic responses, but produce less pollen in a short time span. It is interesting to notice the differences in amide I banding patterns between the populations as amide I has been shown to prevent loss of hydrogen bonds during heat stress, which could be adaptive in the Italian population (Lahlali *et al.*, 2014). Within species, all populations of *P. alpina* had specific spectra whereas the differences were less prominent in *A. odoratum* and *F. ovina*. The distinct profiles of the *P. alpina* populations may be related to the discontinuous distribution pattern of the species (Fjellheim *et al.*, 2015), possibly also in combination with limited pollen production period (Figure 3). This will prevent gene flow between populations which over time will diverge. The opposite is the case for *F. ovina* and *A. odoratum* which have enormous potential for gene flow between populations through prolonged pollen production period and continuous geographic distribution. Norwegian populations of *P. alpina* are more divergent than populations of *F. ovina*, most likely reflecting a difference in gene flow (Jørgensen *et al.*, 2016). The outcrossing temperate forage grass *Phleum pratensis* is a species very similar in biology and distribution as *F. ovina* and *A. odoratum* and show no genetic structure in the entire distribution area reflecting large amount of gene flow (Fjellheim *et al.*, 2015). It is likely that *F. ovina* and *A. odoratum* also exhibit large amount of gene flow which will counteract divergence between populations.

CONCLUSION

The value of FITR as a method of classification on higher taxonomic levels has been shown in several papers (Zimmermann, 2010; Zimmermann & Kohler, 2014; Bağcıoğlu *et al.*, 2015; Zimmermann *et al.*, 2015a). In this study we show that classification of species is extremely accurate also on a subfamily level. Furthermore, our study indicate that pollen composition has an adaptive role and should be investigated in more detail in future studies. Prospected climate change may have an influence on pollen quality and the potential for species to respond to those changes through adaptive evolution remain unexplored. Global climate change is already disrupting plant flowering phenology across the northern hemisphere, with higher temperatures related to earlier onset dates (Menzel *et al.*, 2006). It has been shown that temperature stress can have a damaging effect on development, transport and germination of pollen (Bokszczanin & Fragkostefanakis,

519 2013). As one of the two key components of pollination, the consequence of environmental
520 changes on pollen composition should be investigated in more detail. More studies are needed to
521 study if plasticity in pollen composition is adaptive and to see the direct effects of pollen
522 composition on fitness under different conditions. FTIR coupled with controlled environment
523 studies of plants and effect on fitness of pollen with different composition has the potential to shed
524 new light on the evolutionary significance of plant male function and adaptation.

525 **Acknowledgements**

526 The research was supported by the European Commission through the Seventh Framework
527 Programme (FP7-PEOPLE-2012-IEF project No. 328289).

528 **Supporting Information**

529 It contains additional information on sample set, hierarchical classification, confusion matrices,
530 regression coefficients, and principal component analysis.

531 **REFERENCES**

532 **Alcaraz ML, Montserrat M, Hormaza JI. 2011.** In vitro pollen germination in avocado (Persea americana
533 Mill.): Optimization of the method and effect of temperature. *Scientia Horticulturae* **130**(1): 152-
534 156.
535 **Andersen TS, Bertelsen F. 1972.** Scanning Electron Microscope Studies of Pollen of Cereals and other
536 Grasses. *Grana* **12**(2): 79-86.
537 **Arceo-Gomez G, Abdala-Roberts L, Jankowiak A, Kohler C, Meindl GA, Navarro-Fernandez CM, Parra-**
538 **Tabla V, Ashman TL, Alonso C. 2016.** Patterns of among- and within-species variation in
539 heterospecific pollen receipt: The importance of ecological generalization. *American Journal of*
540 *Botany* **103**(3): 396-407.
541 **Bağcıoğlu M, Zimmermann B, Kohler A. 2015.** A Multiscale Vibrational Spectroscopic Approach for
542 Identification and Biochemical Characterization of Pollen. *PLoS ONE* **10**(9): e0137899.
543 **Baker HG, Baker I. 1979.** Starch in Angiosperm Pollen Grains and Its Evolutionary Significance. *American*
544 *Journal of Botany* **66**(5): 591-600.
545 **Bassani M, Pacini E, Franchi GG. 1994.** Humidity stress responses in pollen of anemophilous and
546 entomophilous species. *Grana* **33**(3): 146-150.
547 **Blackmore S, Wortley AH, Skvarla JJ, Rowley JR. 2007.** Pollen wall development in flowering plants. *New*
548 *Phytologist* **174**(3): 483-498.
549 **Bokszczanin KL, Fragkostefanakis S. 2013.** Perspectives on deciphering mechanisms underlying plant heat
550 stress response and thermotolerance. *Front Plant Sci* **4**: 315.
551 **Burd M. 2008.** The Haig-Westoby model revisited. *American Naturalist* **171**(3): 400-404.
552 **Buta E, Cantor M, Stefan R, Pop R, Mitre I, Buta M, Sestras RE. 2015.** FT-IR Characterization of Pollen
553 Biochemistry, Viability, and Germination Capacity in Saintpaulia H. Wendl. Genotypes. *Journal of*
554 *Spectroscopy*.
555 **Dell'Anna R, Lazzeri P, Frisanco M, Monti F, Malvezzi Campeggi F, Gottardini E, Bersani M. 2009.** Pollen
556 discrimination and classification by Fourier transform infrared (FT-IR) microspectroscopy and
557 machine learning. *Anal Bioanal Chem* **394**(5): 1443-1452.
558 **Dickinson GH, Elleman JC, Doughty J.** Pollen coatings – chimaeric genetics and new functions. *Sexual Plant*
559 *Reproduction* **12**(5): 302-309.

Dominguez E, Mercado JA, Quesada MA, Heredia A. 1999. Pollen sporopollenin: degradation and structural elucidation. *Sexual Plant Reproduction* **12**(3): 171-178.

Fjellheim S, Tanhuanpaa P, Marum P, Manninen O, Rognli OA. 2015. Phenotypic or Molecular Diversity Screening for Conservation of Genetic Resources? An Example from a Genebank Collection of the Temperate Forage Grass Timothy. *Crop Science* **55**(4): 1646-1659.

Gottardini E, Rossi S, Cristofolini F, Benedetti L. 2007. Use of Fourier transform infrared (FT-IR) spectroscopy as a tool for pollen identification. *Aerobiologia* **23**(3): 211-219.

Hedhly A, Hormaza JI, Herrera M. 2004. Effect of temperature on pollen tube kinetics and dynamics in sweet cherry, Prunus avium (Rosaceae). *American Journal of Botany* **91**(4): 558-564.

Hesse M. 2006. Conventional and novel modes of exine patterning in members of the Araceae – the consequence of ecological paradigm shifts? *Protoplasma* **228**(1): 145-149.

Hoekstra FA, Crowe LM, Crowe JH. 1989. Differential desiccation sensitivity of corn and Pennisetum pollen linked to their sucrose contents. *Plant Cell Environ* **12**(1): 83-91.

Holmes-Davis R, Tanaka CK, Vensel WH, Hurkman WJ, McCormick S. 2005. Proteome mapping of mature pollen of Arabidopsis thaliana. *Proteomics* **5**(18): 4864-4884.

Jiang YF, Lahlali R, Karunakaran C, Kumar S, Davis AR, Bueckert RA. 2015. Seed set, pollen morphology and pollen surface composition response to heat stress in field pea. *Plant Cell and Environment* **38**(11): 2387-2397.

Jones CE, Little RJ. 1983. *Handbook of experimental pollination biology*: Scientific and Academic Editions.

Jørgensen MH, Elameen A, Hofman N, Klemsdal S, Malaval S, Fjellheim S. 2016. What's the meaning of local? Using molecular markers to define seed transfer zones for ecological restoration in Norway. *Evolutionary Applications*: n/a-n/a.

Kakani VG, Reddy KR, Koti S, Wallace TP, Prasad PVV, Reddy VR, Zhao D. 2005. Differences in in vitro pollen germination and pollen tube growth of cotton cultivars in response to high temperature. *Annals of Botany* **96**(1): 59-67.

Karaman I, Qannari E, Martens H, Hedemann MS, Knudsen KEB, Kohler A. 2013. Comparison of Sparse and Jack-knife partial least squares regression methods for variable selection. *Chemometrics and Intelligent Laboratory Systems* **122**: 65-77.

Kellogg EA. 2001. Evolutionary history of the grasses. *Plant Physiology* **125**(3): 1198-1205.

Koti S, Reddy KR, Kakani VG, Zhao D, Reddy VR. 2004. Soybean (Glycine max) pollen germination characteristics, flower and pollen morphology in response to enhanced ultraviolet-B radiation. *Annals of Botany* **94**(6): 855-864.

Lahlali R, Jiang Y, Kumar S, Karunakaran C, Liu X, Borondics F, Hallin E, Bueckert R. 2014. ATR–FTIR spectroscopy reveals involvement of lipids and proteins of intact pea pollen grains to heat stress tolerance. *Frontiers in Plant Science* **5**: 747.

Lau T-C, Stephenson AG. 1993. Effects of Soil Nitrogen on Pollen Production, Pollen Grain Size, and Pollen Performance in Cucurbita pepo (Cucurbitaceae). *American Journal of Botany* **80**(7): 763-768.

Lau TC, Stephenson AG. 1994. Effects of Soil-Phosphorus on Pollen Production, Pollen Size, Pollen Phosphorus-Content, and the Ability to Sire Seeds in Cucurbita-Pepo (Cucurbitaceae). *Sexual Plant Reproduction* **7**(4): 215-220.

Lee Rodgers J, Nicewander WA. 1988. Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician* **42**(1): 59-66.

Lidforss B 1899. Weitere Beiträge zur Biologie des Pollens. *Jahrbücher für Wissenschaftliche Botanik*: Verlag von Gebrüder Borntraeger, Berlin, 232-312.

Lindgren F, Hansen B, Karcher W, Sjostrom M, Eriksson L. 1996. Model validation by permutation tests: Applications to variable selection. *Journal of Chemometrics* **10**(5-6): 521-532.

Lora J, Testillano PS, Risueno MC, Hormaza JI, Herrero M. 2009. Pollen development in Annona cherimola Mill. (Annonaceae). Implications for the evolution of aggregated pollen. *Bmc Plant Biology* **9**.

Mander L, Li M, Mio W, Fowlkes CC, Punyasena SW. 2013. Classification of grass pollen through the quantitative analysis of surface ornamentation and texture. *Proceedings of the Royal Society B-Biological Sciences* **280**(1770).

Matsushima R, Hamamura Y, Higashiyama T, Arimura S, Sodmergen, Tsutsumi N, Sakamoto W. 2008. Mitochondrial dynamics in plant male gametophyte visualized by fluorescent live imaging. *Plant and Cell Physiology* **49**(7): 1074-1083.

Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, Alm-KÜBler K, Bissolli P, BraslavskÁ OG, Briede A, et al. 2006. European phenological response to climate change matches the warming pattern. *Global Change Biology* **12**(10): 1969-1976.

Molisch H. 1893. *Zur physiologie des pollens mit besonderer Rucksicht auf die chemotropische Bewegungen der Pollenschlauche.* Wien: Akad. Wiss Mathnaturw.

Pappas CS, Tarantilis PA, Harizanis PC, Polissiou MG. 2003. New method for pollen identification by FT-IR spectroscopy. *Applied Spectroscopy* **57**(1): 23-27.

Pham VT, Herrero M, Hormaza JI. 2015. Effect of temperature on pollen germination and pollen tube growth in longan (Dimocarpus longan Lour.). *Scientia Horticulturae* **197**: 470-475.

Piffanelli P, Ross EJH, Murphy JD. Biogenesis and function of the lipidic structures of pollen grains. *Sexual Plant Reproduction* **11**(2): 65-80.

Rang ZW, Jagadish SVK, Zhou QM, Craufurd PQ, Heuer S. 2011. Effect of high temperature and water stress on pollen germination and spikelet fertility in rice. *Environmental and Experimental Botany* **70**(1): 58-65.

Rosenheim JA, Williams NM, Schreiber SJ. 2014. Parental Optimism versus Parental Pessimism in Plants: How Common Should We Expect Pollen Limitation to Be? *American Naturalist* **184**(1): 75-90.

Roulston TaH, Cane JH, Buchmann SL. 2000. What Governs Protein Content of Pollen: Pollinator Preferences, Pollen-Pistil Interactions, or Phylogeny? *Ecological Monographs* **70**(4): 617-643.

Salih A, Jones AS, Bass D, Cox G. 1997. Confocal imaging of exine as a tool for grass pollen analysis. *Grana* **36**(4): 215-224.

Savitzky A, Golay MJE. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**: 1627ff.

Sears P, Metcalf E. 1926. The behavior of pollen starch in a geranium and its bud sport. *Journal of Genetics* **17**: 33-42.

Sivaguru M, Mander L, Fried G, Punyasena SW. 2012. Capturing the Surface Texture and Shape of Pollen: A Comparison of Microscopy Techniques. *Plos One* **7**(6).

Speranza A, Calzoni GL, Pacini E. 1997. Occurrence of mono- or disaccharides and polysaccharide reserves in mature pollen grains. *Sexual Plant Reproduction* **10**(2): 110-115.

Stanley RG, Linskens HF. 1974. *Pollen: biology, biochemistry, management*: Springer-Verlag.

Suwabe K, Suzuki G, Takahashi H, Shiono K, Endo M, Yano K, Fujita M, Masuko H, Saito H, Fujioka T, et al. 2008. Separated Transcriptomes of Male Gametophyte and Tapetum in Rice: Validity of a Laser Microdissection (LM) Microarray. *Plant and Cell Physiology* **49**(10): 1407-1416.

Szymanska E, Saccenti E, Smilde AK, Westerhuis JA. 2012. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **8**(1): S3-S16.

Torres C. Pollen size evolution: correlation between pollen volume and pistil length in Asteraceae. *Sexual Plant Reproduction* **12**(6): 365-370.

Van Herpen MMA 1986. Biochemical Alterations in the Sexual Partners Resulting from Environmental Conditions before Pollination Regulate Processes after Pollination. In: Mulcahy DL, Mulcahy GB, Ottaviano E eds. *Biotechnology and Ecology of Pollen: Proceedings of the International Conference on the Biotechnology and Ecology of Pollen, 9–11 July, 1985, University of Massachusetts, Amherst, MA, USA*. New York, NY: Springer New York, 131-133.

Van Herpen MMA, Linskens HF. 1981. EFFECT OF SEASON, PLANT AGE AND TEMPERATURE DURING PLANT GROWTH ON COMPATIBLE AND INCOMPATIBLE POLLEN TUBE GROWTH IN PETUNIA HYBRIDA. *Acta Botanica Neerlandica* **30**(3): 209-218.

Wang Z-Y, Ge Y, Scott M, Spangenberg G. 2004. Viability and longevity of pollen from transgenic and nontransgenic tall fescue (Festuca arundinacea) (Poaceae) plants. *American Journal of Botany* **91**(4): 523-530.

Williams EG, Rouse JL. Relationships of pollen size, pistil length and pollen tube growth rates in Rhododendron and their influence on hybridization. *Sexual Plant Reproduction* **3**(1): 7-17.

Williams JH, Mazer SJ. 2016. Pollen-Tiny and ephemeral but not forgotten: New ideas on their ecology and evolution. *American Journal of Botany* **103**(3): 365-374.

Wolters-Arts M, Lush WM, Mariani C. 1998. Lipids are required for directional pollen-tube growth. *Nature* **392**(6678): 818-821.

Zimmerman B, Tafintseva V, Bagcioglu M, Hoegh Berdahl M, Kohler A. 2016. Analysis of Allergenic Pollen by FTIR Microspectroscopy. *Anal Chem* **88**(1): 803-811.

Zimmermann B. 2010. Characterization of pollen by vibrational spectroscopy. *Appl Spectrosc* **64**(12): 1364-1373.

Zimmermann B, Bagcioglu M, Sandt C, 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**(5): 1237-1250.

Zimmermann B, Kohler A. 2014. Infrared Spectroscopy of Pollen Identifies Plant Species and Genus as Well as Environmental Conditions. *PLoS ONE* **9**(4): e95417.

Zimmermann B, Tkalcec Z, Mesic A, Kohler A. 2015b. Characterizing aeroallergens by infrared spectroscopy of fungal spores and pollen. *Plos One* **10**(4): e0124240.

