i. Title: Chemical variations in *Quercus* pollen as a tool for taxonomic identification: implications for long-term ecological and biogeographical research

ii. Running title: Chemical variations in *Quercus* pollen

iii. Authors: Florian Muthreich[1]*, Boris Zimmermann[2], H. John B. Birks[1,3,4], Carlos M. Vila-Viçosa[5,6], Alistair W.R Seddon[1,4]

iv. Affiliations: [1] Department of Biological Sciences, University of Bergen, Bergen, Norway. 2 Faculty of Science and Technology, Norwegian University of Life Sciences, Ås, Norway. [3]Bjerknes Centre for Climate Research, Bergen, Norway. [4]Environmental Change Research Centre, University College London, London, UK.[5]CIBIO - Research Center in Biodiversity and Genetic Resources, Vairão, Portugal [6]MHNC-UP - Science and Natural History Museum of Porto University; Praça Gomes Teixeira; 4099-002 Porto; Portugal

*corresponding author

v. Acknowledgements: We dedicate this paper to the memory of John Flenley (1936-2018) who pioneered so many exciting and novel aspects of pollen analysis and vegetation history. One of us (HJBB) knew John since 1963 and was always stimulated by John's latest ideas and palaeoecological studies. John would have been fascinated by the potential of pollen chemistry in palaeoecology and biogeography.

This work is part of the PollChem project funded with a FRIPRO Grant from the Research Council of Norway (PollChem 214359); the extensive field work was possible through funding from the L. Meltzer Høyskolefond (2017/05/LMH) and the Olaf Grolle Olsen Legat (2017/52/FOL). HJBB is supported by EU H2020 grant 741413 HOPE Humans on Planet Earth.

vi. Abstract

Aim: Fossil pollen is an important tool for understanding biogeographic patterns in the past, but the taxonomic resolution of the fossil-pollen record may be limited to genus or even family level. Chemical analysis of pollen grains has the potential to increase the taxonomic resolution of pollen, but present-day chemical variability is poorly understood. This study aims to investigate whether a phylogenetic signal is present in the chemical variations of *Quercus* L. pollen and to assess the prospects of chemical techniques for identification in biogeographic research.

Location: Portugal

Taxon: Six taxa (five species, one subspecies) of *Quercus* L., *Q. faginea*, *Q. robur*, *Q. robur* ssp. *estremadurensis*, *Q. coccifera*, *Q. rotundifolia* and *Q. suber* belonging to three sections: *Cerris*, *Ilex*, and *Quercus* (Denk, Grimm, Manos, Deng, & Hipp, 2017)

Methods: We collected pollen samples from 297 individual *Quercus* trees across a 4° (~450 km) latitudinal gradient and determined chemical differences using Fourier-transform infrared spectroscopy (FTIR). We used canonical powered partial least-squares regression (CPPLS) and discriminant analysis to describe within- and between-species chemical variability.

Results: We find clear differences in the FTIR spectra from *Quercus* pollen at the section level (*Cerris*: ~98%; *Ilex*: ~100%; *Quercus*: ~97%). Successful discrimination is based on spectral signals related to lipids and sporopollenins. However, discrimination of species within individual *Quercus* sections is more difficult: overall, species recall is ~76% and species misidentifications within sections lie between 18% and 31% of the test-set.

Main Conclusions: Our results demonstrate that subgenus level differentiation of *Quercus* pollen is possible using FTIR methods, with successful classification at the section level. This indicates that operator-independent FTIR approaches can surpass traditional morphological techniques using the light microscope. Our results have implications both for providing new insights into past colonisation pathways of *Quercus*, and likewise for forecasting future responses to climate change. However, before FTIR techniques can be applied more broadly across palaeoecology and biogeography, our results also highlight a number of research challenges that still need to be addressed, including developing sporopollenin-specific taxonomic discriminators and determining a more complete understanding of the effects of environmental variation on pollen-chemical signatures in *Quercus*.

Keywords: Chemical variation, Ecology and Environmental sciences, Fourier transform infrared spectroscopy, Palynology, Pollen, Portugal, *Quercus*

vii. Main text

1. Introduction

Sub-fossil pollen remains preserved in lake sediments or peat bogs have been important tools to reconstruct past floristic, vegetational, and environmental changes for over one hundred years. The biogeographic applications of such reconstructions are varied and wide-ranging. Palaeoecological studies based on fossil pollen have made vital contributions to understanding the broad-scale range dynamics through time, the rates and directions of spread of different plant species, and the location of glacial-stage refugia (see Birks, 2019 for a review). Fossil pollen data can also be used to track relative-niche shifts in association with the emergence of no-analogue climates (Veloz et al., 2012) and forecast future range shifts as a result of climate change (e.g. Nogués-Bravo et al., 2016, 2018).

The basis of all such studies is reliable identifications of fossil pollen to the lowest taxonomic level possible. With detailed identifications, reconstructions and answers to particular biogeographical and ecological questions can similarly be detailed. Indeed, many advances in historical plant geography (e.g. Godwin, 1975; Lang, 1994; Magri et al., 2006; Birks, 2008; Birks, 2014) have been made because of advances in the identification of plant fossils. However, although Quaternary botany (sensu Birks, 2019) has been dominated for over one hundred years by pollen analysis, identifications can only be made to the genus or family level for many taxa. This is limiting the biogeographical information gained from fossil pollen to relatively coarse taxonomic levels.

This issue is of particular relevance for understanding the past, present and future distributions of the genus *Quercus* (oak). *Quercus* contains 22 native species in two subgenera and three sections in Europe (Tutin et al., 1993; Denk et al., 2017), several of which have striking and often distinct geographical distributions today (e.g. Iberia, Balkans, eastern Mediterranean, widespread Mediterranean, Apennine Peninsula, widespread to about 60˚N; Jalas & Suominen, 1976). However, what is known about the history of *Quercus* is almost entirely based on pollen and is thus only at the genus level. Although three pollen-morphological types can, with care, be distinguished by conventional light microscopy (LM) (Beug, 2004) and scanning election microscopy (SEM) (Denk & Grimm, 2009), fossil pollen of *Quercus* is usually determined as either *Quercus* Deciduous or *Quercus* Evergreen types.

This situation has several implications for biogeographic research. Maps of the changing distribution and abundance of oak pollen in the late-glacial and Holocene of Europe (Huntley & Birks, 1983; Simon Brewer et al., 2017) can only confidently be made using the two broad pollen morphotypes (i.e. *Quercus* Deciduous or *Quercus* Evergreen). Palaeo-biomisation methods used to forecast the future responses of Mediterranean ecosystems to climate change have used the same distinction between these two morphotypes (Guiot & Cramer, 2016), whilst a recent attempt to model future responses of *Quercus* responses in Europe using fossil pollen were based on *Quercus* pollen resolved to the genus level (Nogués-Bravo et al., 2016). Since the sensitivity and response of *Quercus* to recent environmental changes is species-specific in Mediterranean ecosystems (Acácio, Dias, Catry, Rocha, & Moreira, 2017), and since *Quercus* macrofossils are very rarely found, any improved understanding of its historical and future biogeography clearly depends on consistent pollen identifications at levels lower than is presently available.

One potential approach lies in the chemical analysis of pollen. Fourier Transform Infrared Spectroscopy (FTIR) is a non-destructive method which is used to infer the chemical composition of a sample based on the fact that different molecular-functional groups have different wavelength-specific absorbances of infrared radiation (Pappas, Tarantilis, Harizanis, & Polissiou, 2003; Ivleva, Niessner, & Panne, 2005; Gottardini, Rossi, Cristofolini, & Benedetti, 2007; Schulte, Lingott, Panne, & Kneipp, 2008; Zimmermann, 2010; Parodi, Dickerson, & Cloud, 2013; Zimmermann & Kohler, 2014; Bağcıoğlu, Zimmermann, & Kohler, 2015). Evidence suggests that the analysis of pollen using FTIR may be a useful tool for differentiating between pollen types extracted from sediment sequences (Julier et al., 2016; Woutersen et al., 2018; Jardine, Gosling, Lomax, Julier, & Fraser, 2019), because pollen-grain chemistry may show biogeographical patterns related to phylogeny and environmental conditions (Depciuch, Kasprzyk, Roga, & Parlinska-Wojtan, 2016; Bağcıoğlu, Kohler, Seifert, Kneipp, & Zimmermann, 2017; Zimmermann et al., 2017). However, although taxonomic differentiation of pollen based on the chemical variations inferred by FTIR shows considerable promise, widespread application of FTIR in biogeography and palaeoecology remains limited. This is because lipids and proteins can be major discriminants of variation in FTIR spectra (Bağcıoğlu et al., 2017; Zimmermann & Kohler, 2014), but these may not be preserved in sediment sequences alongside the sporopollenin-based exines (Zimmermann, Tkalčec, Mešić, & Kohler, 2015). Therefore the relative importance of the different chemical structures (e.g. lipids, proteins, sporopollenins) that are responsible for discrimination between many modern pollen types still needs to be established. Moreover, influence of various abiotic stressors on pollen chemistry can hinder taxonomic differentiation of pollen samples by FTIR (Lahlali et al., 2014; Depciuch et al., 2016; Zimmermann et al., 2017), and this needs to be researched further. In addition, the majority of studies investigating pollen chemistry have used pollen from herbaria, botanical gardens, or university campuses, with a limited number of replicates per location and large number of different species and families. Although this sampling design encourages ease of access, reliable identification, and a broad species range, the number of replicates remains a limiting factor for understanding chemical variations in response to both environmental and taxonomic factors.

