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

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

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

iv. Affiliations: [1]University of Bergen, Bergen, Norway. [2]Norwegian University of Life Sciences, Ås, Norway. [3]CIBIO - Research Center In Biodiversity and Genetic Resources, Vairão, Portugal [4]Bjerknes Centre for Climate Research, Bergen, Norway. [5]Environmental Change Research Centre, University College London, London, UK.

*corresponding author

v. Acknowledgements: We dedicate this paper to the memory of John Flenley (1936-2018) who pioneered so many exciting aspects of pollen analysis and vegetation history. One of us (HJBB) has known John since 1963 and was always stimulated by John's latest ideas and palaeoecological studies.

This work is part of the PollChem project funded with a FRIPRO Grant (PollChem 214359); the extensive field work was possible through funding by the L. Meltzer Høyskolefond (2017/05/LMH) and the Olaf Grolle Olsen Legat (2017/52/FOL).

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, to determine the prospects of chemical techniques as for identification in biogeographic research.

Location: Portugal

Taxon: Six species of *Quercus* L., *Q. faginea*, *Q. robur*, *Q. robur 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. Chemical functional groups associated with variation in the FTIR spectra between different *Quercus* species were determined.

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 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 22% and 31% of the test-set.

Main Conclusions: Our results demonstrate that sub-genus 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 perform equally to traditional morphological techniques. However, although sporopollenins are identified as important functional groups for discrimination between some *Quercus* taxa, the importance of lipids to resolve phylogenetic signatures presents challenges for future biogeographic research because lipids may be less likely to be preserved in sediment sequences. We suggest that future work on further discrimination of isolated sporopollenin components is required before FTIR-based chemical discrimination can be used in long-term ecology and biogeography studies.

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

vii. Main text

1. Introduction

Palaeoecology uses biological proxies (e.g. the sub-fossilised remains of plants and animals) preserved in sediments to reconstruct past biotic changes over decadal to multi-millennial timescales. These reconstructions can be used to address a range of biogeographic questions related to species-range dynamics in response to the Quaternary glaciations (Webb, 1986; Davis & Shaw, 2001), information about long-term successional processes and the establishment and dynamics of plant communities (Delcourt, Delcourt, & Webb, 1982; Ritchie, 1995; Mitchell & Cole, 1998), and temporal and spatial changes in plant-diversity gradients and their links with climate and anthropogenic change (Odgaard, 1999; Haskell, 2001). Fossil pollen and spores are a critical tool to provide insight into these topics. The exine, or outer cell wall, of pollen and spores is made up of a set of complex biomolecules known as sporopollenins (Scott, 1994). Sporopollenins are highly resistant to corrosion and provide a protective outer layer for the genetic material, energy and carbon reserves (carbohydrates and lipids), and proteins (nutrients and pollen-specific transcriptomes) stored within the pollen intine. In palaeoecological research, the sporopollenin-based exine has traditionally been used as the main tool to differentiate between different pollen types (Erdtman & Praglowski, 1959; Fægri & Iversen, 1989). Since the structural complexity of the exine varies considerably between genera and families, analysis of subsequent sediment samples using traditional methods under the light microscope has enabled insights into biogeographic patterns and processes through reconstructions of past landscape and vegetation change (Prentice, Guiot, Huntley, Jolly, & Cheddadi, 1996; Prentice & Webb, 1998; Elenga et al., 2000).