Supporting Information

A greenhouse study for assessing environmental effects on pollen composition by FTIR spectroscopy

Boris Zimmermann[1], Murat Bağcıoğlu[1], Valeria Tafinstseva[1], Achim Kohler[1,2], Mikael Ohlson[3], Siri Fjellheim[4]*

[1]Department of Mathematical Sciences and Technology, Faculty of Environmental Science and Technology, Norwegian University of Life Sciences, 1432, Ås, Norway

[2]Nofima AS, 1430, Ås, Norway

[3]Department of Ecology and Natural Resource Management, Faculty of Environmental Science and Technology, Norwegian University of Life Sciences, 1432, Ås, Norway

[4]Department of Plant Sciences, Faculty of Veterinary Medicine and Biosciences, Norwegian University of Life Sciences, 1432, Ås, Norway

*Corresponding author:
Siri Fjellheim
[4]Department of Plant Sciences
Faculty of Veterinary Medicine and Biosciences
Norwegian University of Life Sciences
Drøbakveien 31, 1432 Ås, Norway.
Tel: +47 6723 2801
Faks: +47 6496 5001
E-mail: siri.fjellheim@nmbu.no

Table of Contents **Page**

	Anthoxanthum odoratum			*Festuca ovina*			*Poa alpina*		
	France	Greece	Finland	Sweden	Finland	Italy	Sweden	Italy	Norway
14 °C +NU	11 (20)	14 (26)	15 (26)	15 (39)	8 (14)	4 (8)	15 (36)	12 (20)	14 (34)
14 °C - NU	11 (19)	14 (28)	15 (28)	15 (20)	8	4 (6)	15 (28)	14 (18)	14 (39)
20 °C +NU	11 (19)	14 (21)	13 (16)	15 (21)	8 (20)	4 (17)	15 (19)	14 (32)	14 (23)
20 °C - NU	11 (21)	14 (28)	14 (20)	15 (19)	8	4 (7)	15 (25)	13 (15)	14 (25)
4 clones	11	14	13	15	5	3	15	11	14

Table S1. List of samples with number of sampled individuals, and total number of pollen samples. Last row designates number of genotypes for which all four clones were sampled.

	Anthoxanthum odoratum			*Festuca ovina*			*Poa alpina*		
	France	Greece	Finland	Sweden	Finland	Italy	Sweden	Italy	Norway
14 °C +NU	11	12	12	12	8	4	12	12	12
14 °C - NU	11	12	12	12	8	4	12	12	12
20 °C +NU	11	12	12	12	8	4*	12	12	12
20 °C - NU	11	12	12	12	8	4	12	12	12
Total	44	48	48	48	32	16	48	48	48

Table S2. List of samples covered by *main FTIR study*. *9 additional pollen samples of the same individual plant, collected at different times, were covered by *FTIR timeline study*.

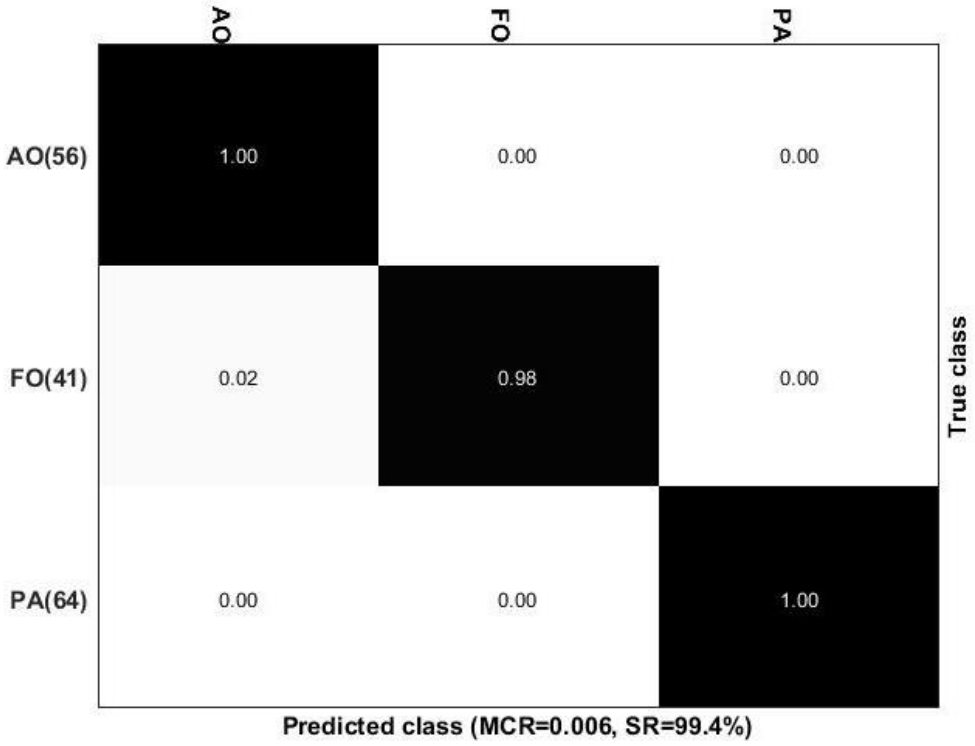


Fig. S1 Confusion matrix for hierarchical classification for 1st level (species); AO: *Anthoxanthum odoratum*, FO: *Festuca ovina*, PA: *Poa alpina*; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

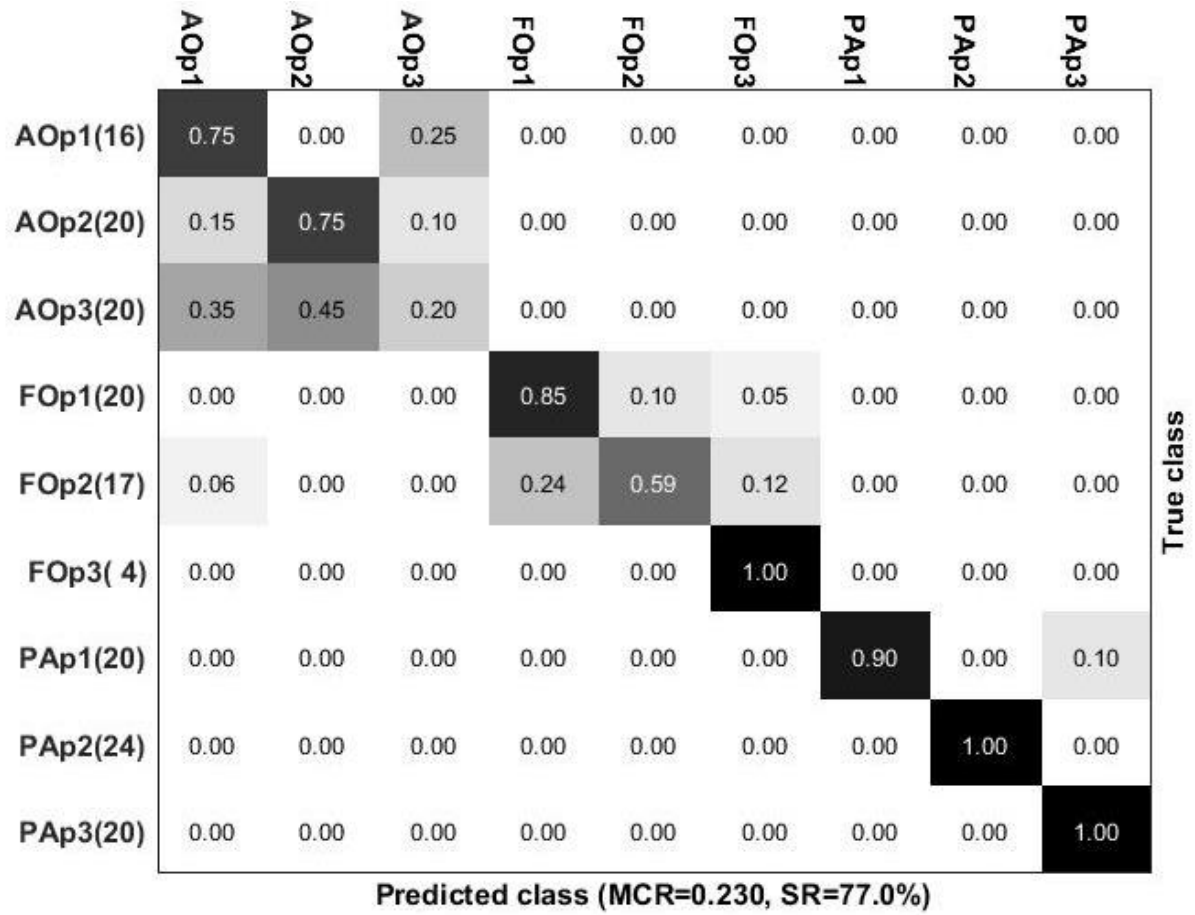


Fig. S2 Confusion matrix for hierarchical classification for 2nd level (population); AO: *Anthoxanthum odoratum* (p1: France, p2: Greece, p3: Finland), FO: *Festuca ovina* (p1: Sweden, p2: Finland, p3: Italy) PA: *Poa alpina* (p1: Sweden; p2: Italy, p3: Norway); number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

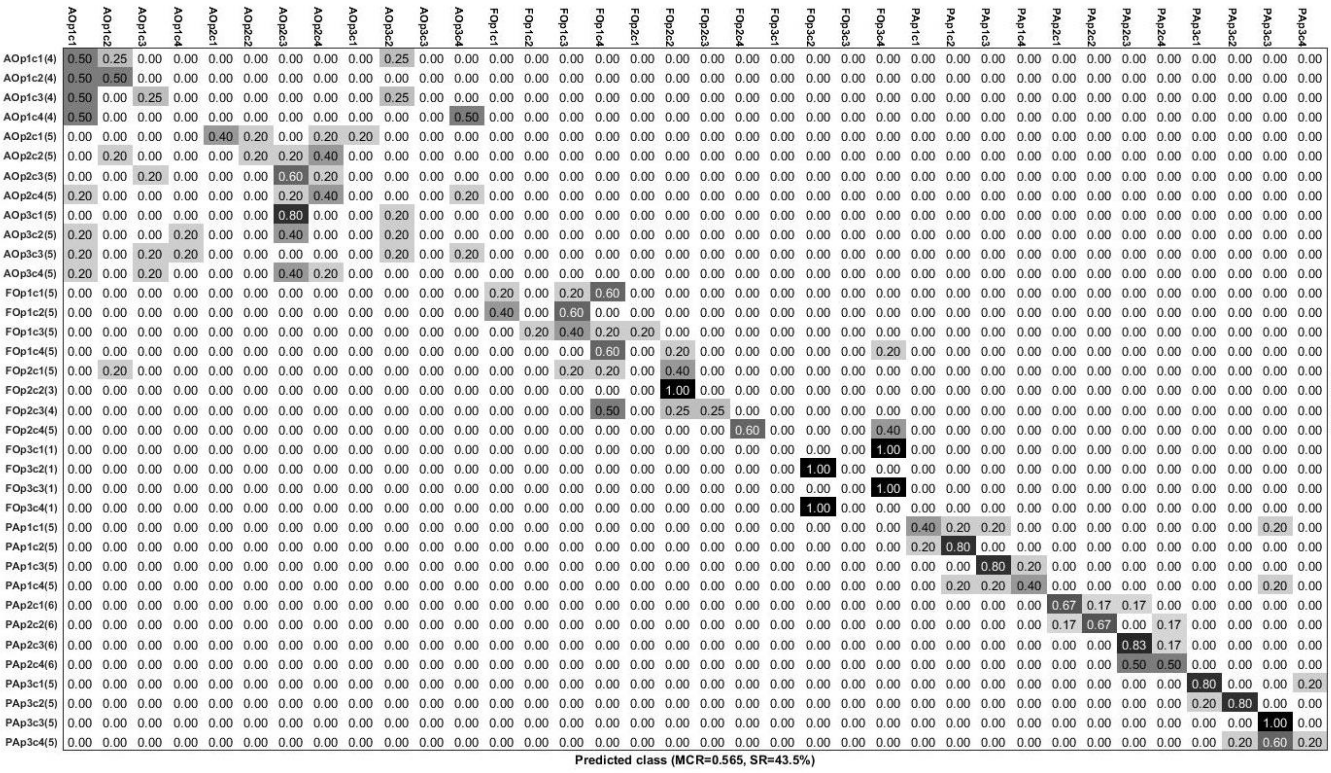


Fig. S3 Confusion matrix for hierarchical classification for 3rd level (growth conditions); AO: *Anthoxanthum odoratum* (p1: France, p2: Greece, p3: Finland), FO: *Festuca ovina* (p1: Sweden, p2: Finland, p3: Italy) PA: *Poa alpina* (p1: Sweden; p2: Italy, p3: Norway); c1: 14°C +NU, c2: 14°C -NU, c3: 20°C +NU, c4: 20°C –NU; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

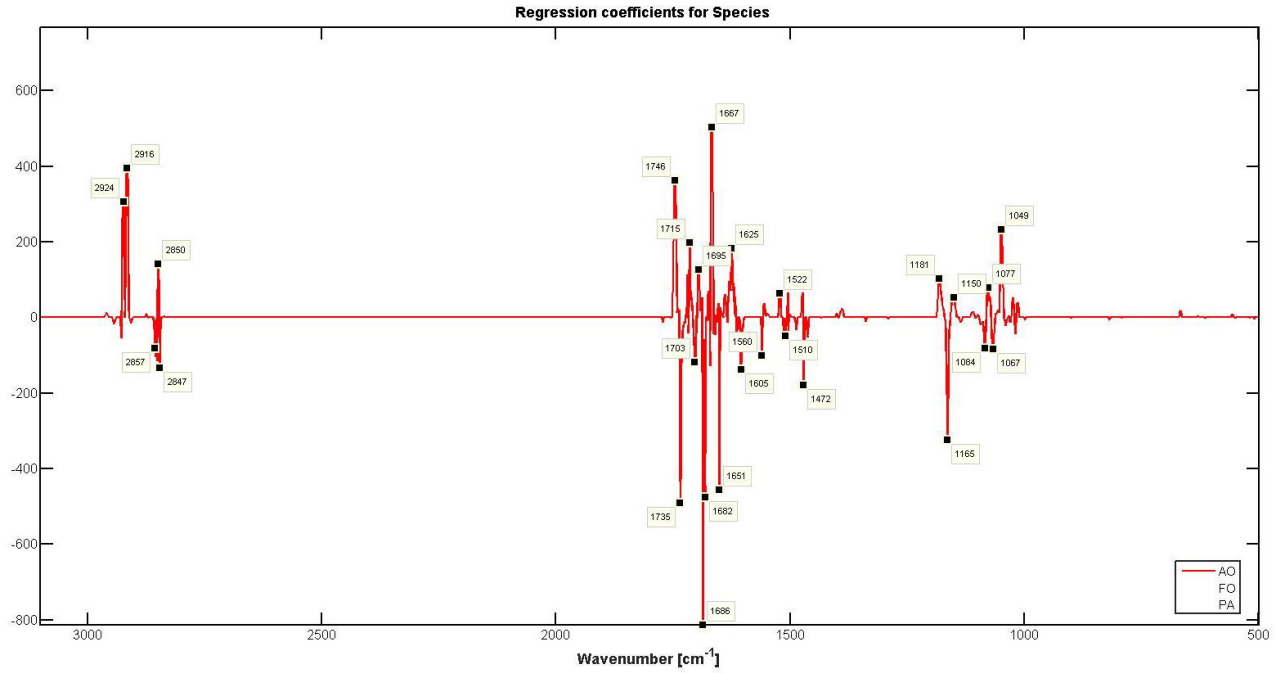


Fig. S4 (Part A) Hierarchical classification for 1st level (species): The regression coefficient for *Anthoxanthum odoratum.*

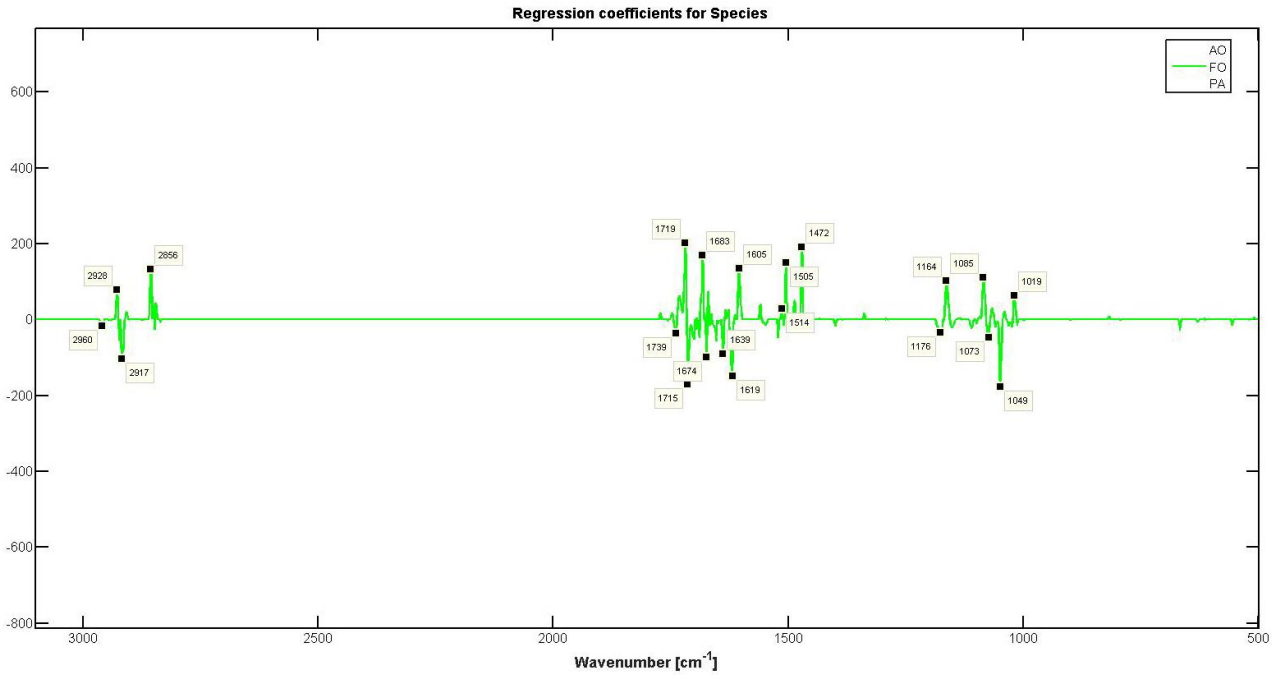


Fig. S4 (Part B) Hierarchical classification for 1st level (species): The regression coefficient for *Festuca ovina.*