This study aims to address the challenges related to understanding the taxonomic variations of pollen chemistry by investigating the relative importance of within- and between-species chemical differences in *Quercus*. Our dataset is unique because it represents the largest collection of closely related species sampled from populations outside botanic gardens across a large bioclimatic and biogeographic gradient in Portugal. We use a combined multivariate- and discriminant- analysis approach to (i) investigate the potential for FTIR as tool to differentiate six taxa of *Quercus* based on pollen, and (ii) determine the main chemical-functional groups responsible for chemical variation observed in the dataset. Addressing these questions represents the first step if we are to successfully use pollen chemistry as a tool to improve our understanding of past biogeographical patterns and history of oaks in Europe.

2. Methods

2.1 Sample collection

We collected pollen samples from 297 individual trees belonging to five *Quercus* species across a 4° (~450 km) latitudinal gradient in Portugal (Figure 1). The *Quercus* species in this study belong to the sections *Cerris*, *Ilex*, and *Quercus* according to Denk et al. (2017) and have different geographic distributions (Table 1). Trees were sampled along gradients of temperature and precipitation to cover a wide range of environmental conditions. A detailed summary of the number of trees sampled at each location is in the Supplementary Material (Table S1).

Table 1 Taxonomy of sampled Quercus *trees and total number of trees sampled (n). Sections are according to Denk et al. (2017)*

Subgenus	Section	Species	Subspecies	n	Distribution
Cerris	*Ilex*	*Q. coccifera*		36	*Q. coccifera* and *Q. rotundifolia* prefer xerophytic conditions and are often co-occurring species. Both are indifferent towards bedrock conditions, but prefer soils without waterlogging, although *Q. coccifera* is a thermophilous species and is less tolerant of winter cold.
Cerris	*Ilex*	*Q. rotundifolia*		38	
Cerris	*Cerris*	*Q. suber*		69	*Q. suber* is distributed across the Mediterranean Basin on siliceous bedrock but is absent from areas with winter cold (Amigo, 2017; Matías, Abdelaziz, Godoy, & Gómez-Aparicio, 2019).
Quercus	*Quercus*	*Q. faginea*		63	*Q. faginea s.l.* has a broad distribution in the Iberian Peninsula (Tschan & Denk, 2012) and is more abundant on limestone with higher summer precipitation, tracking the sub-Mediterranean bioclimatic belt (Sanchez, Benito-Garzon, & Ollero, 2009).
Quercus	*Quercus*	*Q. robur*		76	*Q. robur* has a temperate distribution and is more abundant in north-west Portugal (Jalas & Suominen, 1976). It occurs in regions with sufficient summer rain, and is absent from areas with summer drought (Amigo, 2017; Ülker, Tavsanoglu, & Perktas, 2018).
Quercus	*Quercus*	*Q. robur*	*estremadurensis*	15	

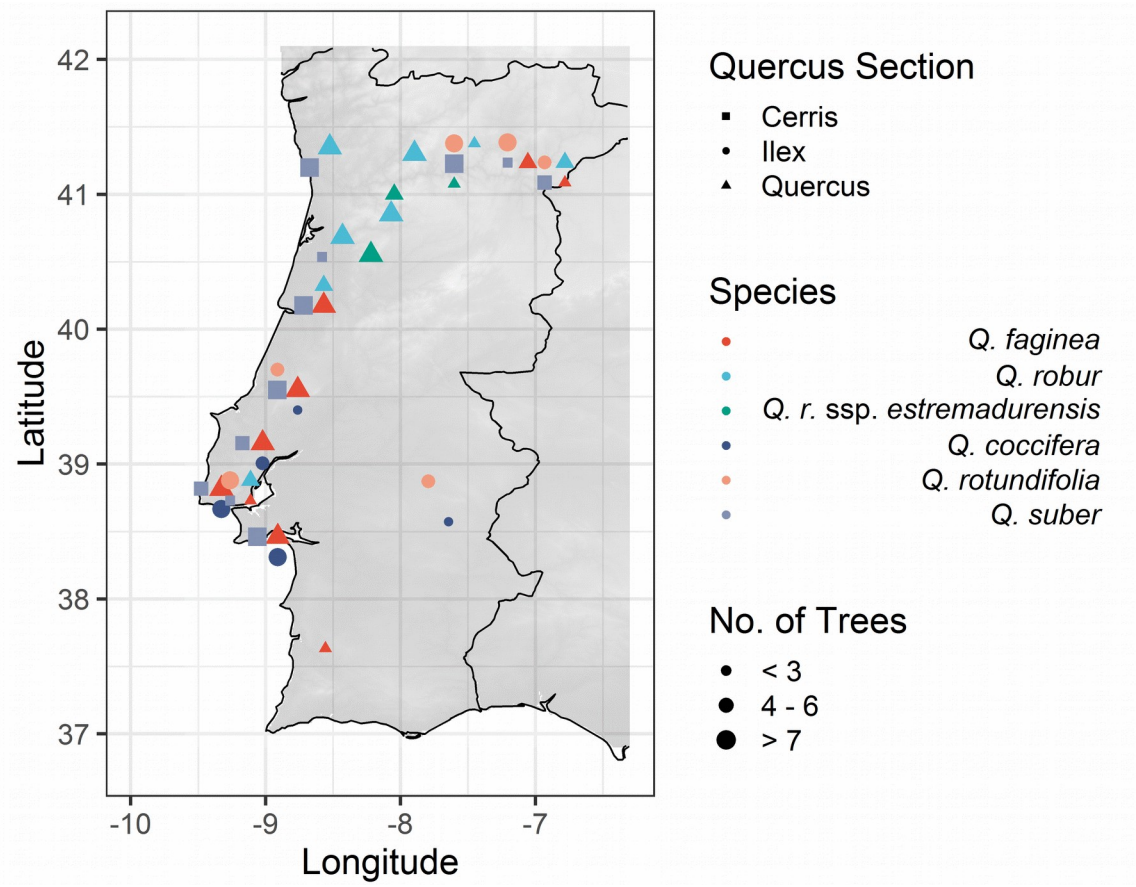

Figure 1. Map of Portugal displaying the locations where trees where sampled. Colour indicates species, shape indicates Quercus section following Denk et al. (2017) and size of the symbol indicates number of trees sampled per location.

All samples were collected in spring 2018 by taking whole-tree composite samples of ca. 30 catkins per individual tree. Several branches were sampled for catkins up to 5 m in height. The catkins were air-dried at room temperature (~23 C) for at least 24h and the pollen was separated from the anthers by light shaking. Pollen was also sieved through 60 µm sieves to remove excess plant material before analysis.

2.2 Pollen-chemistry measurements

Reflectance-infrared spectra were recorded using a Vertex 70 FTIR spectrometer (Bruker Optik GmbH, Germany) with a single reflectance-attenuated total-reflectance (SR-ATR) accessory. The ATR IR spectra were recorded with a total of 32 scans and spectral resolution of 4 cm^{-1} over the range of 4000–600 cm^{-1}, using the horizontal SR-ATR diamond prism with 45° angle of incidence on a High Temperature Golden gate ATR Mk II (Specac, United Kingdom). Approximately 1 mg of dried pollen was deposited onto the ATR crystal for each measurement (3 replicate measurements). Between each measurement a background (reference) spectrum was recorded using the sample-free setup. The OPUS software (Bruker Optik GmbH, Germany) was used for data acquisition and instrument control.