Although traditional morphological approaches continue to yield insights into the patterns and processes of vegetation changes in the past, these methods of identification are time consuming, and this information is often limited by resolving pollen types only to the genus, or even family level. One potential solution lies in the chemical analysis of pollen and spores. Evidence is mounting that analysis of pollen grains using Fourier Transform Infrared Spectroscopy (FTIR) may provide a new tool for differentiating between pollen types extracted from sediment sequences (Julier et al., 2016; Woutersen et al., 2018; Jardine, Gosling, Lomax, Julier, & Fraser, 2019), since pollen-grain chemistry may itself show biogeographic patterns related to phylogeny and environmental conditions (Bağcıoğlu, Kohler, Seifert, Kneipp, & Zimmermann, 2017; Zimmermann et al., 2017). 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 (Gottardini, Rossi, Cristofolini, & Benedetti, 2007; Ivleva, Niessner, & Panne, 2005; Pappas, Tarantilis, Harizanis, & Polissiou, 2003; Parodi, Dickerson, & Cloud, 2013; Schulte, Lingott, Panne, & Kneipp, 2008; Zimmermann, 2010; Zimmermann & Kohler, 2014; Bağcıoğlu, Zimmermann, & Kohler, 2015). For example, Zimmermann (2010) classified pollen from 11 species at comparable accuracy to conventional light-microscopy methods, whilst later work by Zimmermann & Kohler (2014) and Depciuch, Kasprzyk, Drzymała, & Parlinska-Wojtan (2018) expanded on the differences in pollen composition by analysing over 300 different species and identified lipids (specifically triglycerides) as being major sources of variation in the pollen spectra of congeneric species across several diverse genera (*Quercus, Iris, Pinus*, *Betula*, etc). Differences between the sporopollenin-spectral regions of the pollen of 15 morphologically-similar Pinales species (Bağcıoğlu et al., 2015) have also recently been observed, whilst two independent studies have demonstrated that the pollen spectra of pollen of both Poaceae and Nitrariaceae can be used to generate sub-family level differentiation using FTIR (Julier et al., 2016; Woutersen et al., 2018; Zimmermann et al., 2017; P. Jardine et al., 2019).

Taxonomic differentiation of pollen based on the chemical variations inferred by FTIR undoubtedly shows promise, but there are a number of challenges that remain to be solved to enable more extensive application across palaeoecology. First, lipids and proteins have been found to

be major discriminants of variation in FTIR spectra (Bağcıoğlu et al., 2017; Zimmermann & Kohler, 2014), but these may not be preserved in sediment sequences as compared to the sporopollenin. For example, a study on fungal basidiospores, conducted on the archived samples collected within the last 50 years, demonstrated changes in lipid chemistry associated with the increasing storage time (Zimmermann, Tkalčec, Mešić, & Kohler, 2015). Therefore, although variations in sporopollenin components between pollen types have been observed, 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. Second, understanding of the potential impact of environmental variation on the pollen-chemistry signature remains limited. For example, some of the variation in lipid, protein, and carbohydrate chemistry may be explained by environmental influences such as temperature during pollen maturation (Lahlali et al., 2014; Zimmermann & Kohler, 2014; Jiang et al., 2015; Zimmermann et al., 2017), whilst sporopollenin composition may be influenced by external environmental influences such as the plant exposure to UV-B radiation (Rozema et al., 2001; Blokker, Boelen, Broekman, & Rozema, 2006; Willis et al., 2011; Wesley T. Fraser et al., 2011; P. E. Jardine, Abernethy, Lomax, Gosling, & Fraser, 2017; Bell et al., 2018). As a result, the extent to which variation across larger environmental gradients might override taxonomic differences in chemical composition remains unclear. Finally, the majority of studies investigating pollen chemistry have so far used pollen from herbaria, botanical gardens, or university campuses, with a limited number of replicates (usually less than three) per location and large number of different species and families (often more than ten). Although sampling from botanic gardens offers ease of access, reliable identification, and a broad species range, the number of replicates in these analyses remains a limiting factor when investigating environmental and taxonomic variability in pollen chemistry. As a result, extensive sampling across large spatial gradients are required.