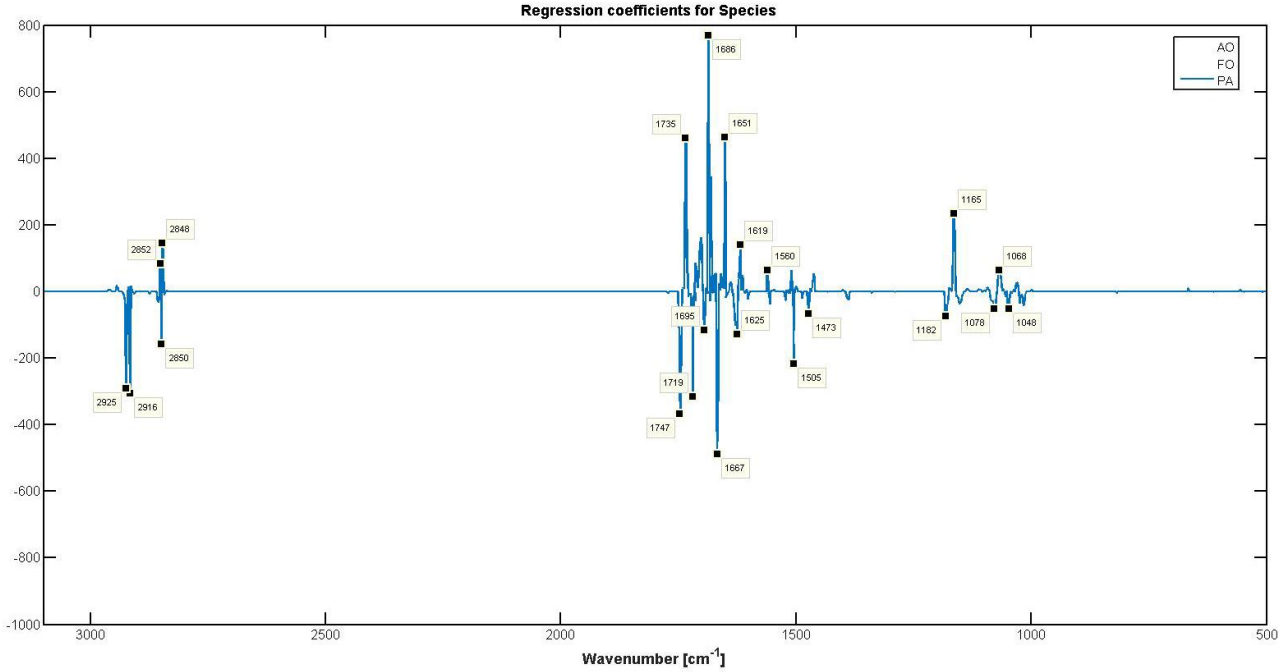


Fig. S4 (Part C) Hierarchical classification for 1st level (species): The regression coefficient for *Poa alpina.*

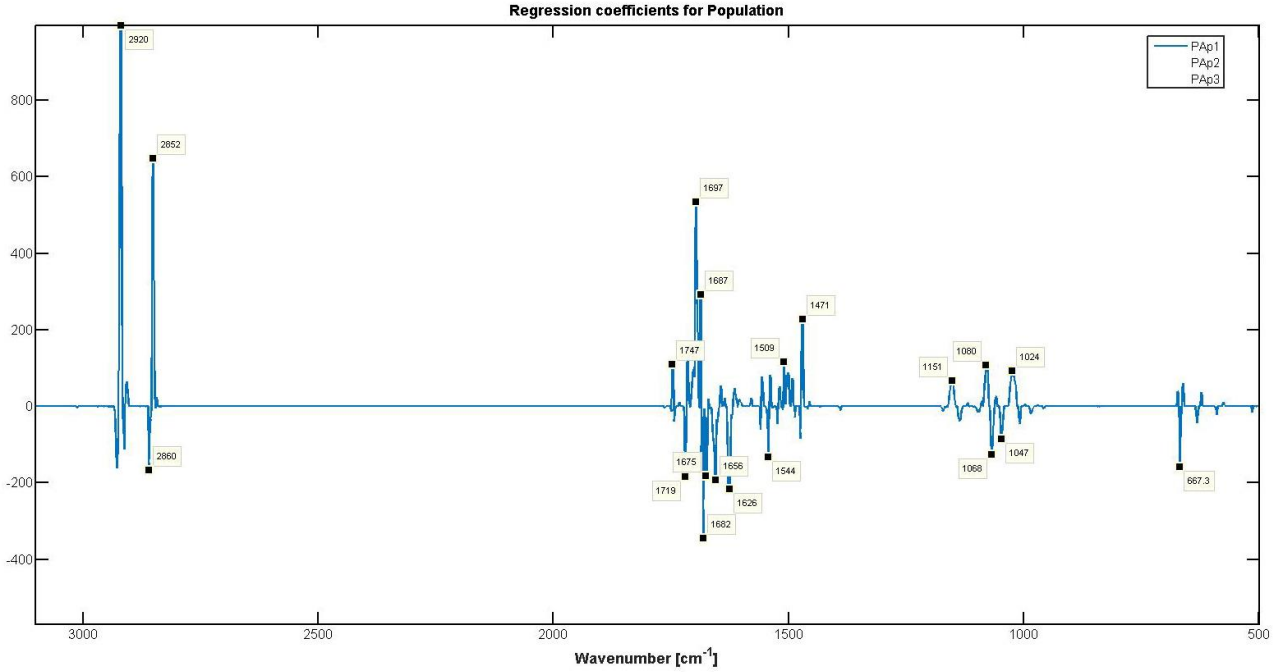


Fig. S5 (Part A) Hierarchical classification for 2nd level (population): The regression coefficient for *Poa alpina*, Sweden.

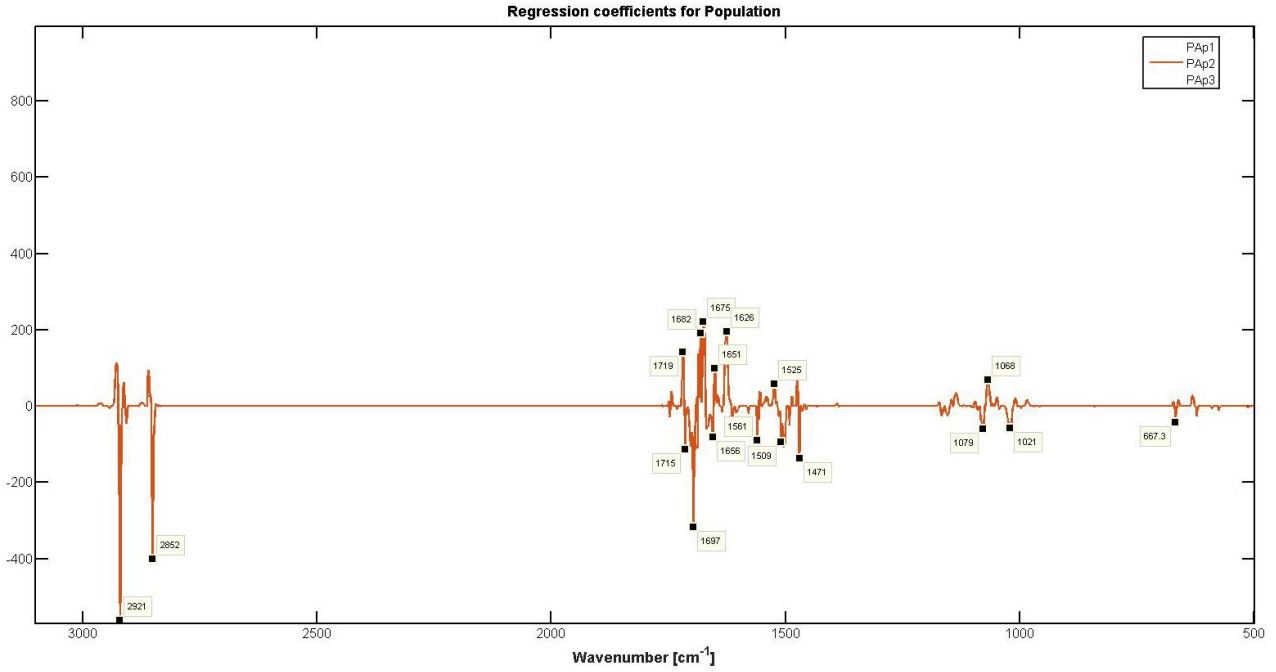


Fig. S5 (Part B) Hierarchical classification for 2nd level (population): The regression coefficient for *Poa alpina*, Italy.

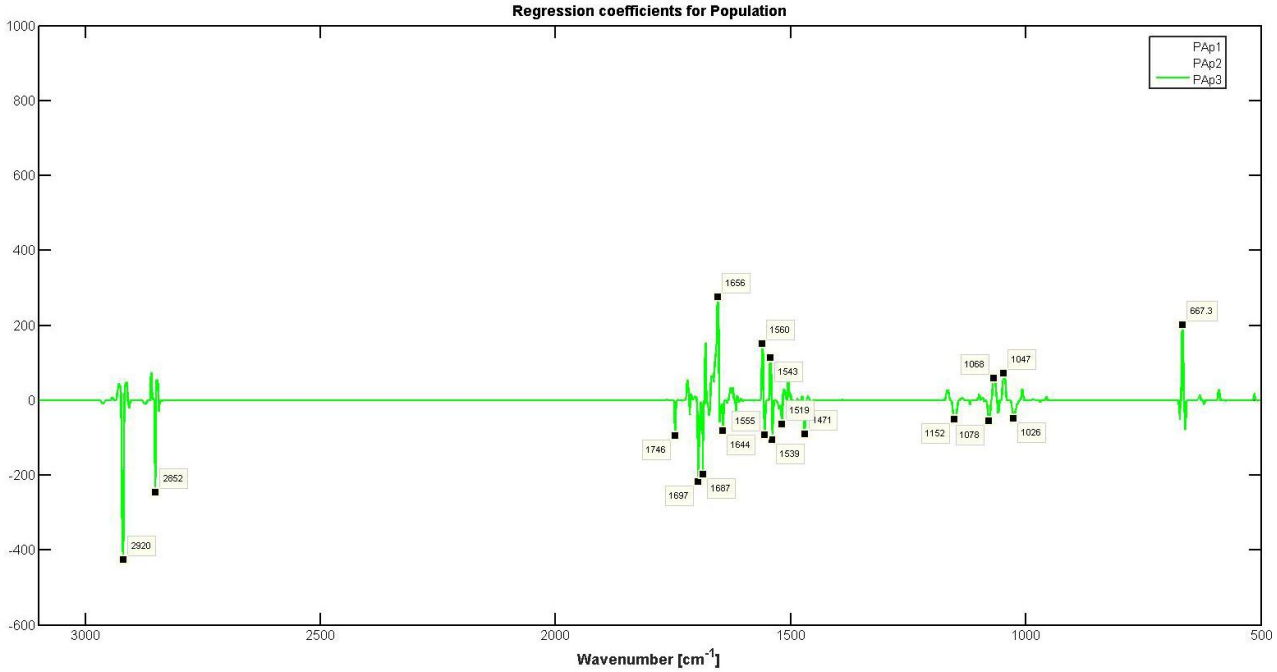


Fig. S5 (Part C) Hierarchical classification for 2nd level (population): The regression coefficient for *Poa alpina*, Norway.

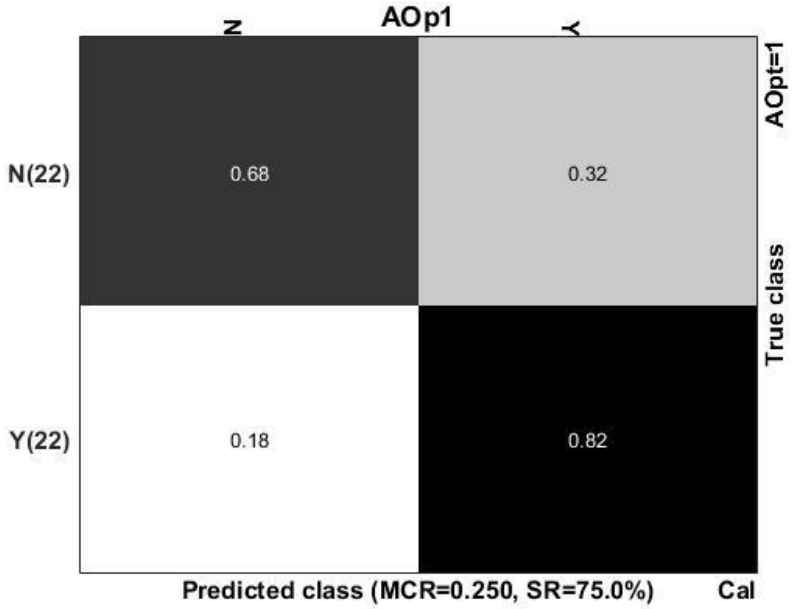


Fig. S6 (Part A) Confusion matrix for classification based on growth conditions; *Anthoxanthum odoratum*, France, N: -NU, Y: +NU; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

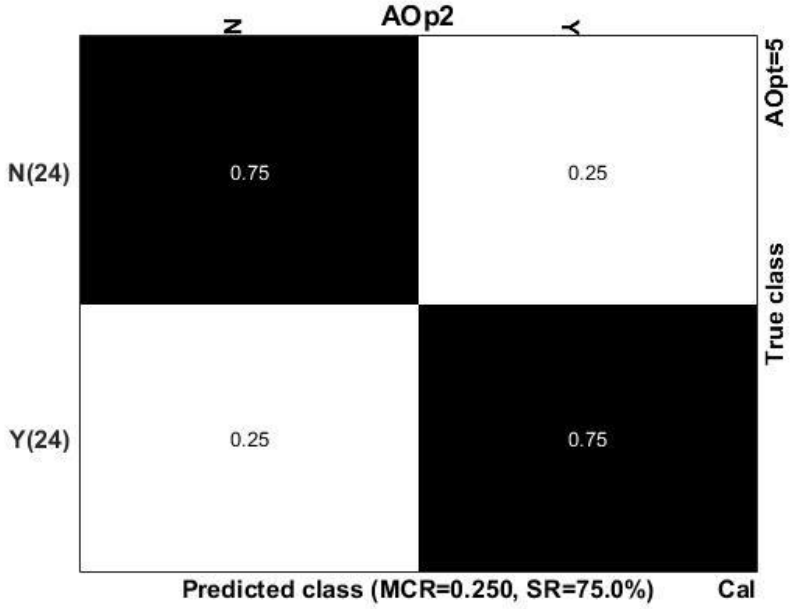


Fig. S6 (Part B) Confusion matrix for classification based on growth conditions; *Anthoxanthum odoratum*, Greece, N: -NU, Y: +NU; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

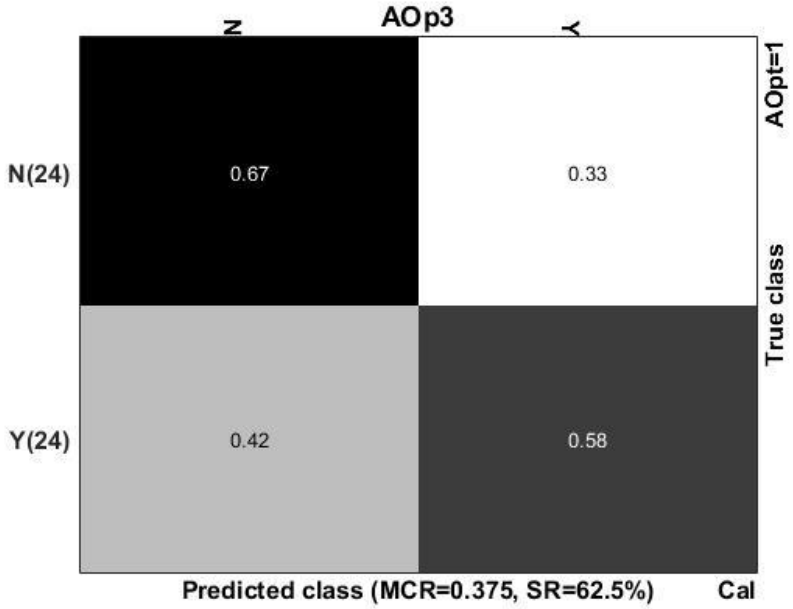


Fig. S6 (Part C) Confusion matrix for classification based on growth conditions; *Anthoxanthum odoratum*, Finland, N: -NU, Y: +NU; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