We pre-processed the spectra, since multivariate-regression methods (e.g. partial least squares; PLS) have been shown to perform better with preprocessed spectra in other studies (Zimmermann & Kohler, 2013; Woutersen et al., 2018).The processing of the spectra consisted of smoothing and calculation of

the second-derivative using the Savitzky-Golay algorithm, as implemented by the EMSC package (Liland, 2017). The settings of the Savitzky-Golay smoothing algorithm (Edwards & Willson, 1974; Savitzky & Golay, 1964) were: 2nd degree polynomial and a window-size of 11. The second-derivative spectra were constrained between 700 and 1900 cm^{-1} and normalized using extended multiplicative signal correction (EMSC), a MSC model extended by a linear and quadratic component (Liland, 2017). For further analyses, the mean of the measurement replicates (three) was calculated for each tree (resulting in one spectra per tree). We follow peaks of interest that have been attributed to chemical-functional groups according to Pappas et al. (2003), Gottardini et al. (2007), Schulte et al. (2008), Zimmermann (2010) and Zimmermann & Kohler (2014) (Table 2).

Table 2 Wavenumber of peaks attributed to specific functional groups in spectra of fresh pollen and the compounds most representative for them (Pappas et al., 2003; Gottardini et al., 2007; Schulte et al., 2008; Zimmermann, 2010, p. 20109; Zimmermann & Kohler, 2014 compiled in Zimmermann, Bağcıoğlu, et al., 2015). A (*) marks wavenumbers which are shared by more than one compound (Bağcıoğlu et al., 2015).

Compounds	**Wavenumber (cm^{-1})**	**Functional group**
Lipids	1745; 1175*	C=O stretch
Triglycerides &	1462	CH_2 deformation
	1170	C-O-C stretch
Phospholipids	722; 955*	CH_2 rocking
Proteins	1640; 1650	Amide I: C=O stretch
	1535; 1550	Amide II: NH deformation and C-N stretch
Carbohydrates	1200-900	C–O–C and C–OH stretch
Cellulose	1107; 1055; 1028	
Amylose	1076; 995*	
Sporopollenin	1605; 1515; 1175*; 852 833 and 816	Aromatic rings in phenylpropanoid subunits

2.3 Statistical analyses

For the exploration of within- and between-species chemical variability, we fitted a PLS model combined with canonical correlation analysis (CPPLS) to the processed mean spectra (2nd derivative) to predict species identity. This analysis was implemented in the “pls” package (Mevik, Wehrens, Liland, & Hiemstra, 2019) in R version 3.6.0 (R Core Team, 2019). The PLS family of models has been shown to be powerful in multivariate analyses of FTIR-spectral data (Wold, Sjöström, & Eriksson, 2001; Liland, Mevik, Rukke, Almøy, & Isaksson, 2009;

Telaar, Nürnberg, & Repsilber, 2010; Zimmermann et al., 2017). The CPPLS method improves the extraction of predictive information by estimating optimal latent variables in comparison to standard PLS regression (Mehmood & Ahmed, 2016). Unlike standard PLS, CPPLS weights the contribution of the explanatory variables (wavenumbers), which weakens the contribution of non-relevant wavenumbers to optimise the covariance between response (species) and explanatory variables (wavenumbers). Indahl, Liland, & Næs (2009) show improved accuracy and increased explained variance of CPPLS compared to conventional PLSR using spectral data.

To assess the classification performance of the CPPLS, the dataset was randomly split into training and test sets using a 60%/40% split. This split was repeated 100 times to create 100 versions of the dataset (folds) with different training/test splits. A CPPLS model was fitted for each fold and the extracted component scores were used to predict species identity using limited discriminant analyses. The performance of the classifier in predicting the test set was averaged over the folds and summarised in a confusion matrix (Table 3).

3. Results

3.1 Chemical variations in *Quercus*

Assessment of the mean spectra of the five *Quercus* species and one subspecies reveals clear differences in absorbance between the major intrageneric lineages (sections) at wavelengths associated with specific chemical functional groups (Figure 2). For example, the lipid peak absorbance at ~1750 cm^{-1} is weaker in section *Ilex* compared to the other sections, whilst the sporopollenin and carbohydrate absorbance bands (at 1516, 1168, 833 cm^{-1} and 985 cm^{-1}, respectively) are noticeably lower in absorbance in the taxa belonging to section *Ilex*. Note, however, that although clear spectral differences exist at the section level, it is more difficult to separate variations between species *within* different sections (Figure 2).

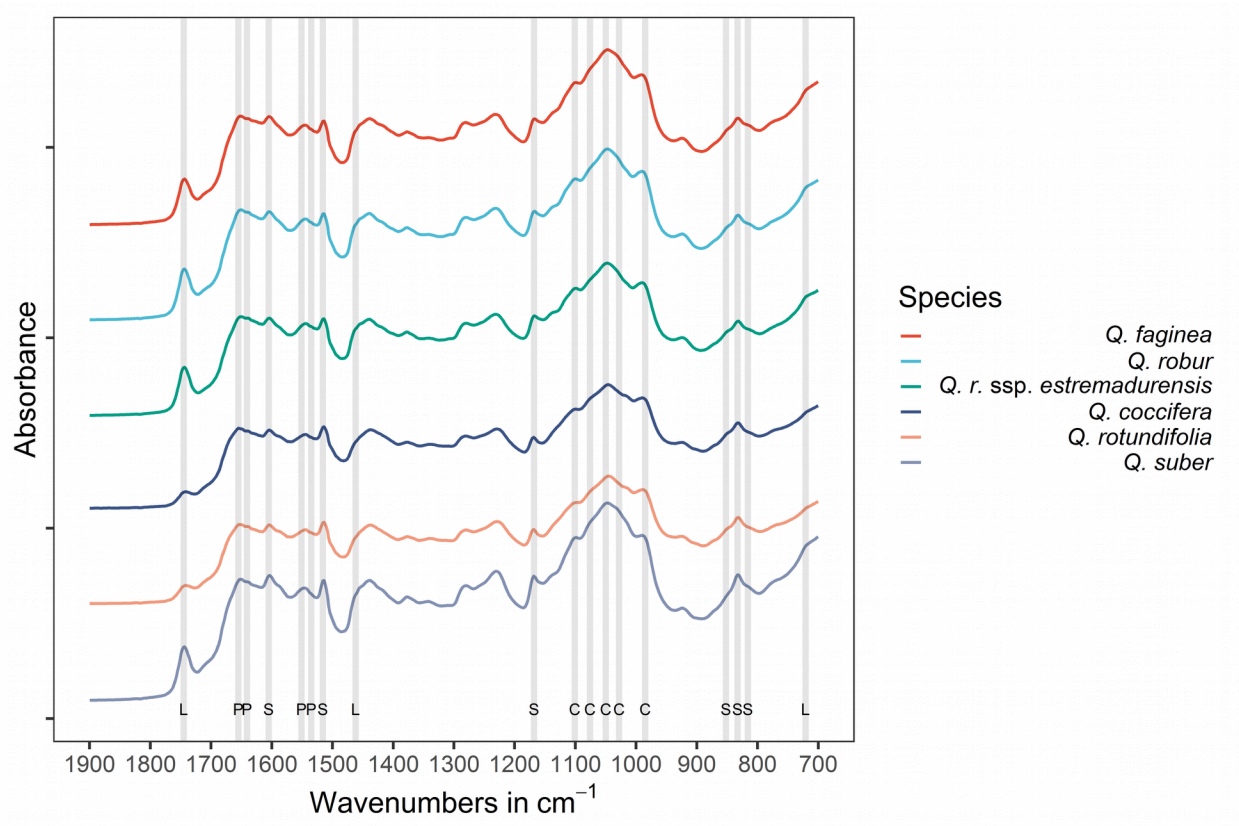

Figure 2 Mean absorbance spectra of Quercus *species and notable peak locations. Lipids (L), protein (P), sporopollenin (S), and carbohydrates (C) using wavenumbers given in Table 2. Spectra are offset.*

The observations made following assessment of the mean spectra are confirmed by the analysis using CPPLS (Figure 3a). Here, the three sections of *Quercus* can be clearly separated using the variance explained by two CPPLS components (Figure 3a). For example, the section *Ilex* scores are negative along the first component (26 % of the variation), whilst individuals of sections *Quercus* and *Cerris* have positive scores along this axis. The sections *Quercus* and *Cerris* can mainly be separated along the second principal component (section *Cerris* with positive scores and section *Quercus* with negative scores). However, between-species level variations within species in the same section are harder to differentiate, because there is a large overlap between species of the sections *Ilex* and *Quercus*. This overlap is reduced on later components, where the different species separate within their respective sections. On components 3 and

4 *Q. coccifera* and *Q. rotundifolia* can be separated along the third component, while *Q. robur* and *Q. faginea* show some separation along the fourth component (Figure 3b). In total, the two first components explain ~37% of the variation in the dataset and separate the samples into the *Quercus* sections, while the next two components explain a further ~9.7% (Figure S1).