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* pollen. *Quercus* makes an excellent study system for this purpose because it has a wide geographical distribution (Colombo, Lorenzoni, & Grigoletto, 1983; Rushton, 1993; Manos, Doyle, & Nixon, 1999; Petit, Bodénès, Ducousso, Roussel, & Kremer, 2003). Investigating the chemical differences between different *Quercus* pollen types also has implications for palaeoecological research, particularly for studies based in the Mediterranean region. *Quercus* pollen is often abundant in Mediterranean-pollen records (e.g. Carrión, Parra, Navarro, & Munuera, 2000; Brewer, Cheddadi, de Beaulieu, & Reille, 2002; Petit et al., 2002), but despite its high species diversity, pollen from the genus *Quercus* can only generally be separated into three morphologically distinctive types in pollen sequences based on differences in surface ornamentation of the grains. Given the importance of incorporating long-term ecological information to improve biodiversity forecasts of environmental change (Dawson, Jackson, House, Prentice, & Mace, 2011), the ability to differentiate *Quercus* to species level in the pollen record would greatly improve our understanding of the ecological dynamics underlying Mediterranean ecosystems in the past and future (e.g. Guiot & Cramer, 2016).

Here, we use a total of 297 trees sampled across a climatic and topographic gradient in Portugal to investigate within and between species variability in *Quercus* pollen chemistry at a regional scale. Our dataset is unique because, to our knowledge, it represents the largest collection of closely related species sampled from populations outside botanic gardens, with the specific goal of testing the possibilities of taxonomic differentiation using FTIR. Using this dataset, we aim to investigate whether within- and between-species variability can be used to differentiate sections and species of *Quercus*. Addressing these questions represents the first step if we are to successfully use pollen chemistry as a tool to improve our understanding of past biogeographic patterns and processes in plants.

2. Methods

2.1 Sample collection

We collected pollen samples from 297 individual trees belonging to six *Quercus* species across a 4° (~450 km) latitudinal gradient in Portugal, from Porto in the north to Lisbon in the south and the Spanish border east of Vila Nova de Foz Côa. The *Quercus* species in this study belong to the sections Cerris, Ilex, and Quercus according to Denk et al., (2017) (Table 1) and have different geographic distributions (Figure 1). *Q. robur,* for example, has a temperate distribution and is more abundant in north-west Portugal. It occurs in regions with sufficient summer rain, and is absent from areas with summer drought (Amigo, 2017; Ülker, Tavsanoglu, & Perktas, 2018). *Q. faginea s.l.* has a broad distribution in the Mediterranean Basin and 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). *Q. suber* is distributed across the whole study region on siliceous bedrock but is absent from areas with winter cold (Amigo, 2017; Matías, Abdelaziz, Godoy, & Gómez-Aparicio, 2019). *Q. coccifera* and *Q. rotundifolia* prefer xerophytic conditions and are often co-occuring species. Both are indifferent towards bedrock conditions, but prefer soils without waterlogging, although *Q. coccifera* is a thermophile species and is less tolerant of winter cold. A detailed summary of the number of trees sampled at each location is in the Supplementary Material (Table S1).

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

Subgenus	Section	Species	Subspecies	n
Cerris	Ilex	*Q. coccifera*		36
Cerris	Ilex	*Q. rotundifolia*		38
Cerris	Cerris	*Q. suber*		69
Quercus	Quercus	*Q. faginea*		63
Quercus	Quercus	*Q. robur*		76
Quercus	Quercus	*Q. robur*	*estremadurensis*	15

Trees were sampled along gradients of temperature and precipitation to cover a wide range of environmental conditions. Climatic conditions across the study region vary, with oceanic conditions along the coast, and a drier Mediterranean climate in the south and towards the Spanish border. East of Porto is the area with the highest precipitation in Portugal before the upper regions of the Douro river valley become semi-Mediterranean (Figure 1). The underlying soil changes from predominantly silica-rich bedrock in the north to clay in the south and a pocket of calcareous bedrock around Sintra (Rivas-Martinez, Penas, del Rio, Gonzales, & Rivas-Saenz, 2017).