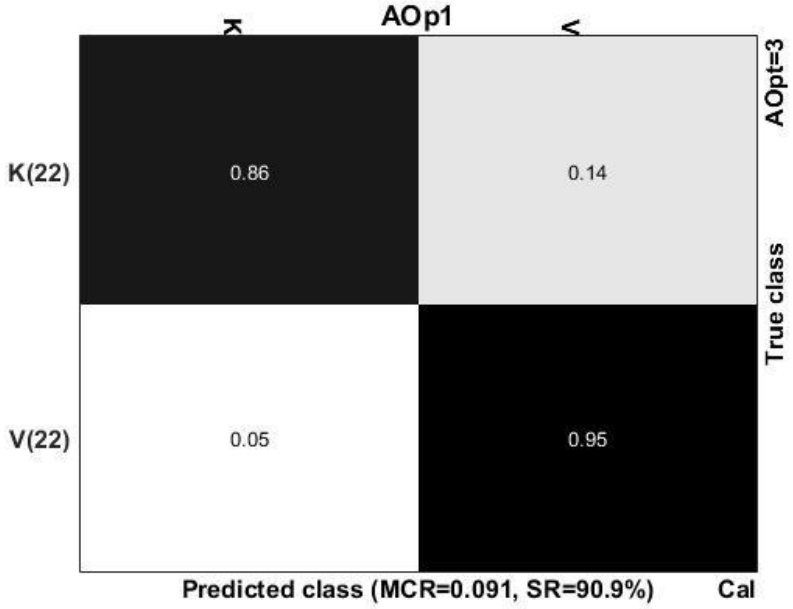


Fig. S6 (Part D) Confusion matrix for classification based on growth conditions; *Anthoxanthum odoratum*, France, K: 14 °C, V: 20 °C; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

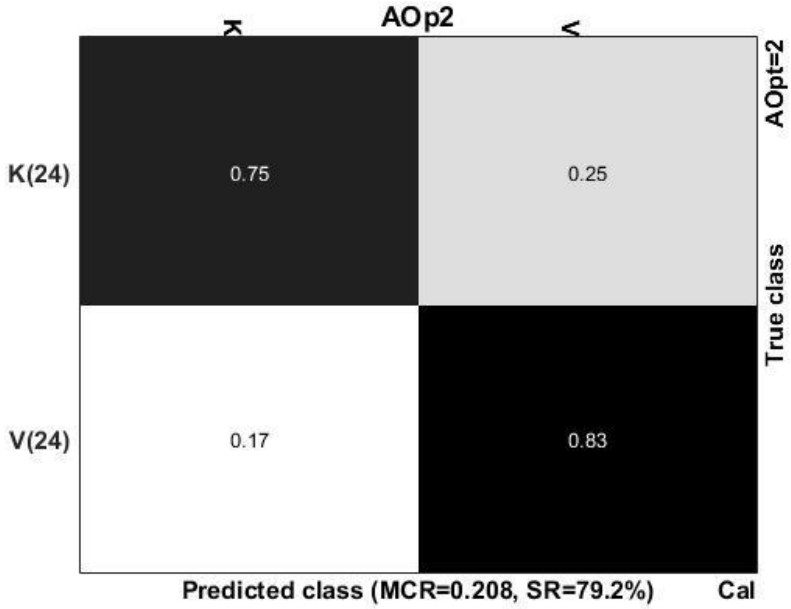


Fig. S6 (Part E) Confusion matrix for classification based on growth conditions; *Anthoxanthum odoratum*, Greece, K: 14 °C, V: 20 °C; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

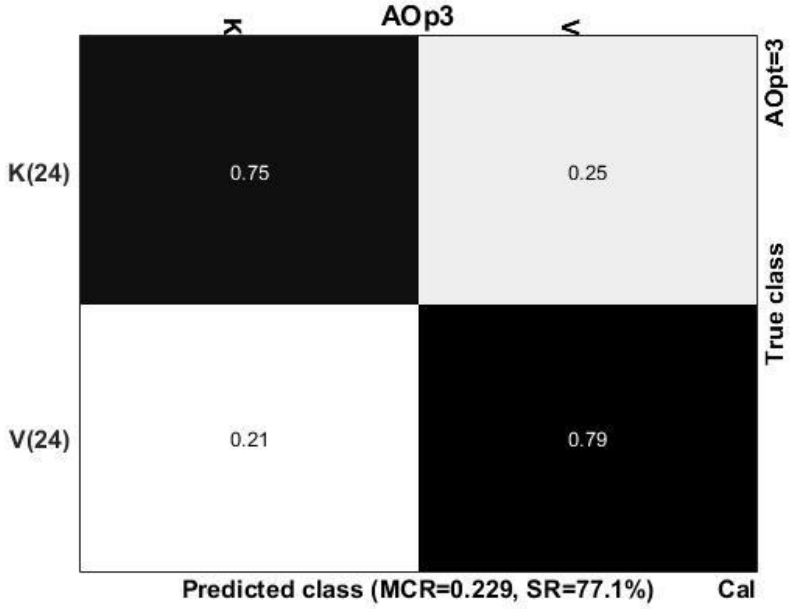


Fig. S6 (Part F) Confusion matrix for classification based on growth conditions; *Anthoxanthum odoratum*, Finland, K: 14 °C, V: 20 °C; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

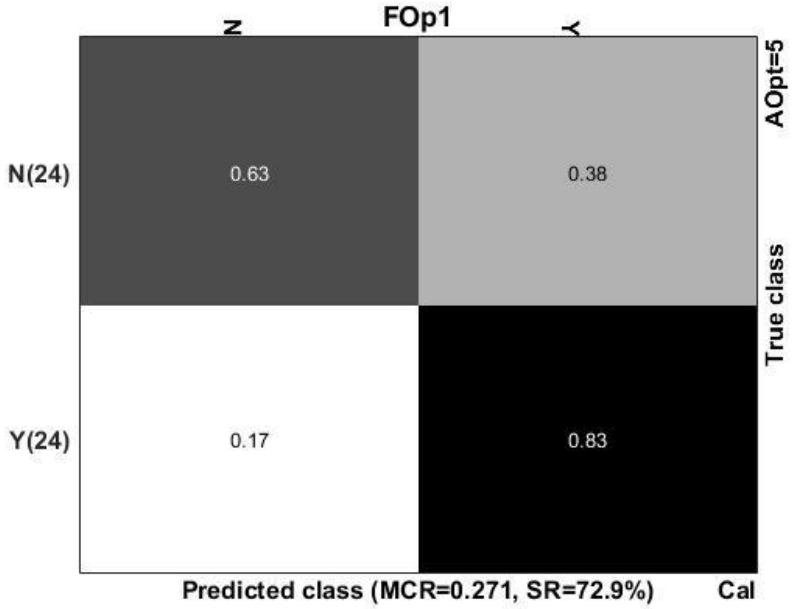


Fig. S7 (Part A) Confusion matrix for classification based on growth conditions; *Festuca ovina*, Sweden, N: -NU, Y: +NU; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

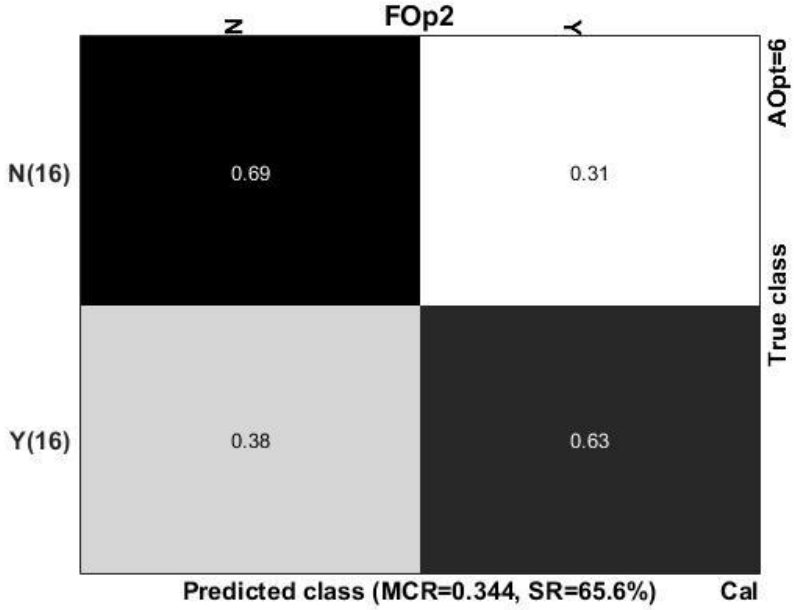


Fig. S7 (Part B) Confusion matrix for classification based on growth conditions; *Festuca ovina*, Finland, N: -NU, Y: +NU; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

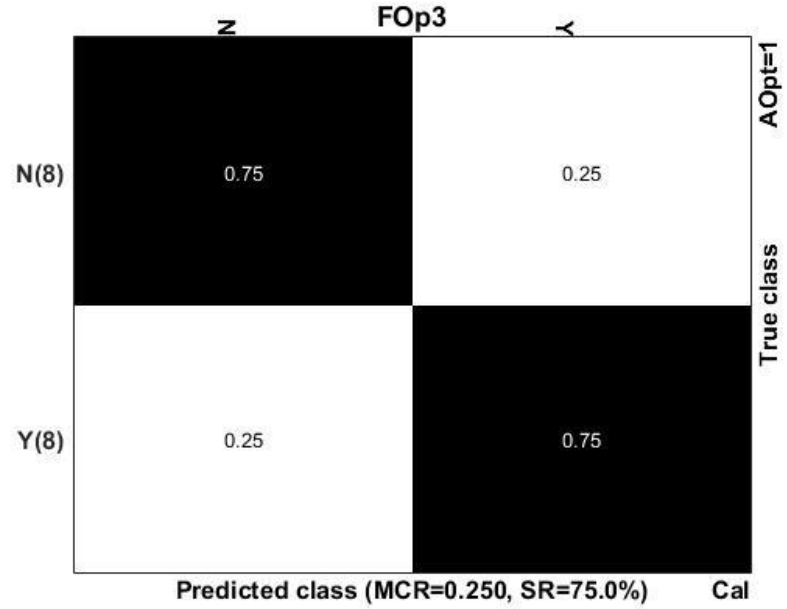


Fig. S7 (Part C) Confusion matrix for classification based on growth conditions; *Festuca ovina*, Italy, N: -NU, Y: +NU; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

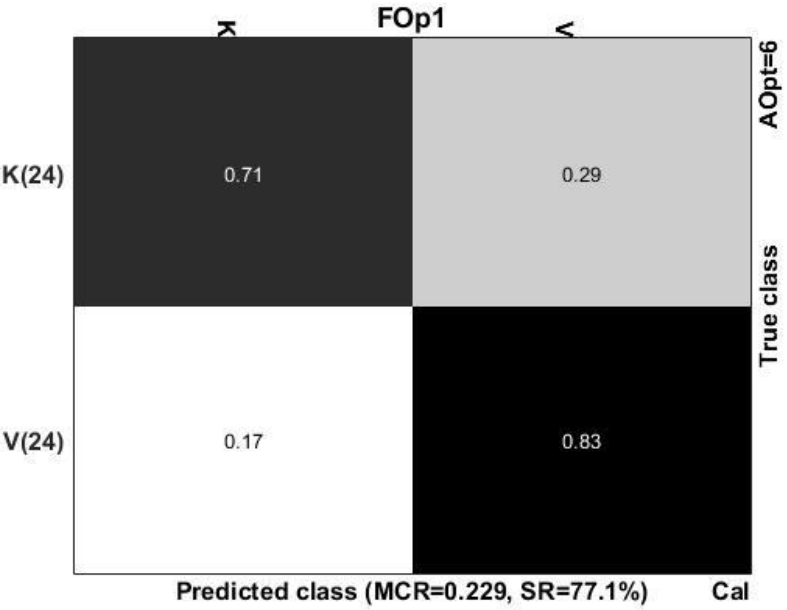


Fig. S7 (Part D) Confusion matrix for classification based on growth conditions; *Festuca ovina*, Sweden, K: 14 °C, V: 20 °C; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

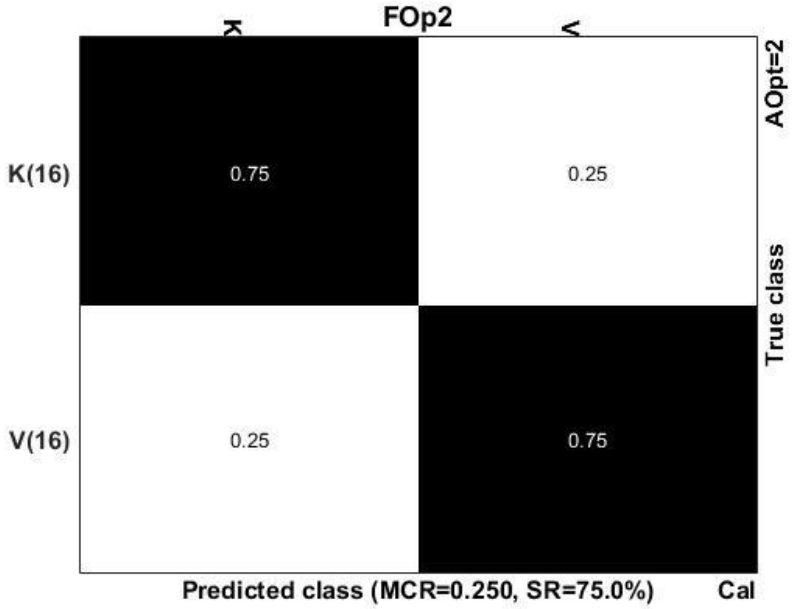


Fig. S7 (Part E) Confusion matrix for classification based on growth conditions; *Festuca ovina*, Finland, K: 14 °C, V: 20 °C; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

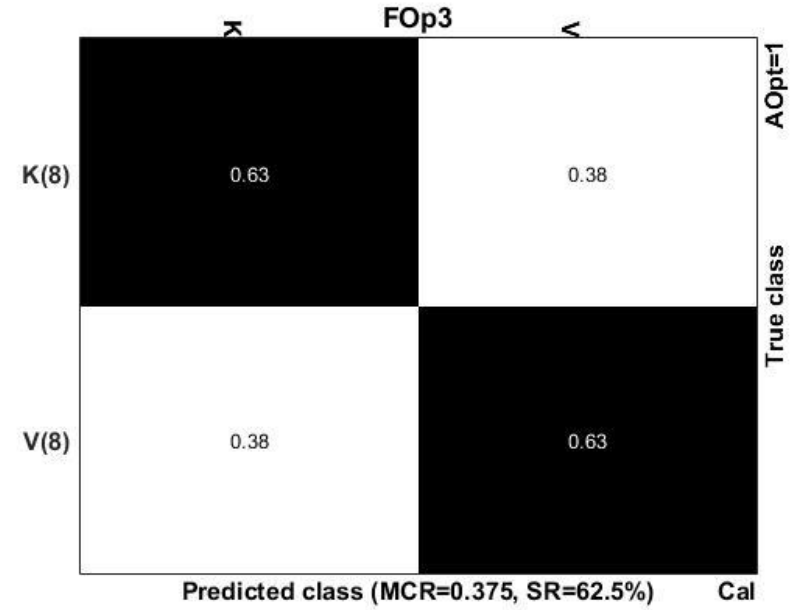


Fig. S7 (Part F) Confusion matrix for classification based on growth conditions; *Festuca ovina*, Italy, K: 14 °C, V: 20 °C; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