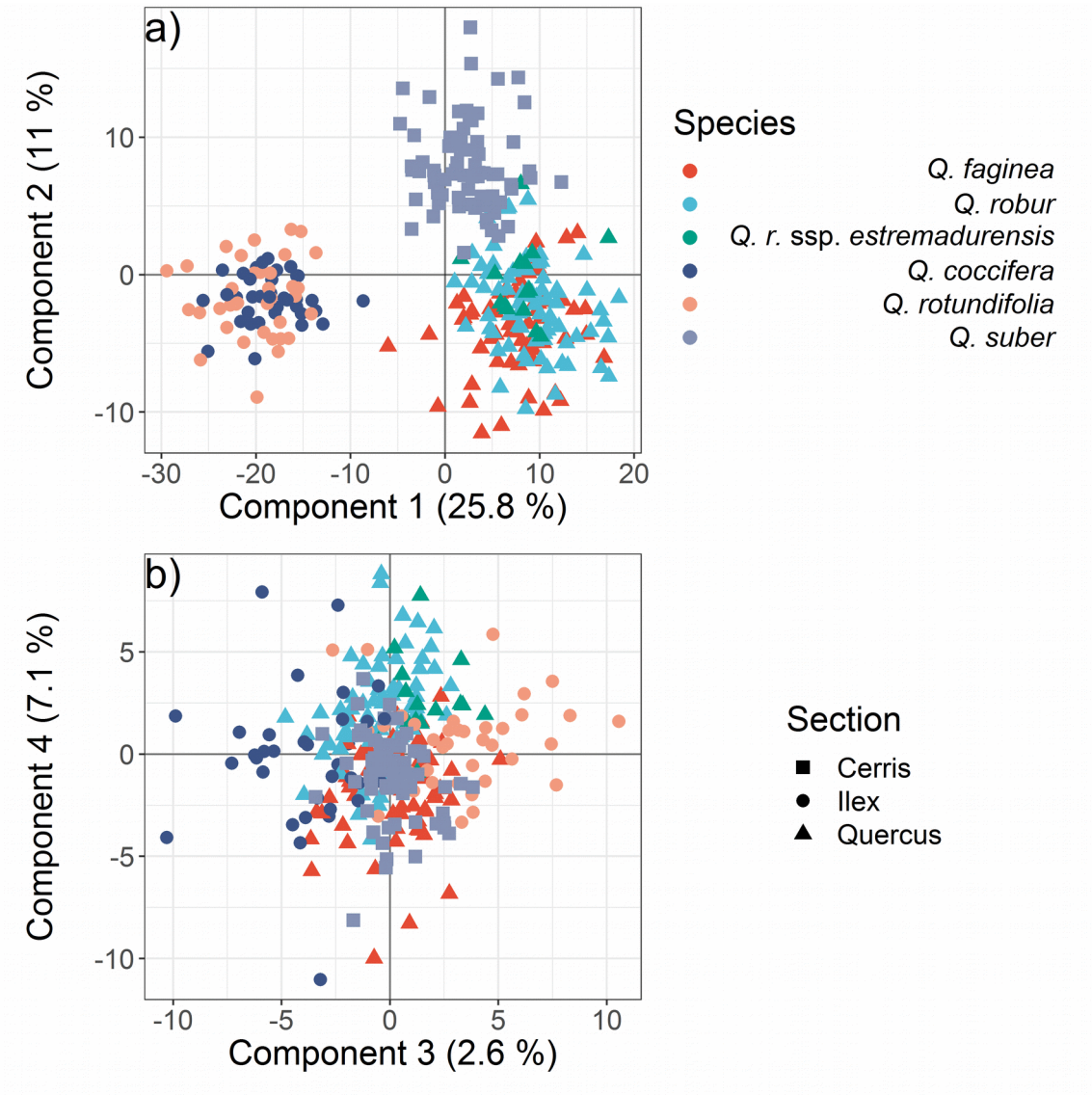

Figure 3 Ordination of components 1-4 of canonical powered partial least squares (CPPLS) models (100-fold cross-validation). Each point represents a sampled tree. The mean training scores for each sample over the 100-folds were calculated. Proportion of variance explained by each component in parentheses. Colour indicates species, while symbol indicates Quercus *section.*

Since specific absorbance peaks in the FTIR spectra can be related to chemical functional groups (summarised in Table 2), it is possible to identify which chemical functional groups have the highest influence in explaining patterns of variation observed in our dataset (Figure 4). In general, lipids have the highest loadings (i.e. greatest influence) on component 1. The peak at 995 cm^{-1} in pollen can also be attributed to phospholipids (Bağcıoğlu et al., 2015) in addition to carbohydrates. Sporopollenin peaks have high positive loadings on component 1 and mostly negative loadings on component 2. The 852 cm^{-1} sporopollenin peak shows high positive loadings on component 2 in contrast to the other sporopollenin peaks (816, 833, 1175, 1516, and 1605 cm^{-1}). Components 3 and 4 have higher loadings for carbohydrates and protein peaks. Taken together, these results indicate that sporopollenin and lipids are strong drivers of the main sources of variation in the dataset and congeneric discrimination is achievable

based on FTIR at least at section level.

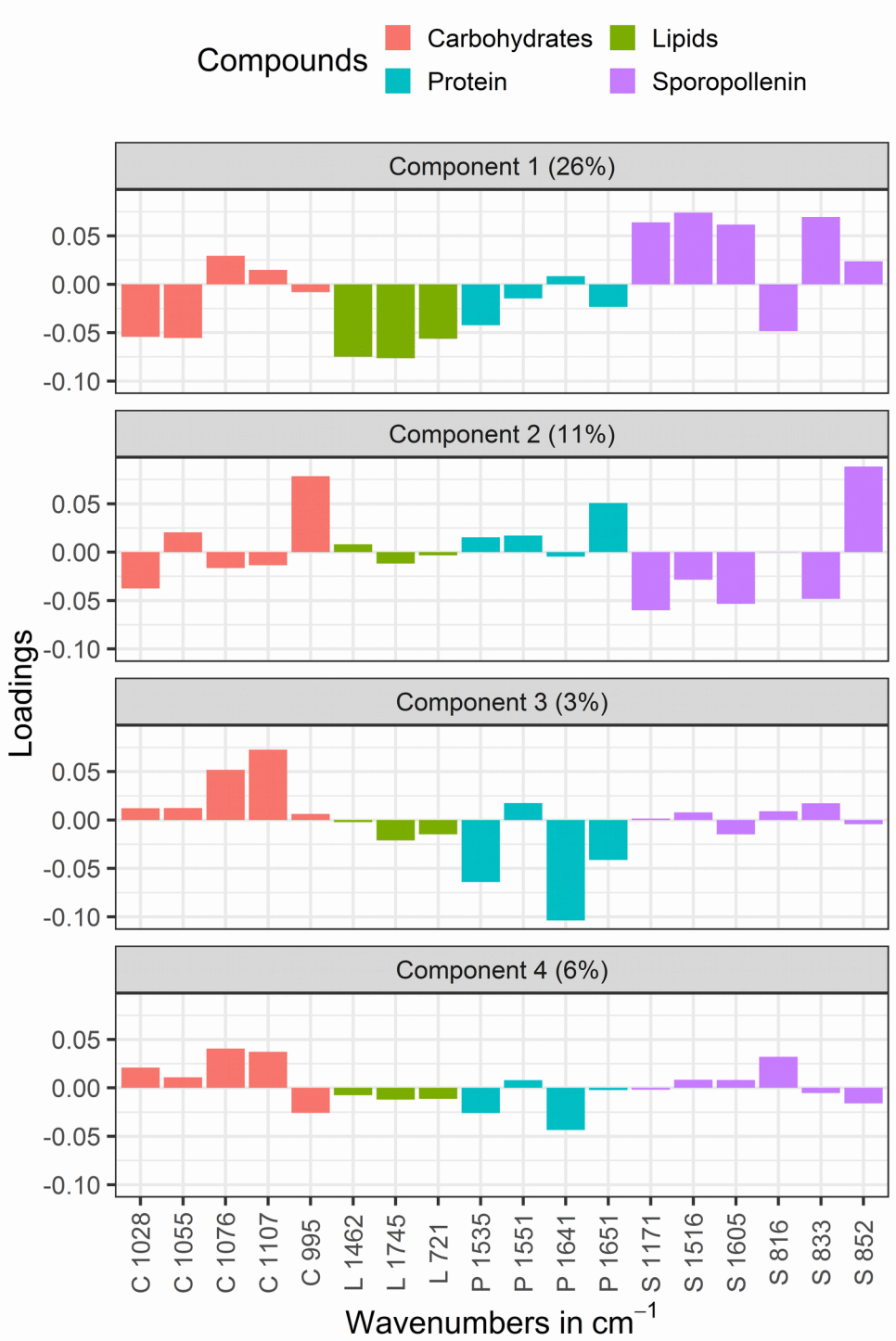

Figure 4 Loadings plot of classification of canonical powered partial least squares (CPPLS) model. Lipids (L), protein (P), sporopollenin (S), and carbohydrates (C) using wavenumbers given in Table 2. High absolute loading indicates a high importance of a given wavenumber for the corresponding component. Loadings are chosen in such a way as to describe as much as possible of the covariance between the variables (wavenumbers) and the response (species). Proportion of variance explained by each component in parentheses.