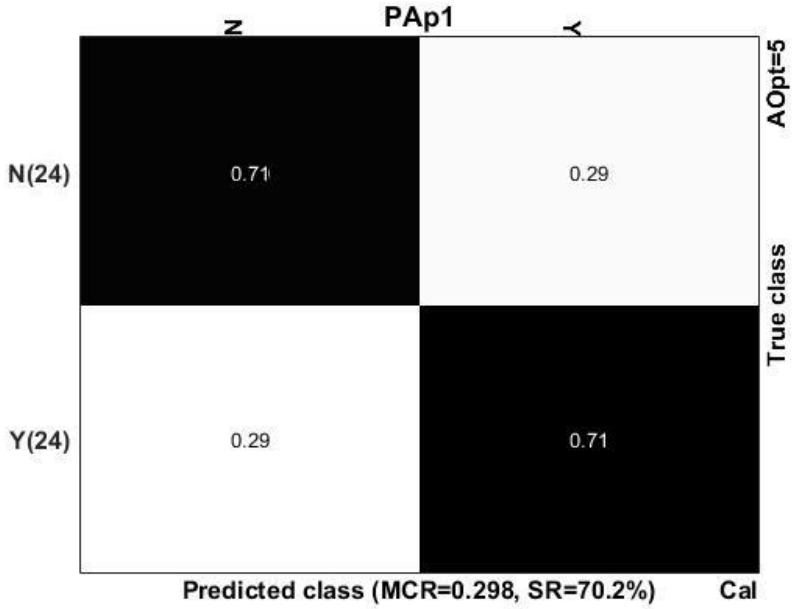


Fig. S8 (Part A) Confusion matrix for classification based on growth conditions; *Poa alpina*, Sweden, N: -NU, Y: +NU; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

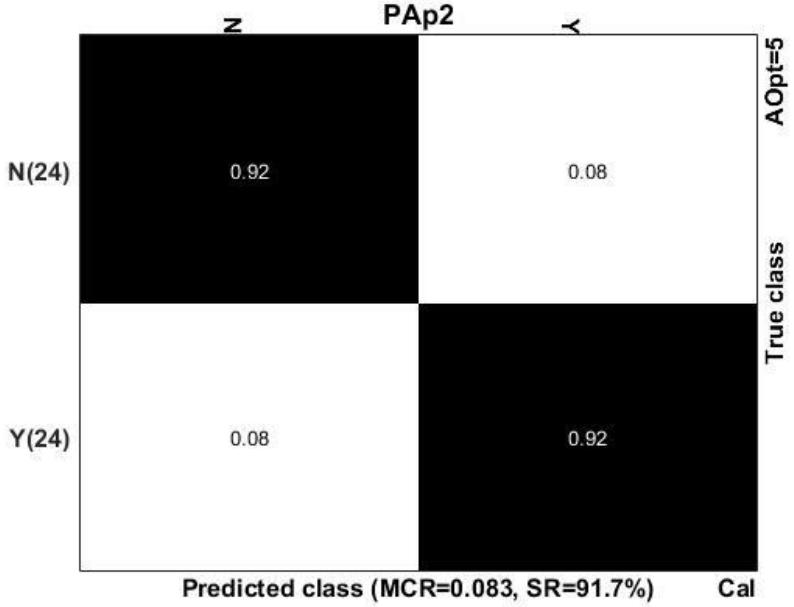


Fig. S8 (Part B) Confusion matrix for classification based on growth conditions; *Poa alpina*, Italy, N: -NU, Y: +NU; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

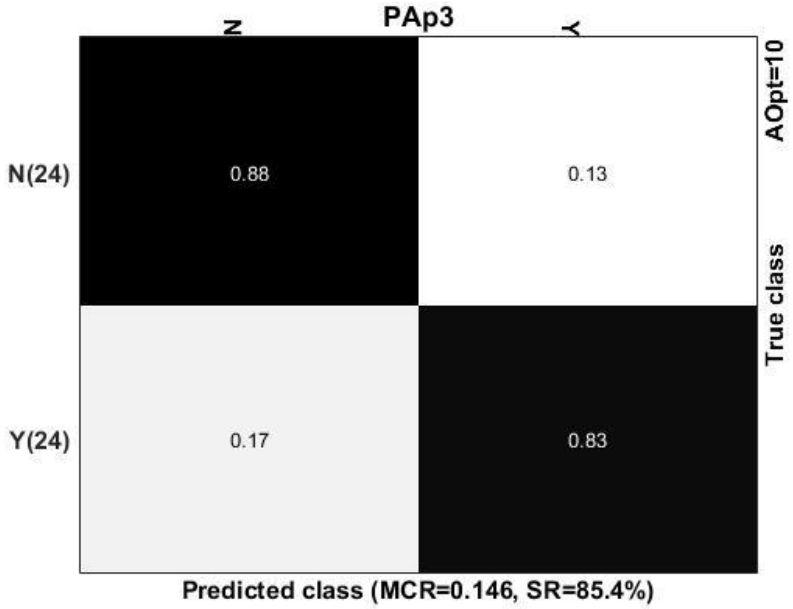


Fig. S8 (Part C) Confusion matrix for classification based on growth conditions; *Poa alpina*, Norway, N: -NU, Y: +NU; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

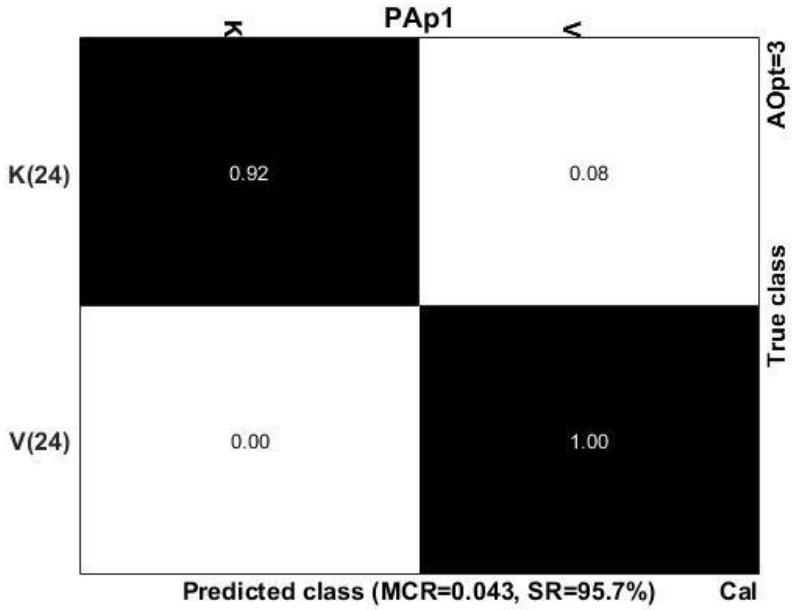


Fig. S8 (Part D) Confusion matrix for classification based on growth conditions; *Poa alpina*, Sweden, K: 14 °C, V: 20 °C; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

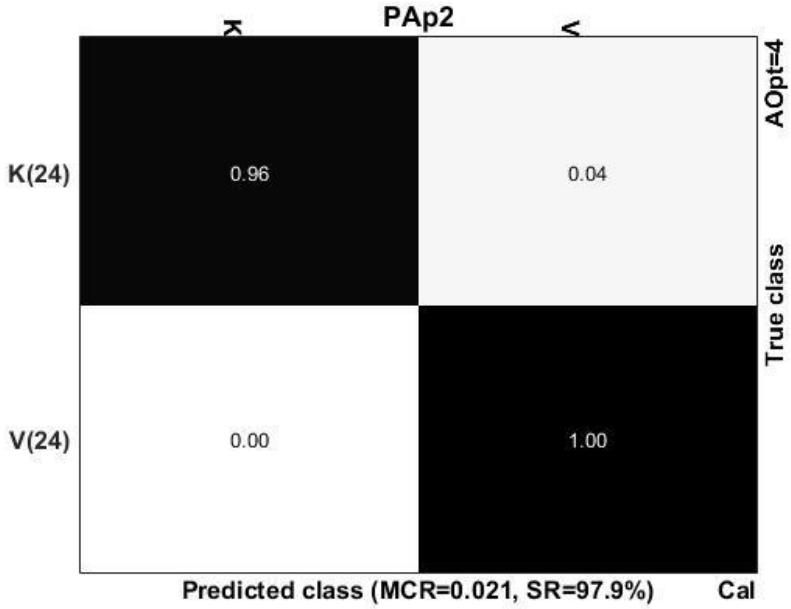


Fig. S8 (Part E) Confusion matrix for classification based on growth conditions; *Poa alpina*, Italy, K: 14 °C, V: 20 °C; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

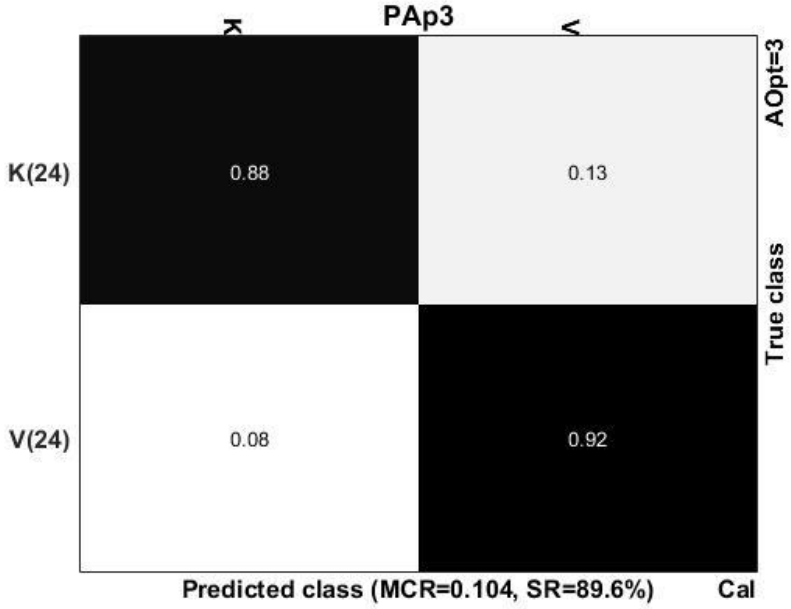


Fig. S8 (Part F) Confusion matrix for classification based on growth conditions; *Poa alpina*, Norway, K: 14 °C, V: 20 °C; number of spectra in parenthesis; MCR: misclassification rate, SR: success rate.

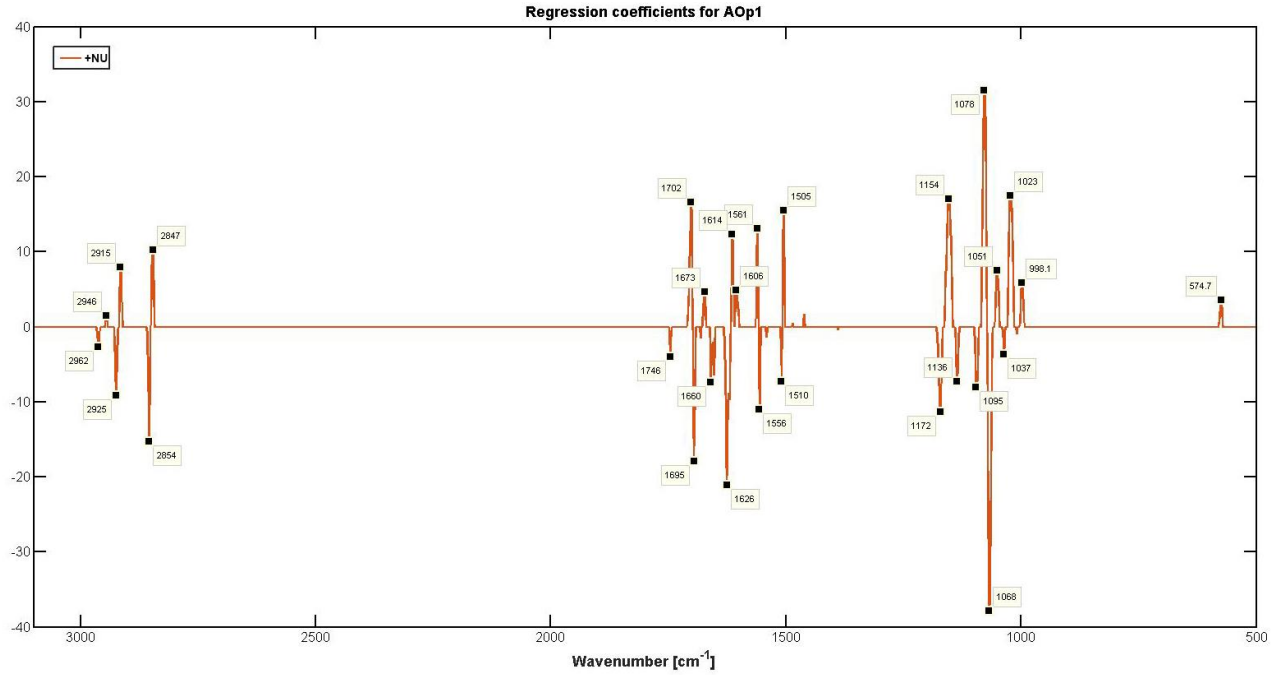


Fig. S9 (Part A) Classification based on growth conditions: The regression coefficient for *Anthoxanthum odoratum*, France, +NU; the regression coefficient for –NU is inverse.

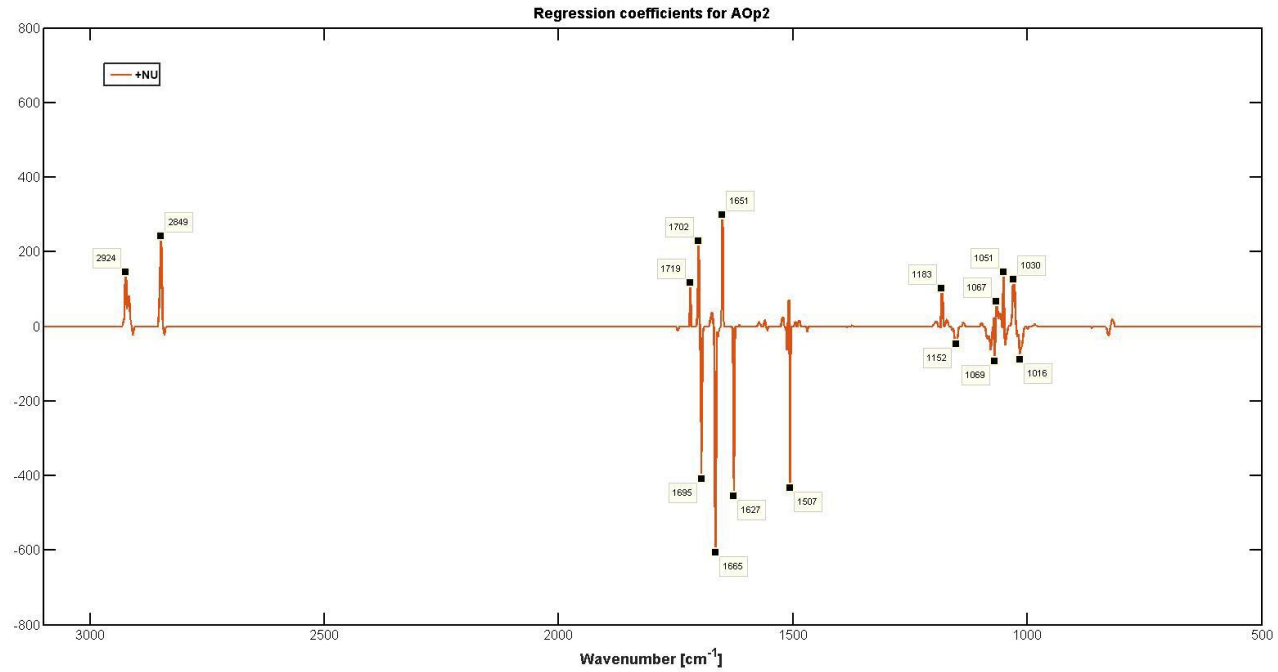


Fig. S9 (Part B) Classification based on growth conditions: The regression coefficient for *Anthoxanthum odoratum*, Greece, +NU; the regression coefficient for –NU is inverse.