3.2 Discriminant analysis

Using four components (explaining ~45% of the variance) the confusion matrix of the classification CPLS model shows clear differentiation between the *Quercus* sections, with some misidentified spectra (<~2 ±3) from section *Quercus* and *Q. suber* (Table 3). *Quercus robur* ssp. *estremadurensis* has by far the worst accuracy in the model and is most often identified as its parent species *Quercus robur*, possibly due to the limited number of samples in the dataset (Table 1). In general, species misidentifications are contained within the different *Quercus* sections and lie between 18% and 31% of the test-set samples (within-section misidentifications: 18% of *Q. robur* as *Q. faginea*; 31% of *Q. faginea* as *Q. robur*; 25% of *Q. coccifera* as *Q. rotundifolia*; 25% of *Q. rotundifolia* as *Q. coccifera*). Overall species accuracy within sections ranges from 65% to 76% in the Ilex and Quercus sections. Increasing the components available to the model to 10 (~57% explained variance) increases species accuracy by 5-10 percentage points in both sections (*Quercus*, *Ilex*) (Table S2). As demonstrated with our ordination plots (Figure 3), differentiation of the three sections of *Quercus* is possible using 37% of the variance in the spectral data, but these results indicate difficulties in differentiation between species of the same section.

Table 3 Confusion matrix of linear discriminant analysis on the test sets using 4 components of the fitted canonical powered partial least squares (CPPLS) model. Predictions as rows and Reference as columns. Values given as % of spectra that were predicted as species and sum to 100% columnwise, e.g. for Q. robur 18% of spectra were predicted as Q. faginea, 76% were correctly classified as Q. robur and 5% as estremadurensis. Abbreviations; Q. r. estr: Q. robur ssp. *estremadurensis, Q. rotund.: Q. rotundifolia. Green signifies section Quercus, blue is section Ilex and red is section Cerris.*

Pred/Ref	***Q. faginea***	***Q. robur***	***Q. r. estr.***	***Q. coccifera***	***Q. rotund.***	***Q. suber***
Q. faginea	65 ±12	18 ±8	11 ±13	0	0	1 ±1
Q. robur	31 ±12	76 ±10	81 ±18	0	0	1 ±2
Q. r. estr.	1 ±2	5 ±6	6 ±13	0	0	0
Q. coccifera	2 ±3	0	0	75 ±14	25 ±11	0
Q. rotund.	1 ±2	0	0	25 ±14	75 ±11	0
Q. suber	3 ±4	1 ±3	2 ± 6	0	0	99 ±2

4. DISCUSSION

4.1 Separation of *Quercus* according to chemical variation

Recent research has shown that spectroscopic methods such as FTIR are effective at differentiating pollen species between distantly related families and/or genera using their chemical composition (Gottardini et al., 2007; Dell'Anna et al., 2009; Zimmermann, 2010; Zimmermann & Kohler, 2014; Zimmermann, Bağcıoğlu, et al., 2015; Zimmermann, Tafintseva, Bağcıoğlu, Berdahl, & Kohler, 2016; Julier et al., 2016; Zimmermann et al., 2017; Woutersen et al., 2018; Jardine et al., 2019). Our results build on these previous studies to reveal the

potential for chemical variations in pollen to distinguish infrageneric variation between species in pollen samples from 297 individuals from Portugal belonging to three different *Quercus* sections (*Cerris*, *Ilex*, *Quercus*). We identify a clear separation at the *Quercus* section level (Fig. 3 and Table 3), using two components of a PLS model and 37% of the explained variance in the spectral data. One component (component 1) can be used to differentiate the sections *Ilex* and *Quercus*, while component 2 can be used to separate section *Cerris*. Combined, these two components achieve the performance equivalent to SEM methods, where *Quercus* pollen can be confidently identified to section level.

Despite finding that classification at the section level is possible using FTIR approaches, there is considerable overlap in variation between species of the same section. Furthermore, classification performance does not improve when using a more complex model in which the number of components used increases from four to ten. Using this more complex model, which explains ~57% of the variance (compared to ~45% in the four-component model), classification accuracies remain roughly similar within *Quercus* sections (Table S2). For example, *Q. coccifera* and *Q. rotundifolia* have a recall of ~75% with both 4 and 10 components. Similarly, approximately one third of *Q. robur* and *Q. faginea* samples (both belonging to section *Quercus*) are misclassified as the other species. Thus, while our results indicate that subgeneric classification of *Quercus* pollen is possible at the section level using FTIR, we still find it difficult to distinguish between more closely related (i.e. within-section) pollen types.

These findings are approximately in line with other studies, which have performed species classification using FTIR. For example, both Julier et al. (2016) and Jardine et al. (2019) report classification successes of ~80% and ~85% rates, respectively, using an FTIR analysis of cryptic morphospecies within the family Poaceae. Their studies are based on a combination of specimens of mainly non-congeneric grass species (except two species of *Oryza,* Julier et al., 2016, and four species of *Triticum*, Jardine et al. 2019). In both these studies, classification success is lower for the samples belonging to congeneric species and higher for the more-distantly related pollen types (i.e. those species belonging to different genera). In another study, Woutersen et al. (2018) report ~95% recall on largely congeneric species in the Nitrariaceae family using single grain FTIR, but also note that lack of environmental variability (pollen from one individual per species) could have led to an overestimation of classification success. In contrast, Zimmermann et al. (2017) achieve ~100% accuracy on species identification and 75% accuracy on identification of origin using hierarchical PLSR on pollen from three species of Poaceae (*Festuca ovina*, *Anthoxanthum odoratum, Poa alpina*) of different genera and origins (Sweden, Norway, Finland) grown under controlled conditions (45 individuals per species). Such a high classification success on taxa grown in controlled conditions demonstrates the strong phylogenetic signature that can be observed using FTIR. Our results also demonstrate these strong phylogenetic differences in FTIR spectra (i.e. the ability to differentiate between *Quercus*-section level variability), but our study additionally demonstrates the difficulty of distinguishing between-species level variability, even when relatively large subsets of samples are used.

4.2 Key chemical drivers of variation within *Quercus* spp. pollen

Given the result that identification is possible at the section level, a key question that follows is which of the chemical components of the pollen grain are mostly responsible for the difference between the three sections of *Quercus* under FTIR? In our study, we find that lipids are one of the most important functional groups in diagnosing samples belonging to section *Ilex*. The waveband

at 1750 cm^{-1} is particularly important in this regard. Previous research has shown that this waveband is a strong indicator for triglyceride lipids (Bağcıoğlu et al., 2017). Our results also confirm previous findings by Zimmermann & Kohler (2014), who show extreme variations in the relative content of triglycerides and find this waveband to be an important separator between *Iris*, *Quercus,* and *Pinus* pollen types. Indeed, our results extend the inferences made in that previous study by demonstrating a subgeneric level variability of the relative lipid content. Specifically, we identify a distinctly lower amount of lipids in pollen sampled from individuals within the Ilex section (Figure 2).

In addition to the importance of triglyceride lipids as a tool for chemical separation, we also find that wavebands representing building blocks of sporopollenin (Table 3) are important for differentiating taxa on the first two components of our CPPLS analysis. For example, the peaks at 833, 852, 1516 and 1605 cm^{-1} are associated with building blocks of sporopollenin (Bağcıoğlu et al., 2015) and have the highest loadings on component 2, which is used in this study to isolate *Q. suber*. Peaks at 833 and 852 cm^{-1} are related to different types of phenylpropanoid building blocks and our results suggest relative differences in their abundance within the sporopollenin of *Q. suber* compared to the other species. In addition to lipid variation, aromatic peaks at 833, 852, 1516 and 1605 cm^{-1} also have a high loading on component 1, which can be used to separate the *Ilex* section pollen from the other sections. Thus, both lipids and sporopollenins are important functional groups to different pollen between the three sections in our dataset.