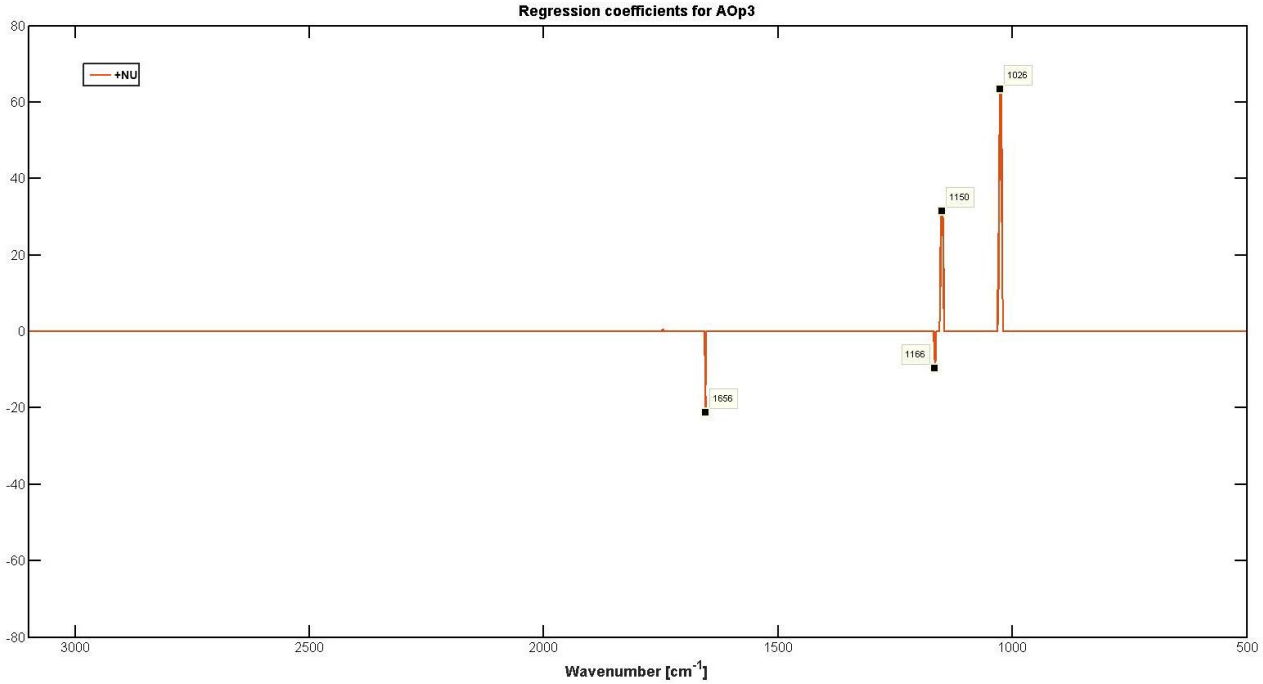


Fig. S9 (Part C) Classification based on growth conditions: The regression coefficient for *Anthoxanthum odoratum*, Finland, +NU; the regression coefficient for –NU is inverse.

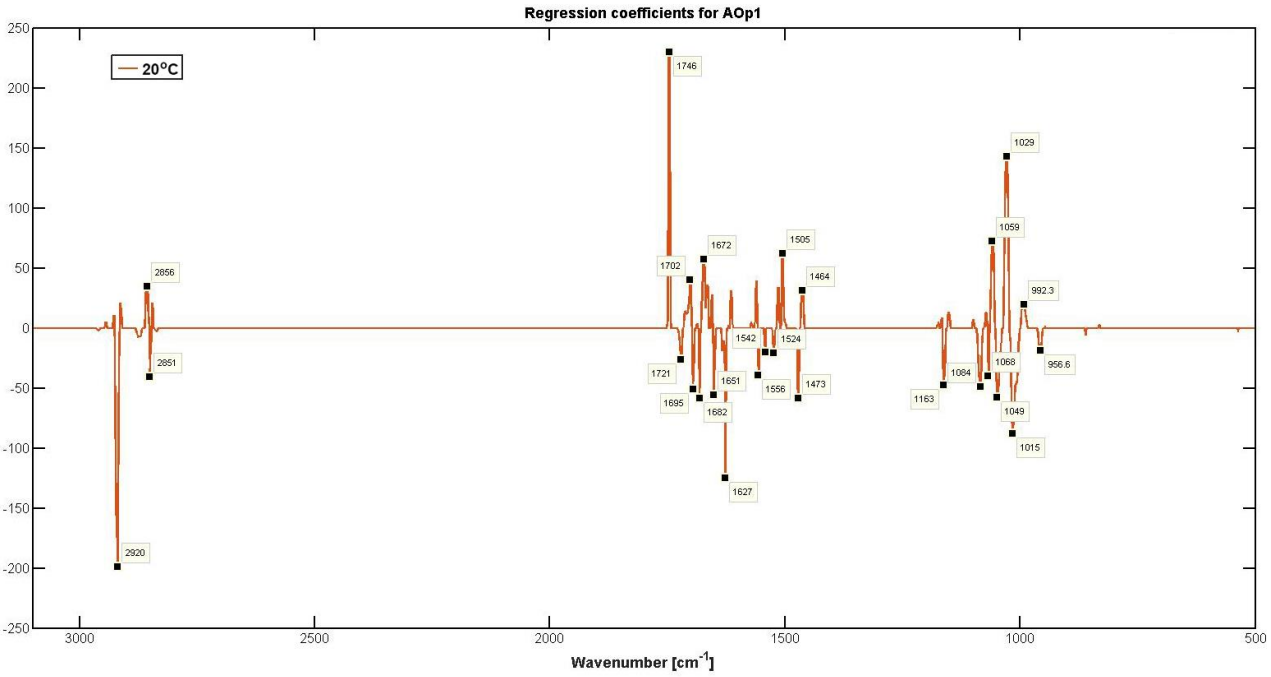


Fig. S9 (Part D) Classification based on growth conditions: The regression coefficient for *Anthoxanthum odoratum*, France, 20°C; the regression coefficient for 14°C is inverse.

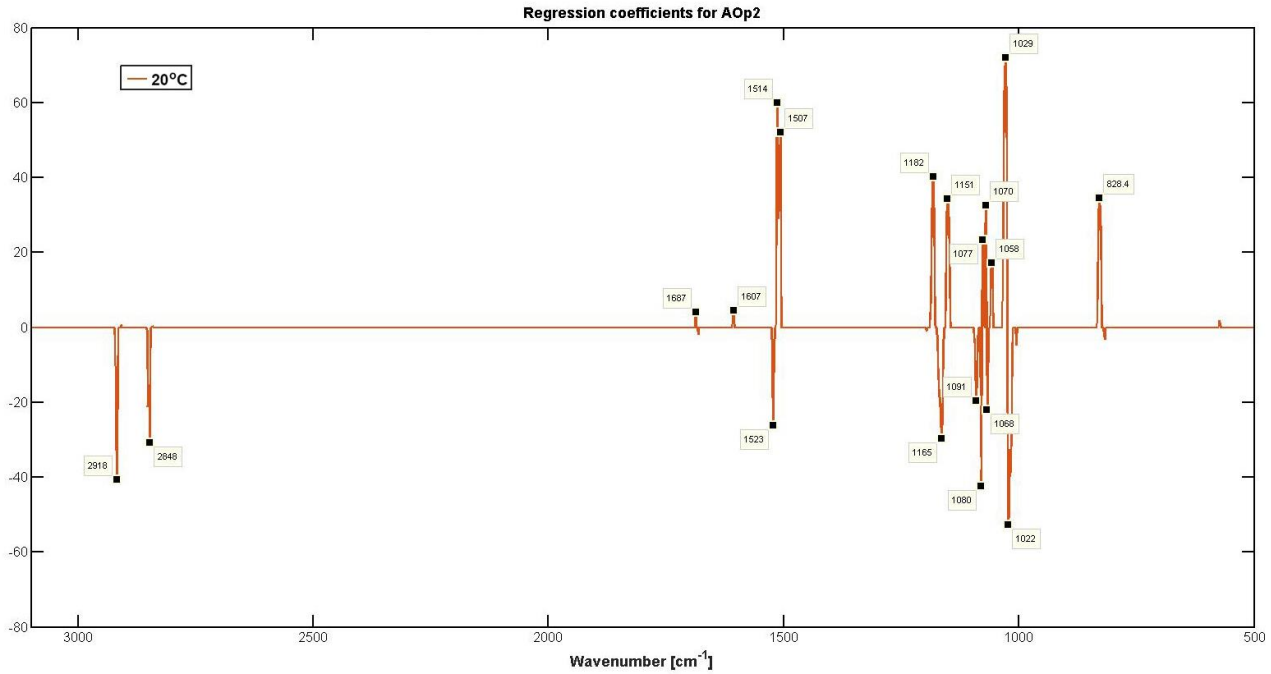


Fig. S9 (Part E) Classification based on growth conditions: The regression coefficient for *Anthoxanthum odoratum*, Greece, 20°C; the regression coefficient for 14°C is inverse.

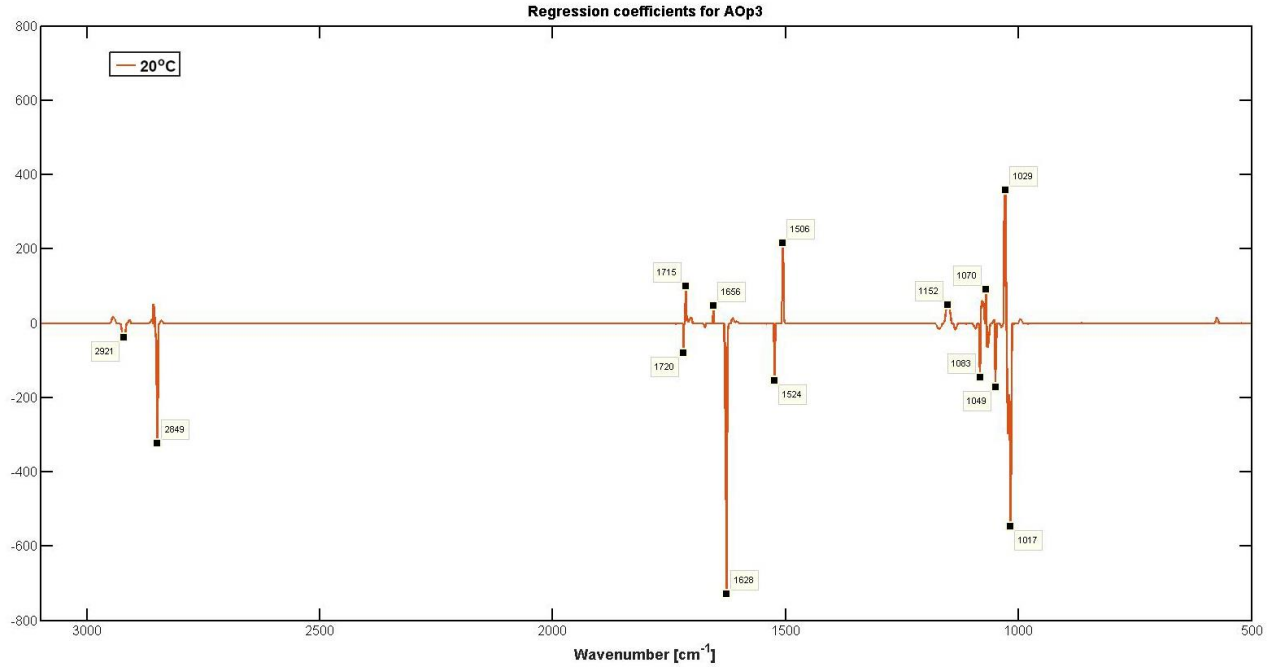


Fig. S9 (Part F) Classification based on growth conditions: The regression coefficient for *Anthoxanthum odoratum*, Finland, 20°C; the regression coefficient for 14°C is inverse.

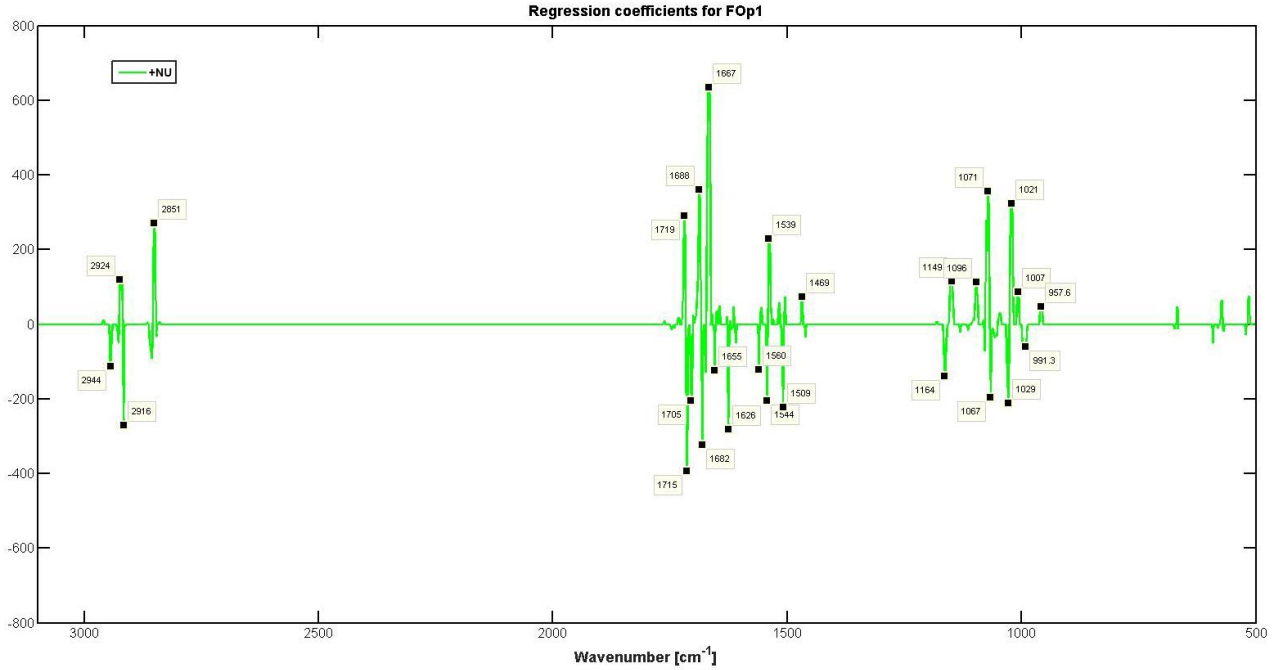


Fig. S10 (Part A) Classification based on growth conditions: The regression coefficient for *Festuca ovina*, Sweden, +NU; the regression coefficient for –NU is inverse.

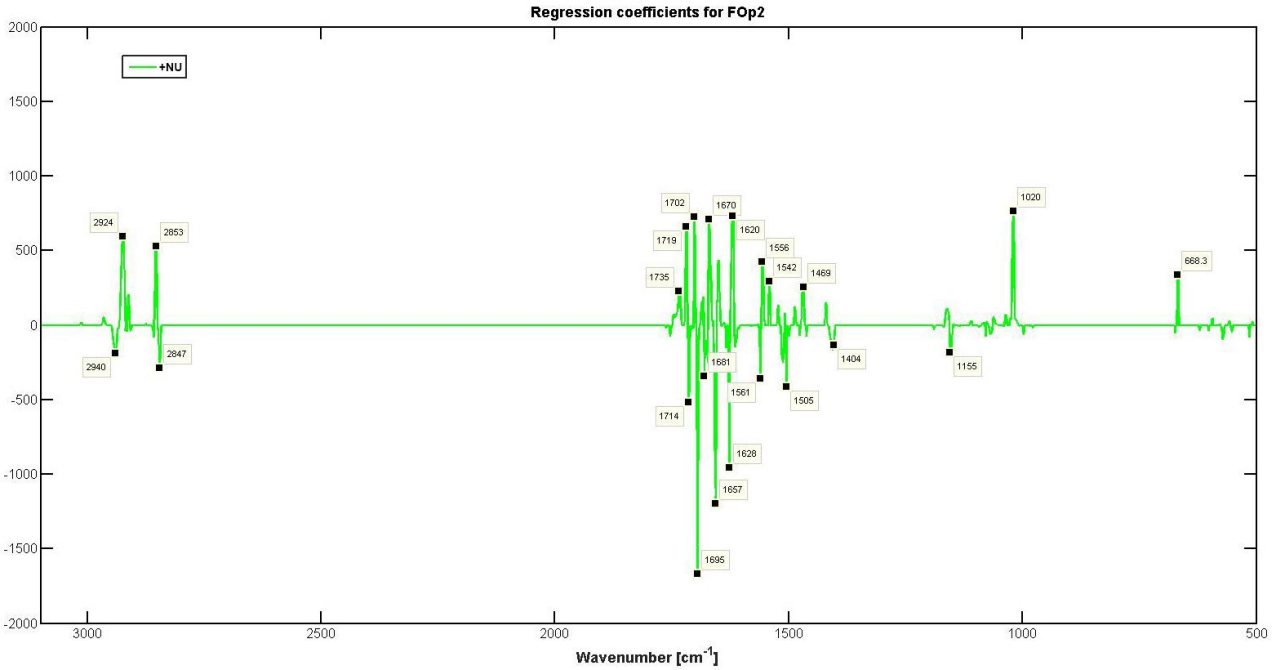


Fig. S10 (Part B) Classification based on growth conditions: The regression coefficient for *Festuca ovina*, Finland, +NU; the regression coefficient for –NU is inverse.

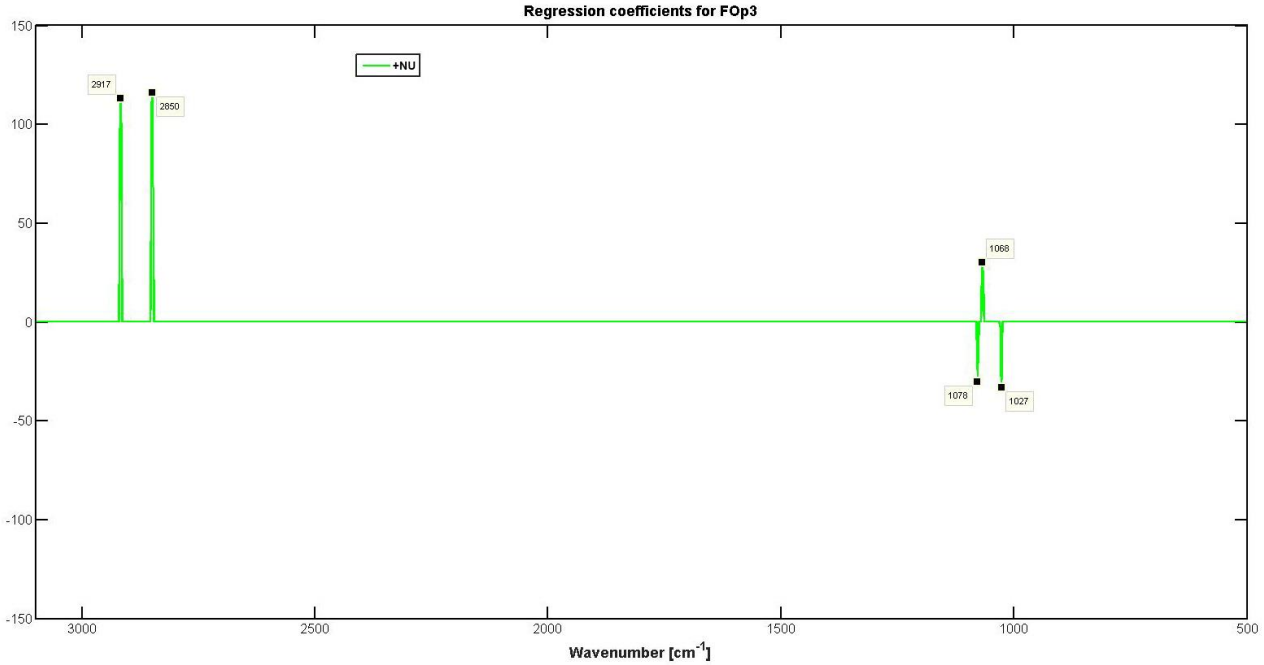


Fig. S10 (Part C) Classification based on growth conditions: The regression coefficient for *Festuca ovina*, Italy, +NU; the regression coefficient for –NU is inverse.

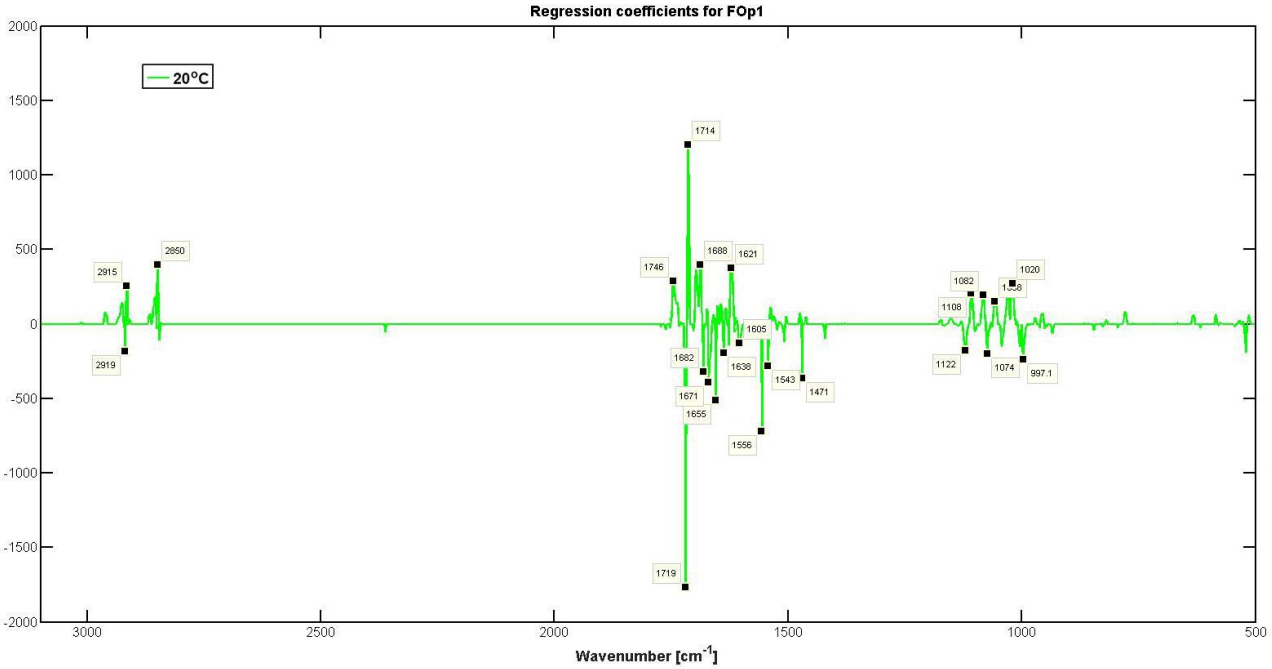


Fig. S10 (Part D) Classification based on growth conditions: The regression coefficient for *Festuca ovina*, Sweden, 20°C; the regression coefficient for 14°C is inverse.

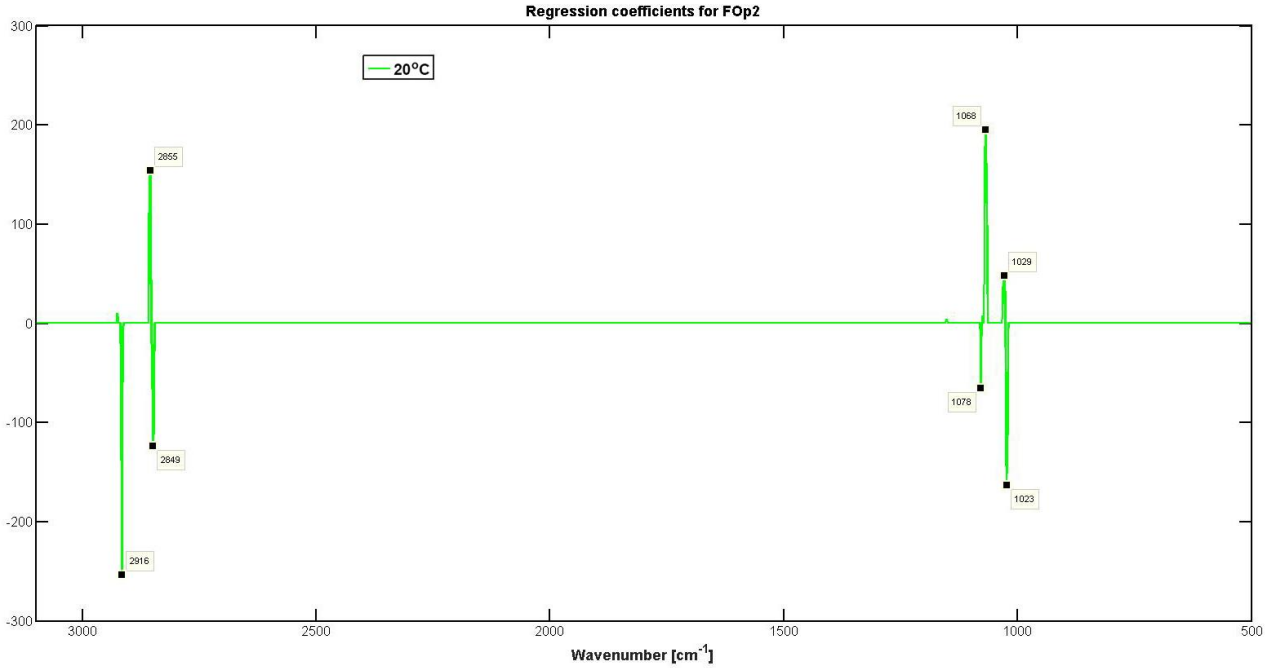


Fig. S10 (Part E) Classification based on growth conditions: The regression coefficient for *Festuca ovina*, Finland, 20°C; the regression coefficient for 14°C is inverse.

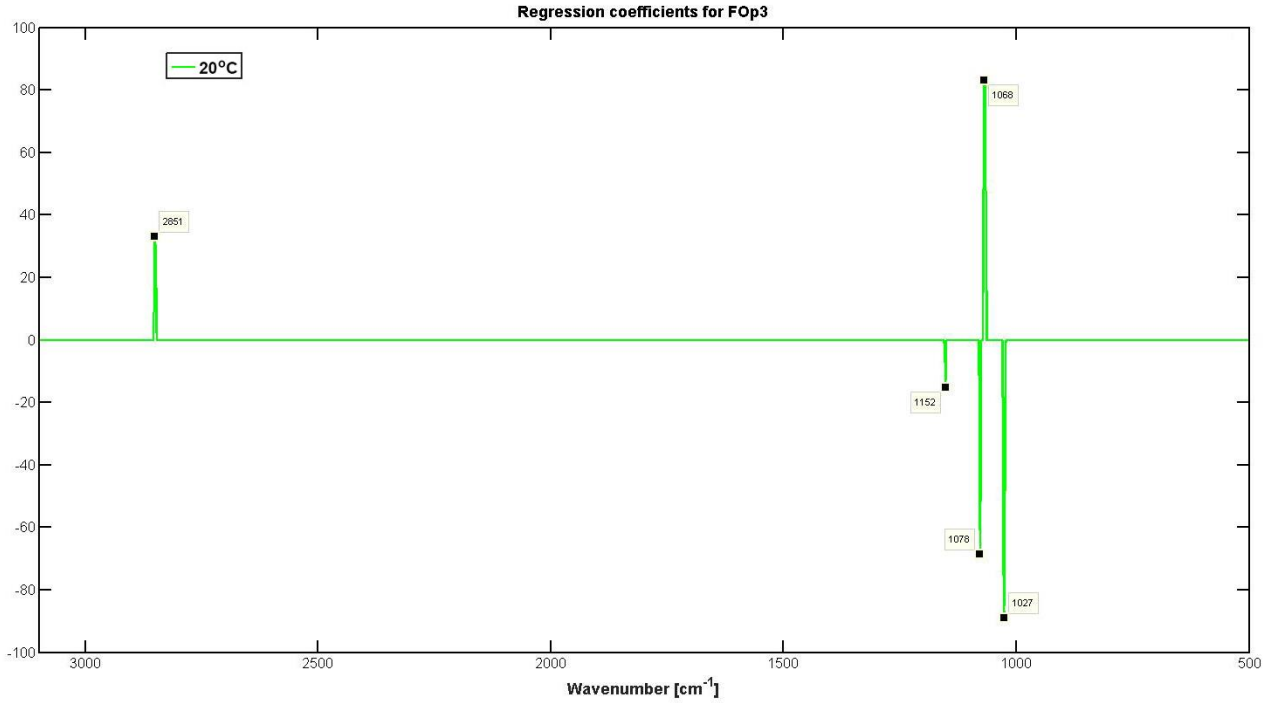


Fig. S10 (Part F) Classification based on growth conditions: The regression coefficient for *Festuca ovina*, Italy, 20°C; the regression coefficient for 14°C is inverse.

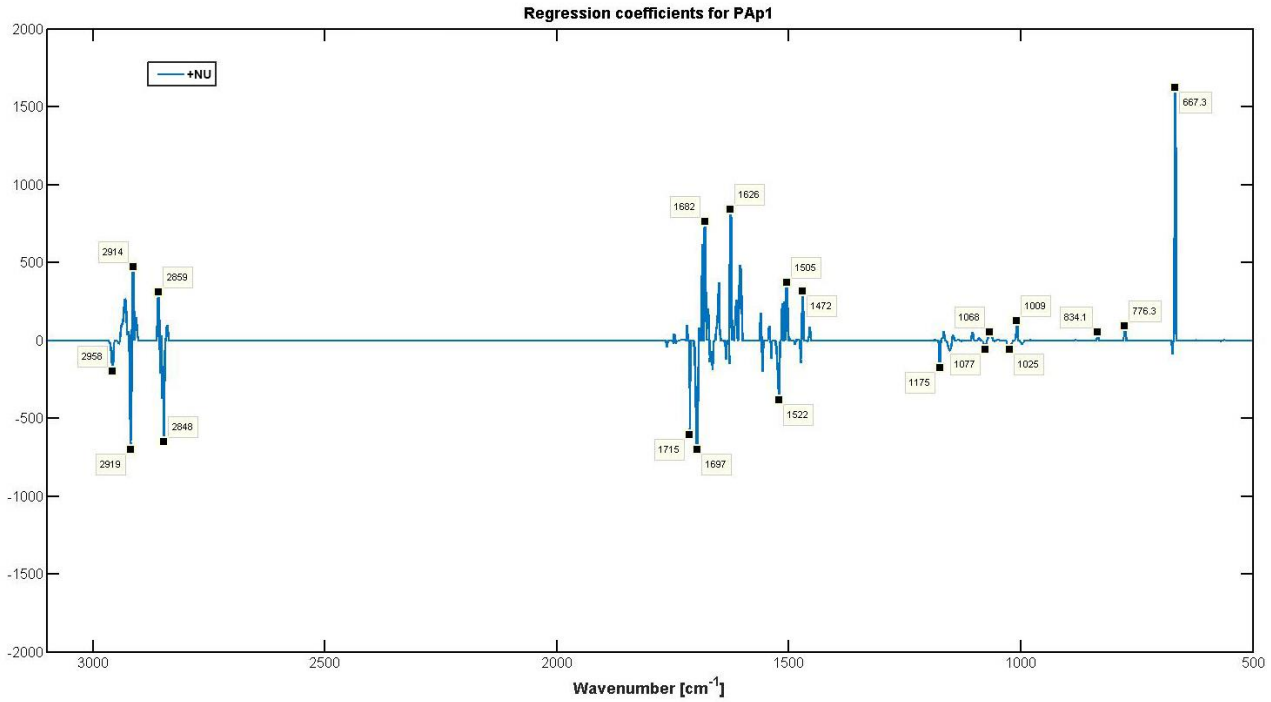


Fig. S11 (Part A) Classification based on growth conditions: The regression coefficient for *Poa alpina*, Sweden, +NU; The regression coefficient for –NU is inverse.

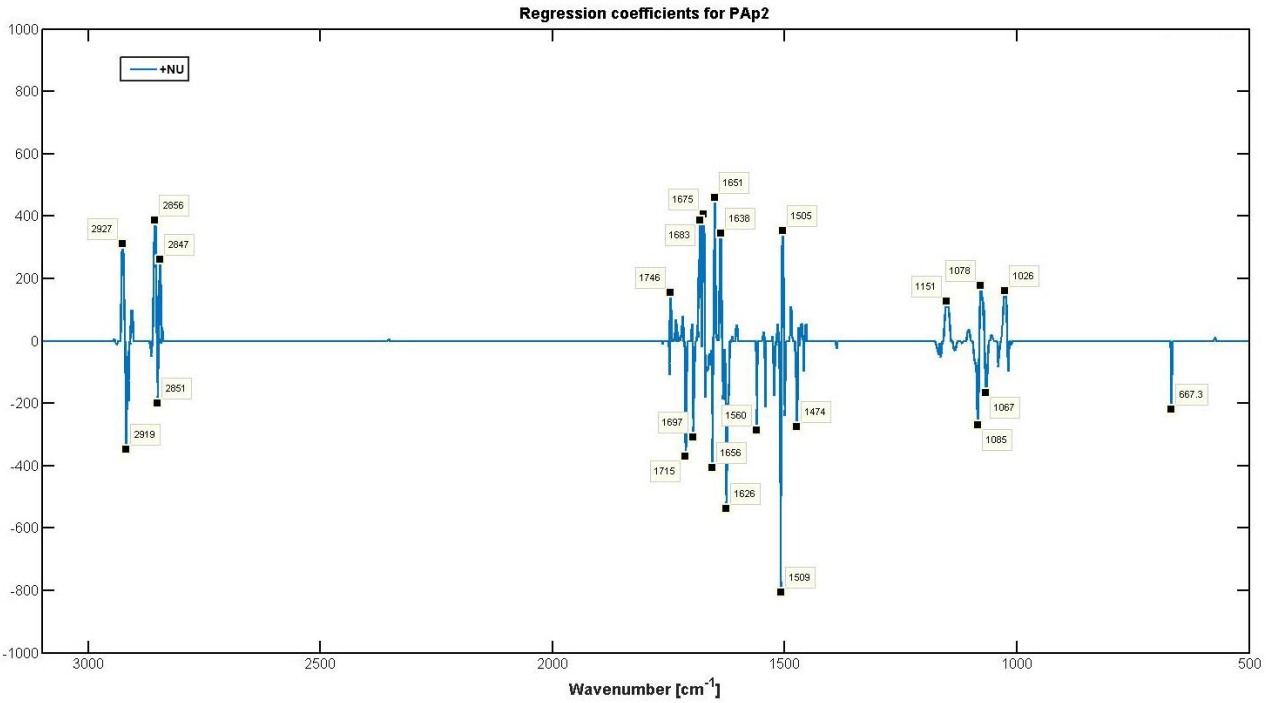


Fig. S11 (Part B) Classification based on growth conditions: The regression coefficient for *Poa alpina*, Italy, +NU; The regression coefficient for –NU is inverse.