Our observations of different chemical compositions of sporopollenin mirror the sequence of the development of the pollen wall in different *Quercus* sections described by Solomon, (1983a, 1983b) and Denk & Grimm, (2009). Evolutionary, pollen of section *Ilex* represent the earliest, primitive state of *Quercus* pollen, with a microrugulate pattern on the pollen exine surface. A set of secondary sporopollenins are then added to this surface during exine formation in pollen of sections *Cerris* and *Quercus* (Denk & Grimm, 2009). It is possible that these key differences in structure and formation of the exine between the section *Ilex* and other *Quercus* pollen grains are responsible for the key differences sporopollenin chemistry identified using FTIR. More detailed work on the composition of sporopollenin of different genera and how this affects pollen grain structural elements (e.g. Li, Phyo, Jacobowitz, Hong, & Weng, 2019) is needed for this finding to be confirmed.

Finally, protein and carbohydrate peaks (Carbohydrates: 1107, 1028, 1076 cm^{-1}; Proteins: 1535, 1641 cm^{-1}) have the highest loadings on components 3 and 4 and are partly responsible for the partial distinction of species within the same section. These peaks represent amylose and cellulose as carbohydrates and amide functional groups within proteins (Table 2). For example, variation along component 3 contributes to the separation of *Q. robur* from *Q. faginea*. However, overall, protein and carbohydrate peaks have the least influence for explaining the variance of classification success, and most of the taxonomically important information we used to distinguish between the species is stored in the lipids and sporopollenin components of the pollen chemistry.

4.3 Implications for understanding past and future *Quercus* dynamics

We investigated the potential for chemical separation of *Quercus* pollen because, despite the high diversity (22 species) of *Quercus* in Europe, fossil pollen of this genus is still most commonly determined as either *Quercus* Deciduous or *Quercus* Evergreen types (e.g. Huntley & Birks, 1983; Brewer et al.,

2017). As a result, there may be much detail missing in our current understanding about past *Quercus* dynamics which could be improved through methods that result in refined taxonomic resolution. Indeed, our results indicate the potential for FTIR to surpass traditional LM methods used in palynology, and work at a comparable level to SEM (Denk & Grimm, 2009; Denk & Tekleva, 2014; Grímsson et al., 2015; Grímsson, Grimm, Meller, Bouchal, & Zetter, 2016). However, the extensive automated-classification possibilities offered by future IR analysis (Mondol et al., 2019), and in the ease of sample preparation and data collection, may mean it will be easier to expand these technologies compared to the more time consuming SEM methods in the long-term.

The ability to differentiate at higher taxonomic resolution would enhance our understanding of past trajectories of co-occurring *Quercus* sections, in particular for understanding the expansion of *Quercus* since the Last Glacial Maximum (e.g. S. Brewer, Cheddadi, de Beaulieu, & Reille, 2002). FTIR techniques may also be useful for older interglacial sequences, where identification of *Quercus* pollen to section is often not possible due to degradation (Tzedakis, 1994). This would complement studies that use genetic methods to reconstruct colonization pathways, which have higher taxonomic resolution and compliment the palynological data, but lack the temporal resolution that pollen records provide (Petit et al., 2002).

In addition, a number of studies have highlighted the need to incorporate long-term ecological information to improve biodiversity forecasts of environmental change (Dawson, Jackson, House, Prentice, & Mace, 2011). Rates of temperature increases in the Mediterranean are projected to outpace the rest of the temperate regions in Europe and are predicted to rapidly change the associated biomes in the region (Giorgi & Lionello, 2008; Guiot & Kaniewski, 2015; Guiot & Cramer, 2016), but the consequences of this change for Mediterranean oak forests remain uncertain (Lindner et al., 2014; Acácio et al., 2017). A number of studies have integrated pollen data in order to reduce uncertainties when forecasting the biotic responses of *Quercus* to climate change in the future (Nogues Bravo et al. 2016; Guiot & Cramer, 2016). However, like the palaeoecological studies discussed above, the limited taxonomic resolution used may bias projections. For example, species distribution models based solely on *Quercus* pollen, were only able to estimate niche-environment relationships at the genus level. Extensive application of FTIR techniques may therefore provide a bridge between long-term ecological and modern biogeographic approaches.

Nevertheless, despite the potential shown in our FTIR approach, our findings still reveal a number of challenges before vibrational methods can be rolled out across biogeographic and palaeoecological applications. First, our results are still unable to resolve at the species level, and so although taxonomic resolution would be refined using FTIR approaches, in many cases palaeoecological studies would still lack the taxonomic precision of other biogeographic tools (e.g. phylogenetic analysis). Second, our results are based on fresh pollen sampled from modern taxa, and lipids were the main functional groups used to differentiate between taxa in this study, followed by sporopollenins (Figure 4). Although the preservation and stability of sporopollenins in fossil sequences are well established (Fraser et al., 2012), the extent to which lipids are preserved in chemical sequences in subfossil pollen sequences remains uncertain. Variations in sporopollenin functional groups were still responsible for differentiation between the three main *Quercus* sections, but the ability for these functional compounds to be used as taxonomic tools in isolation is yet to be established. In the future it may be more beneficial to focus

on variations of sporopollenins in pollen, perhaps through the use of Raman spectroscopy, which preferentially targets the vibration of non-polar bonds in sporopollenins and so may be able to achieve finer-scale differentiation of sporopollenin building blocks (Merlin, 2009).

Third, in this study we used bulk pollen samples to infer differences using FTIR, but fossil-pollen samples would require single-grain measurements since pollen grains are difficult to separate from other organic material within the sediment matrix. Single-grain FTIR spectra are less reproducible than bulk, mostly due to spectral anomalies caused by scattering and by non-radial symmetry of certain pollen types (Zimmermann, Bağcıoğlu, et al., 2015; Zimmermann, 2018). Although these issues have been addressed by adjusting experimental settings and by implementing numerical correction methods (Zimmermann et al., 2016; Zimmermann, 2018), future work is needed to test whether the patterns we observe at the bulk level can be replicated using single-grain FTIR measurements.

Finally, our study shows the importance of using large numbers of replicates in the pollen samples to account for the large amounts of chemical variation present in the chemical spectra, even within replicate species. The large numbers of samples and high levels of replication here (i.e. 50 ± 23 tree replicates per species) are a major advantage over previous studies, which have featured either fewer replicates (<5)(Jardine et al., 2019; Julier et al., 2016; Woutersen et al., 2018) or fewer/no congeneric species (Julier et al., 2016; Zimmermann et al., 2017). Although we do find clear signals in the data linked to systematics (Figure 3 and Table 3), we also find that ~60 % of the total variation remains unexplained. One probable reason for the unexplained variation observed in our study may be linked to the environmental controls on pollen chemistry. Previous studies have suggested plasticity of pollen chemistry to climate and other environmental variables (Zimmermann & Kohler, 2014; Depciuch et al., 2016; Bağcıoğlu et al., 2017; Depciuch, Kasprzyk, Sadik, & Parlińska-Wojtan, 2017; Zimmermann et al., 2017). The other probable reason is the intra-species variation between the genotypes of different populations as well as within populations (Zimmermann et al., 2017). This suggests it will be critical to understand the other factors which can account for this variation if these pollen-chemistry techniques can be successfully applied to fossil sequences.

5. Conclusions

We investigated the chemical variation in pollen sampled from 297 individuals of *Quercus* using FTIR to investigate whether this technique could enable taxonomic discrimination of modern *Quercus* pollen. Our results achieved excellent (~97%) recall to section level, showing that subgenus level differentiation of pollen samples is possible using IR methods. However, despite these promising results at the section level, more detailed, species-level differentiation was complicated by overlapping variation in the chemical composition of closely related species.

We also aimed to identify whether variations in specific functional groups are responsible for any taxonomic discrimination in the data. Here, we found lipids and sporopollenins to be key determinants between different *Quercus* sections. Although the sporopollenin functional groups are identified as important for discrimination between *Quercus* taxa, isolating the effect of these sporopollenin groups from the effects of other functional groups which may not

be preserved in sediment sequences (e.g. lipids) still present a challenge. In addition, testing the application on single-grain *Quercus* samples, and developing a more complete understanding of the effects of environmental variation on pollen-chemical signatures in *Quercus* is required. Taken together, our findings build on previous studies and show that, whilst FTIR approaches on modern *Quercus* pollen can perform at a similar level to SEM techniques, future work on the discrimination of sporopollenin components is required before FTIR can become a more widespread tool in long-term ecology and biogeography. Thus, our study represents a valuable step forward in improving our understanding of variation in pollen chemical composition and its application in long-term ecology and biogeography.

x. Data Availability: Data and code for analysis are available as supplementary material on a github repository (https://github.com/FM-uib/Quercus_Portugal_FTIR) for review. Both will be uploaded to a dryad repository.

xi. References

Acácio, V., Dias, F. S., Catry, F. X., Rocha, M., & Moreira, F. (2017). Landscape dynamics in Mediterranean oak forests under global change: Understanding the role of anthropogenic and environmental drivers across forest types. *Global Change Biology*, *23*(3), 1199–1217. https://doi.org/10.1111/gcb.13487

Amigo, J. (2017). The lowlands and midlands of Northwestern Atlantic Iberia. In J. Loidi (Ed.), *The Vegetation of the Iberian Peninsula: Volume 1* (pp. 191–250). Springer International Publishing.

Bağcıoğlu, M., Kohler, A., Seifert, S., Kneipp, J., & Zimmermann, B. (2017). Monitoring of plant–environment interactions by high-throughput FTIR spectroscopy of pollen. *Methods in Ecology and Evolution*, *8*(7), 870–880. https://doi.org/10.1111/2041-210X.12697

Bağcıoğlu, M., Zimmermann, B., & Kohler, A. (2015). A multiscale vibrational spectroscopic approach for identification and biochemical characterization of pollen. *PLoS One*, *10*(9). https://doi.org/10.1371/journal.pone.0137899

Beug, H. (2004). *Leitfaden der Pollenbestimmung für Mitteleuropa und angrenzende Gebiete.* München: Dr Friedrich Pfeil Verlag.

Birks, H. H. (2008). The Late-Quaternary history of arctic and alpine plants. *Plant Ecology & Diversity*, *1*(2), 135–146. https://doi.org/10.1080/17550870802328652

Birks, H. J. B. (2014). Challenges in the presentation and analysis of plant-macrofossil stratigraphical data. *Vegetation History and Archaeobotany*, *23*(3), 309–330. https://doi.org/10.1007/s00334-013-0430-2

Brewer, S., Cheddadi, R., de Beaulieu, J. L., & Reille, M. (2002). The spread of deciduous Quercus throughout Europe since the last glacial period. *Forest Ecology and Management*, *156*(1), 27–48. https://doi.org/10.1016/S0378-1127(01)00646-6

Brewer, S., Giesecke, T., Davis, B. A. S., Finsinger, W., Wolters, S., Binney, H., … Bradshaw, R. H. (2017). Late-glacial and Holocene European pollen data. *Journal of Maps*, *13*(2), 921–928. https://doi.org/10.1080/17445647.2016.1197613

Dawson, T. P., Jackson, S. T., House, J. I., Prentice, I. C., & Mace, G. M. (2011). Beyond Predictions: Biodiversity Conservation in a Changing Climate. *Science*, *332*(6025), 53–58. https://doi.org/10.1126/science.1200303

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. *Analytical and Bioanalytical Chemistry*, *394*(5), 1443–1452. https://doi.org/10.1007/s00216-009-2794-9

Denk, T., & Grimm, G. W. (2009). Significance of Pollen Characteristics for Infrageneric Classification and Phylogeny in Quercus (Fagaceae). *International Journal of Plant Sciences*, *170*(7), 926–940. https://doi.org/10.1086/600134

Denk, T., Grimm, G. W., Manos, P. S., Deng, M., & Hipp, A. L. (2017). An updated infrageneric classification of the Oaks: Review of previous taxonomic schemes and synthesis of evolutionary patterns. In E. Gil-Pelegrín, J. J. Peguero-Pina, & D. Sancho-Knapik (Eds.), *Oaks Physiological Ecology. Exploring the Functional Diversity of Genus Quercus L.* (pp. 13–38). https://doi.org/10.1007/978-3-319-69099-5_2

Denk, T., & Tekleva, M. V. (2014). Pollen morphology and ultrastructure of Quercus with focus on Group Ilex (=Quercus subgenus Heterobalanus (Oerst.) Menitsky): Implications for oak systematics and evolution. *Grana*, *53*(4), 255–282. https://doi.org/10.1080/00173134.2014.918647

Depciuch, J., Kasprzyk, I., Roga, E., & Parlinska-Wojtan, M. (2016). Analysis of morphological and molecular composition changes in allergenic Artemisia vulgaris L. pollen under traffic pollution using SEM and FTIR spectroscopy. *Environmental Science and Pollution Research International*, *23*(22), 23203–23214. https://doi.org/10.1007/s11356-016-7554-8

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

Edwards, T. H., & Willson, P. D. (1974). Digital Least Squares Smoothing of Spectra. *Applied Spectroscopy*, *28*(6), 541–545.

Fraser, W. T., Scott, A. C., Forbes, A. E. S., Glasspool, I. J., Plotnick, R. E., Kenig, F., & Lomax, B. H. (2012). Evolutionary stasis of sporopollenin biochemistry revealed by unaltered Pennsylvanian spores. *New Phytologist*, *196*(2), 397–401. https://doi.org/10.1111/j.1469-8137.2012.04301.x

Giorgi, F., & Lionello, P. (2008). Climate change projections for the Mediterranean region. *Global and Planetary Change*, *63*(2), 90–104. https://doi.org/10.1016/j.gloplacha.2007.09.005

Godwin, H. (1975). *The history of the British flora: A factual basis for phytogeography.* (2nd ed.). Cambridge: Cambridge University Press.

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. https://doi.org/10.1007/s10453-007-9065-z

Grímsson, F., Grimm, G. W., Meller, B., Bouchal, J. M., & Zetter, R. (2016). Combined LM and SEM study of the middle Miocene (Sarmatian) palynoflora from the Lavanttal Basin, Austria: Part IV. Magnoliophyta 2 – Fagales to Rosales. *Grana*, *55*(2), 101–163. https://doi.org/10.1080/00173134.2015.1096566

Grímsson, F., Zetter, R., Grimm, G. W., Pedersen, G. K., Pedersen, A. K., & Denk, T. (2015). Fagaceae pollen from the early Cenozoic of West Greenland: Revisiting Engler's and Chaney's Arcto-Tertiary hypotheses. *Plant Systematics and Evolution*, *301*, 809–832. https://doi.org/10.1007/s00606-014-1118-5

Guiot, J., & Cramer, W. (2016). Climate change: The 2015 Paris Agreement thresholds and Mediterranean basin ecosystems. *Science*, *354*(6311), 465–468. https://doi.org/10.1126/science.aah5015

Guiot, J., & Kaniewski, D. (2015). The Mediterranean Basin and Southern Europe in a warmer world: What can we learn from the past? *Frontiers in Earth Science*, *3*, 28. https://doi.org/10.3389/feart.2015.00028

Huntley, B., & Birks, H. J. B. (1983). *Atlas of past and present pollen maps for Europe, 0-13,000 years ago*. Cambridge University Press.

Indahl, U. G., Liland, K. H., & Næs, T. (2009). Canonical partial least squares—A unified PLS approach to classification and regression problems. *Journal of Chemometrics*, *23*(9), 495–504. https://doi.org/10.1002/cem.1243

Ivleva, N. P., Niessner, R., & Panne, U. (2005). Characterization and discrimination of pollen by Raman microscopy. *Analytical and Bioanalytical Chemistry*, *381*(1), 261–267. https://doi.org/10.1007/s00216-004-2942-1

Jalas, l & Suominen, l (ed.) 1976: Atlas Florae Europaeae. 3. Salicaceae to Balanophoraceae. - 128 pp. Helsinki.

Jardine, P., Gosling, W., Lomax, B., Julier, A., & Fraser, W. (2019). Chemotaxonomy of domesticated grasses: A pathway to understanding the origins of agriculture. *Journal of Micropalaeontology*, *38*, 83–95. https://doi.org/10.5194/jm-38-83-2019

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

Lahlali, R., Jiang, Y., Kumar, S., Karunakaran, C., Liu, X., Borondics, F., … 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*. https://doi.org/10.3389/fpls.2014.00747

Lang, G. (1994). *Quartäre Vegetationsgeschichte Europas.* Jena: Gustav Fischer Verlag.

Li, F.-S., Phyo, P., Jacobowitz, J., Hong, M., & Weng, J.-K. (2019). The molecular structure of plant sporopollenin. *Nature Plants*, *5*(1), 41–46. https://doi.org/10.1038/s41477-018-0330-7

Liland, K. H. (2017). Package “EMSC” (Version 0.9.0) [R]. Retrieved from https://cran.r-project.org/web/packages/EMSC/EMSC.pdf

Liland, K. H., Mevik, B.-H., Rukke, E.-O., Almøy, T., & Isaksson, T. (2009). Quantitative whole spectrum analysis with MALDI-TOF MS, Part II: Determining the concentration of milk in mixtures. *Chemometrics and Intelligent Laboratory Systems*, *99*(1), 39–48. https://doi.org/10.1016/j.chemolab.2009.07.008

Lindner, M., Fitzgerald, J. B., Zimmermann, N. E., Reyer, C., Delzon, S., van der Maaten, E., … Hanewinkel, M. (2014). Climate change and European forests: What do we know, what are the uncertainties, and what are the implications for forest management? *Journal of Environmental Management*, *146*, 69–83. https://doi.org/10.1016/j.jenvman.2014.07.030

Magri, D., Vendramin, G. G., Comps, B., Dupanloup, I., Geburek, T., Gömöry, D., … Roure, J. M. (2006). A new scenario for the Quaternary history of European beech populations: Palaeobotanical evidence and genetic consequences. *New Phytologist*, *171*(1), 199–221.