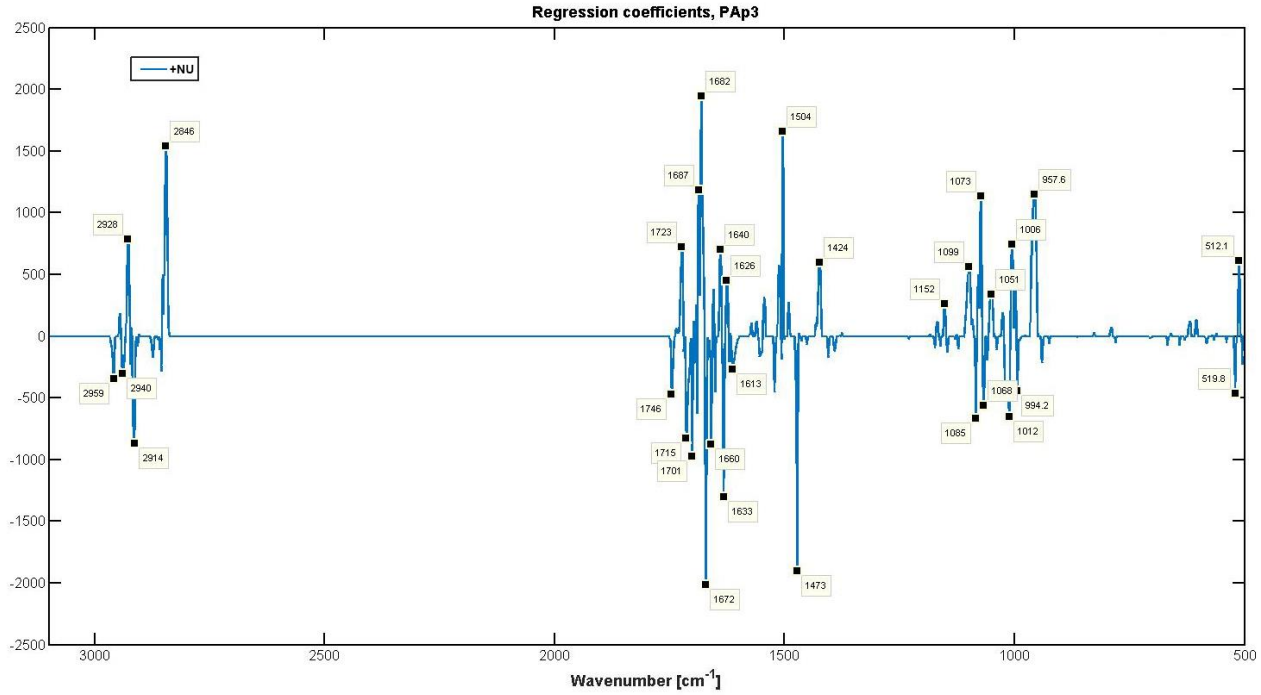


Fig. S11 (Part C) Classification based on growth conditions: The regression coefficient for *Poa alpina*, Norway, +NU; The regression coefficient for –NU is inverse.

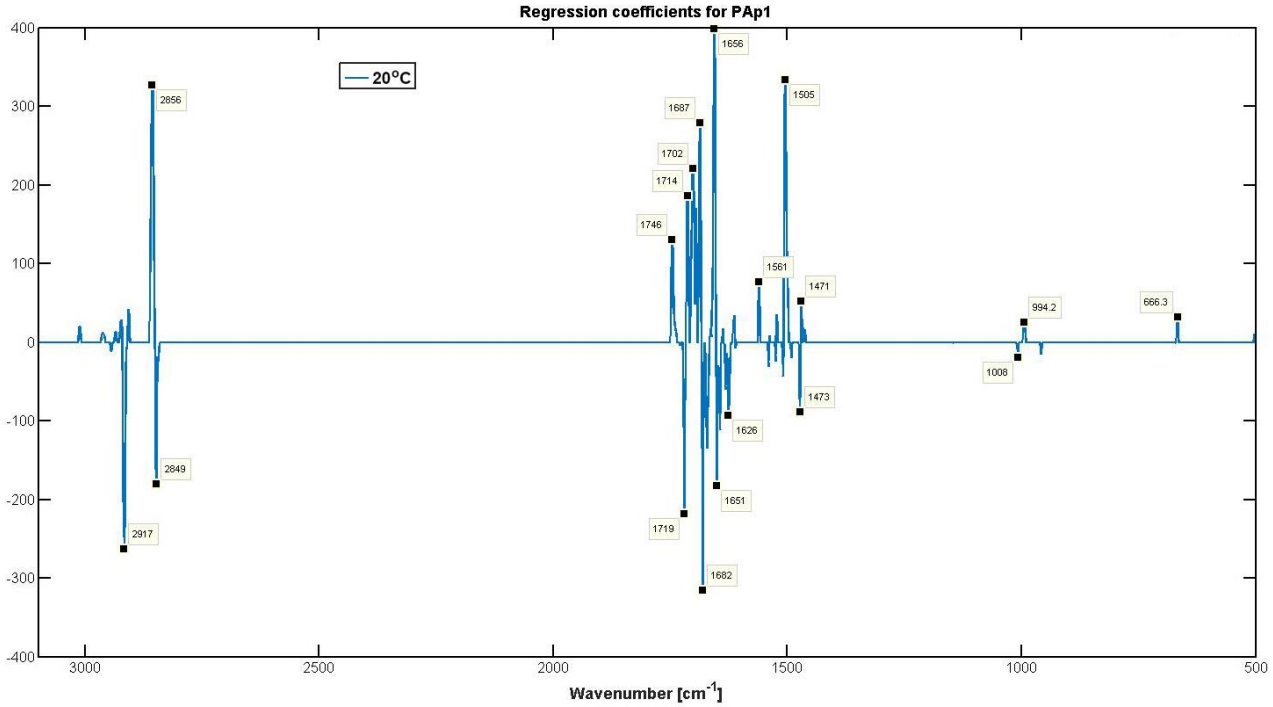


Fig. S11 (Part D) Classification based on growth conditions: The regression coefficient for *Poa alpina*, Sweden, 20°C; The regression coefficient for 14°C is inverse.

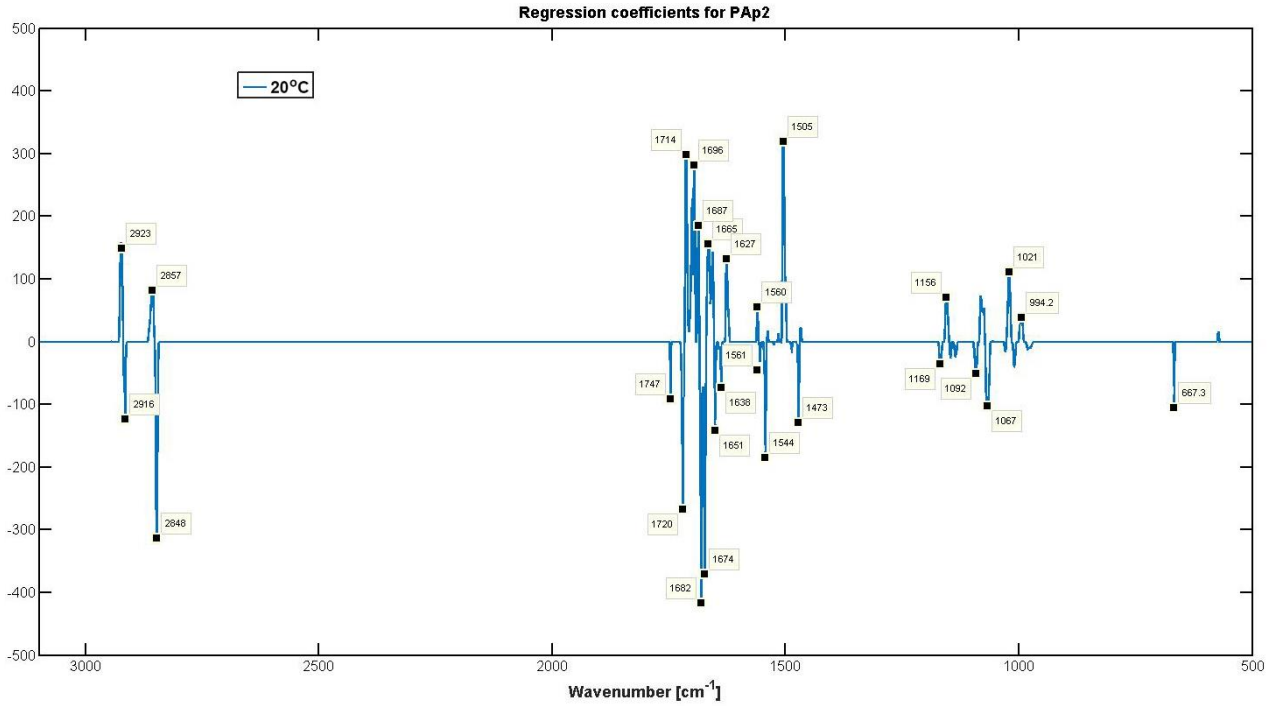


Fig. S11 (Part E) Classification based on growth conditions: The regression coefficient for *Poa alpina*, Italy, 20°C; The regression coefficient for 14°C is inverse.

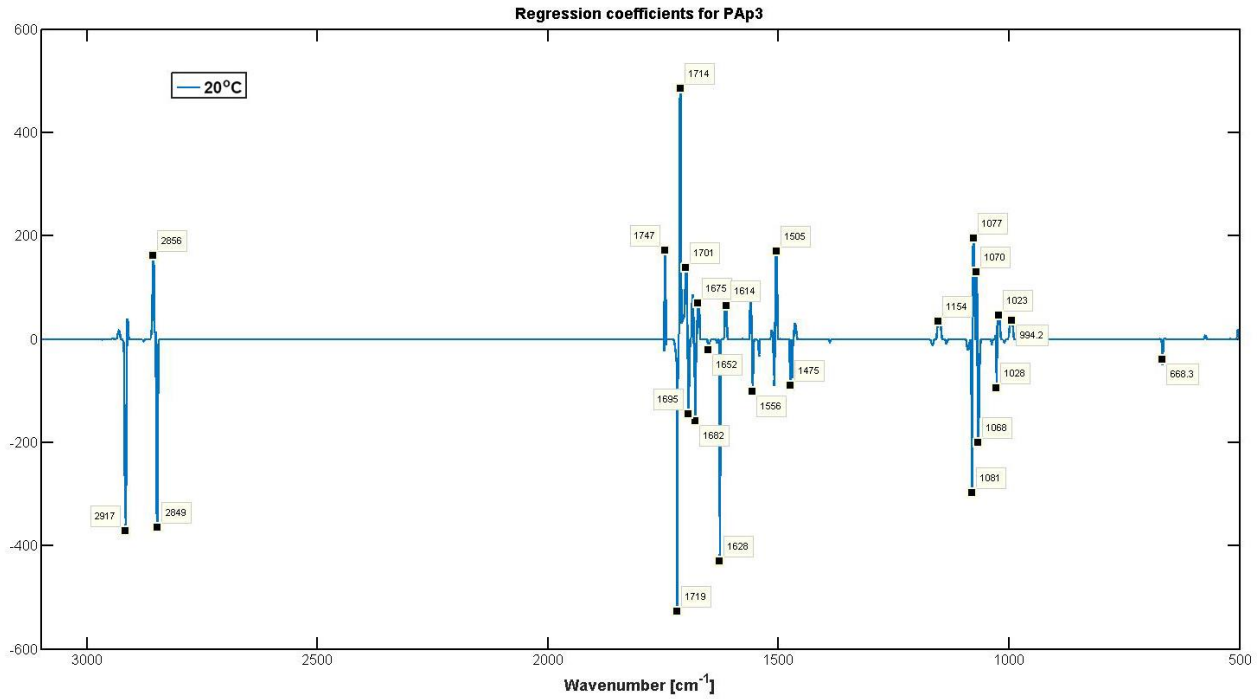


Fig. S11 (Part F) Classification based on growth conditions: The regression coefficient for *Poa alpina*, Norway, 20°C; The regression coefficient for 14°C is inverse.

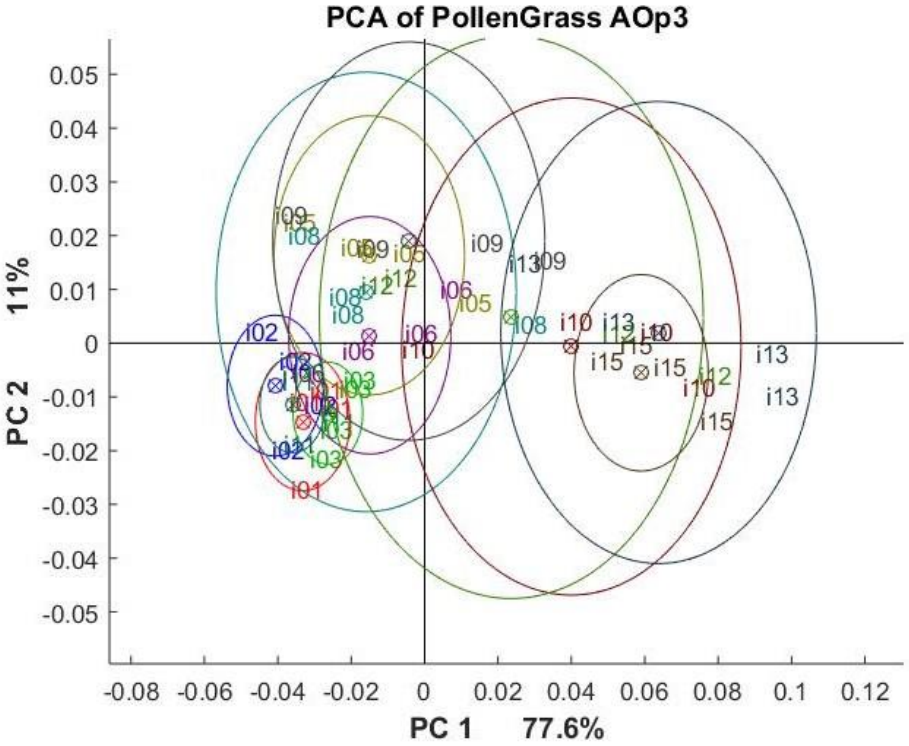


Fig. S12 (Part A) Principal component analysis for *Anthoxanthum odoratum*, Finland, showing clustering based on genotype.

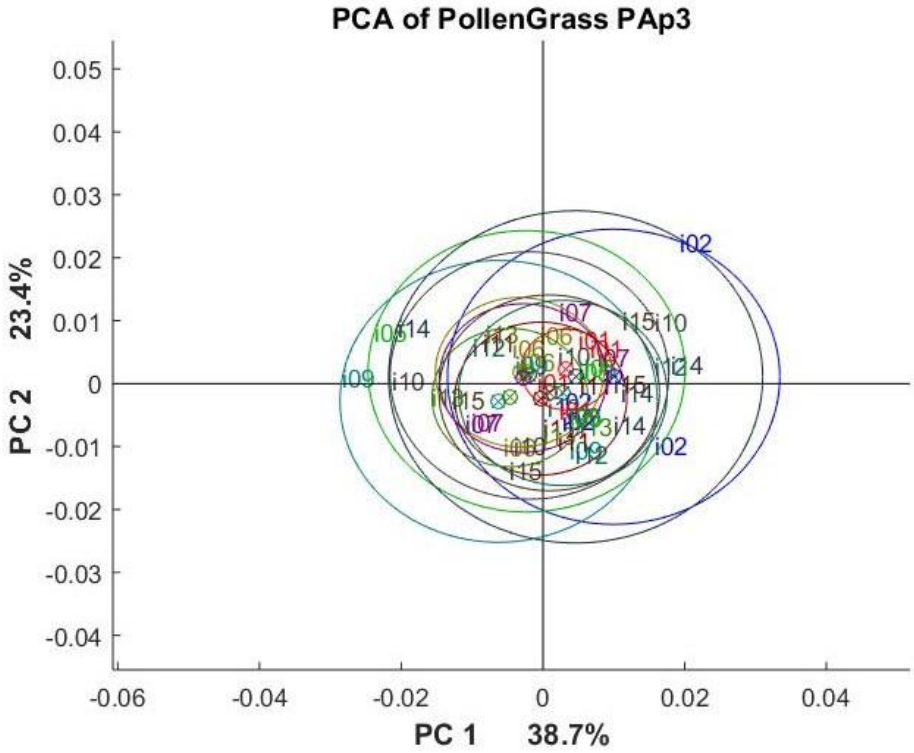


Fig. S12 (Part B) Principal component analysis for *Poa alpina*, Norway, showing clustering based on genotype.