Matías, L., Abdelaziz, M., Godoy, O., & Gómez-Aparicio, L. (2019). Disentangling the climatic and biotic factors driving changes in the dynamics of Quercus suber populations across the species‘ latitudinal range. *Diversity and Distributions*, *25*(4), 524–535. https://doi.org/10.1111/ddi.12873

Mehmood, T., & Ahmed, B. (2016). The diversity in the applications of partial least squares: An overview. *Journal of Chemometrics*, *30*(1), 4–17. https://doi.org/10.1002/cem.2762

Merlin, J. C. (2009). Resonance Raman spectroscopy of carotenoids and carotenoid-containing systems. *Pure and Applied Chemistry*, *57*(5), 785–792. https://doi.org/10.1351/pac198557050785

Mevik, B.-H., Wehrens, R., Liland, K. H., & Hiemstra, P. (2019). Partial least squares and principal component regression (Version 2.7-1) [Windows R]. Retrieved from http://mevik.net/work/software/pls.html, https://github.com/bhmevik/pls

Mondol, Patel, Rüger, Stiebing, Kleiber, Henkel, … Schie. (2019). Application of High-Throughput Screening Raman Spectroscopy (HTS-RS) for Label-Free Identification and Molecular Characterization of Pollen. *Sensors*, *19*(20), 4428. https://doi.org/10.3390/s19204428

Nogués-Bravo, D., Rodriguez-Sanchez, F., Orsini, L., de Boer, E., Jansson, R., Morlon, H., … Jackson, S. T. (2018). Cracking the code of biodiversity responses to past Climate Change. *Trends in Ecology & Evolution*, *33*(10), 765–776. https://doi.org/10.1016/j.tree.2018.07.005

Nogués-Bravo, D., Veloz, S., Holt, B. G., Singarayer, J., Valdes, P., Davis, B., … Rahbek, C. (2016). Amplified plant turnover in response to climate change forecast by Late Quaternary records. *Nature Climate Change*, *6*(12), 1115–1119. https://doi.org/10.1038/nclimate3146

Pappas, C. S., Tarantilis, P. A., Harizanis, P. C., & Polissiou, M. G. (2003). New method for pollen identification by FT-IR spectroscopy. *Applied Spectroscopy*, *57*(1), 23–27. https://doi.org/10.1366/000370203321165160

Parodi, G., Dickerson, P., & Cloud, J. (2013). Pollen Identification by Fourier Transform Infrared Photoacoustic Spectroscopy. *Applied Spectroscopy*, *67*(3), 342–348. https://doi.org/10.1366/12-06622

Petit, R. J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., … Kremer, A. (2002). Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management*, *156*(1), 49–74. https://doi.org/10.1016/S0378-1127(01)00634-X

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.R-project.org/

Sanchez, R., Benito-Garzon, M., & Ollero, H. (2009). Present and future extension of the Iberian submediterranean territories as determined from the distribution of marcescent oaks. *Plant Ecology*, *204*, 189–205. https://doi.org/10.1007/s11258-009-9584-5

Savitzky, Abraham., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, *36*(8), 1627–1639. https://doi.org/10.1021/ac60214a047

Schulte, F., Lingott, J., Panne, U., & Kneipp, J. (2008). Chemical Characterization and Classification of Pollen. *Analytical Chemistry*, *80*(24), 9551–9556. https://doi.org/10.1021/ac801791a

Solomon, A. M. (1983a). Pollen Morphology and Plant Taxonomy of Red Oaks in Eastern North America. *American Journal of Botany*, *70*(4), 495–507. https://doi.org/10.1002/j.1537-2197.1983.tb07877.x

Solomon, A. M. (1983b). Pollen Morphology and Plant Taxonomy of White Oaks in Eastern North America. *American Journal of Botany*, *70*(4), 481–494. https://doi.org/10.1002/j.1537-2197.1983.tb07876.x

Telaar, A., Nürnberg, G., & Repsilber, D. (2010). Finding biomarker signatures in pooled sample designs: A simulation framework for methodological comparisons. *Advances in Bioinformatics*. https://doi.org/10.1155/2010/318573

Tschan, G. F., & Denk, T. (2012). Trichome types, foliar indumentum and epicuticular wax in the Mediterranean gall oaks, Quercus subsection Galliferae (Fagaceae): Implications for taxonomy, ecology and evolution. *Botanical Journal of the Linnean Society*, *169*(4), 611–644. https://doi.org/10.1111/j.1095-8339.2012.01233.x

Tutin, T. G., Burges, N. A., Chater, A. O., Edmondson, J. R., Heywood, V. H., Moore, D. M., ... Webb, D. A. (Eds.). (1993). Flora Europaea 1. Psilotaceae to Platanaceae (2nd edition). Cambridge: Cambridge University Press.

Tzedakis, P. C. (1994). Vegetation Change through Glacial-Interglacial Cycles: A Long Pollen Sequence Perspective. *Philosophical Transactions: Biological Sciences*, *345*(1314), 403–432. Retrieved from JSTOR.

Ülker, E., Tavsanoglu, C., & Perktas, U. (2018). Ecological niche modelling of pedunculate oak (Quercus robur) supports the “expansion-contraction” model of Pleistocene biogeography. *Biological Journal of the Linnean Society*, 1–10. https://doi.org/10.1093/biolinnean/blx154/4769765

Veloz, S. D., Williams, J. W., Blois, J. L., He, F., Otto-Bliesner, B., & Liu, Z. (2012). No-analog climates and shifting realized niches during the late quaternary: Implications for 21st-century predictions by species distribution models. *Global Change Biology*, *18*(5), 1698–1713. https://doi.org/10.1111/j.1365-2486.2011.02635.x

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1

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

Zimmermann, B. (2010). Characterization of pollen by vibrational spectroscopy. *Applied Spectroscopy*, *64*(12), 1364–1373. https://doi.org/10.1366/000370210793561664

Zimmermann, B. (2018). Chemical characterization and identification of Pinaceae pollen by infrared microspectroscopy. *Planta*, *247*(1), 171–180. https://doi.org/10.1007/s00425-017-2774-9

Zimmermann, B., Bağcıoğlu, M., Sandt, C., & Kohler, A. (2015). 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. https://doi.org/10.1007/s00425-015-2380-7

Zimmermann, B., Bağcıoğlu, M., Tafinstseva, V., Kohler, A., Ohlson, M., & Fjellheim, S. (2017). A high-throughput FTIR spectroscopy approach to assess adaptive variation in the chemical composition of pollen. *Ecology and Evolution*, *7*(24), 10839–10849. https://doi.org/10.1002/ece3.3619

Zimmermann, B., & Kohler, A. (2013). Optimizing Savitzky-Golay parameters for improving spectral resolution and quantification in infrared spectroscopy. *Applied Spectroscopy*, *67*(8), 892–902. https://doi.org/10.1366/12-06723

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

Zimmermann, B., Tafintseva, V., Bağcıoğlu, M., Berdahl, H. M., & Kohler, A. (2016). Analysis of allergenic pollen by FTIR microspectroscopy. *Analytical Chemistry*, *88*(1), 803–811. https://doi.org/10.1021/acs.analchem.5b03208

Zimmermann, B., Tkalčec, Z., Mešić, A., & Kohler, A. (2015). Characterizing aeroallergens by infrared spectroscopy of fungal spores and pollen. *PLOS ONE*, *10*(4), e0124240. https://doi.org/10.1371/journal.pone.0124240

xii. Biosketch: Florian Muthreich is a palaeoecologist interested in developing new methods of pollen classification. This work represents a component of his PhD work at the University of Bergen University within the PollChem project (https://www.uib.no/en/rg/EECRG/98775/pollchem). In this project he and other authors collaborate to explore pollen chemistry applications in biogeography and long-term ecology.

Author contributions: FM, AWRS, BZ and HJBB conceived the idea; FM and CMVV conducted the fieldwork and collected the data; FM, BZ, and AWRS analysed the data; and FM and AWRS led the writing with assistance from BZ, HJBB, and CMVV