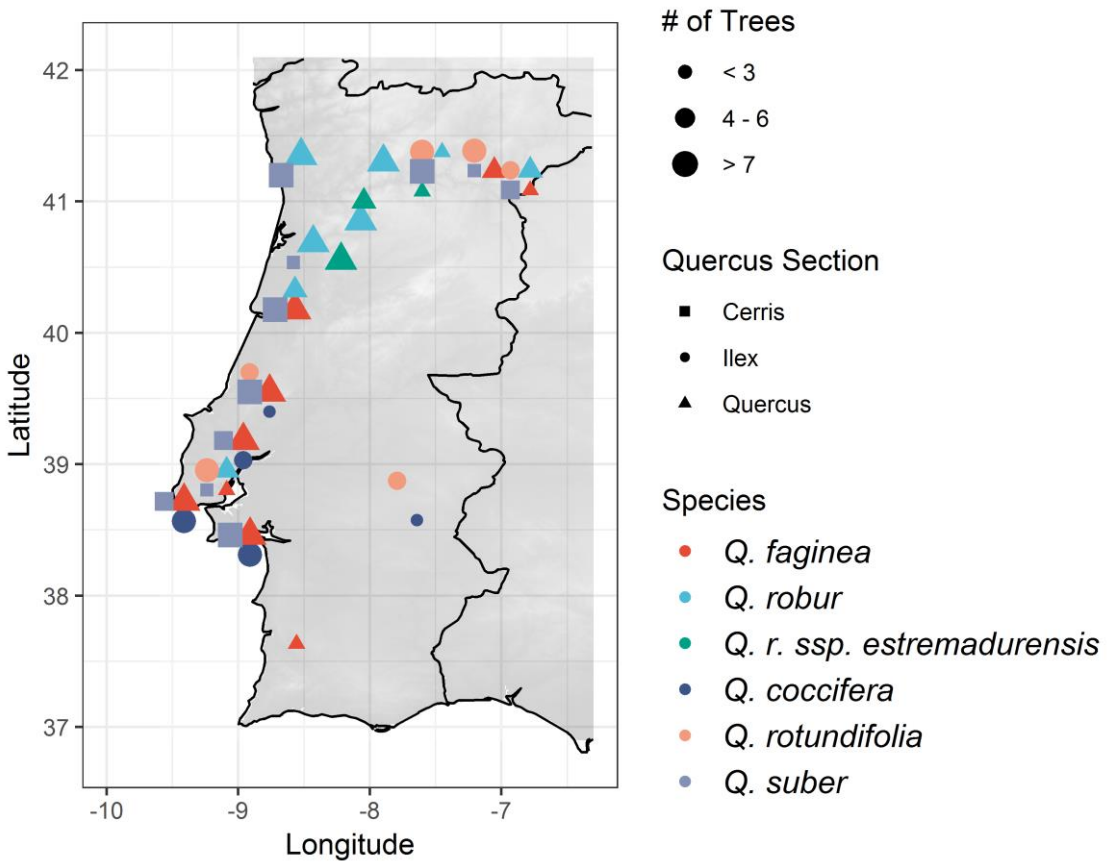

Figure 1. Map of Portugal displaying the locations where trees where sampled. Colour indicates species, shape indicates Quercus section following Denk et al. (2017) and size of the symbol indicates number of trees sampled per location.

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 applied minimal processing on the spectra, because the spectra were reproducible with three replicate analyses. All further processing was performed on the mean spectra of the replicate analyses. The processing of the spectra included a single pass of Extended Multiplicative Signal Correction (EMSC) using the EMSC package (Liland, 2017) as well as second-derivative calculation and smoothing using the Savitzky-Golay smoothing algorithm (Edwards & Willson, 1974; Savitzky & Golay, 1964). It has been shown that multivariate-regression methods (e.g. partial least squares; PLS)

perform better with preprocessed spectra, as shown by classification of FTIR spectral pollen data in other studies (Zimmermann & Kohler, 2013; Woutersen et al., 2018). For further analyses, the spectra were constrained between 700 and 1800 cm^{-1}. 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
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	
Cellulose	1107; 1055; 1028	C–O–C and C–OH stretch
Amylose	1076; 995*	
Sporopollenin	1605; 1515; 1175*; 852 833 and 816	Aromatic rings in phenylpropanoid acids

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 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 (spectra wavenumbers), which weakens the contribution of non-relevant wavenumbers in the spectra to optimise the covariance between response (species) and explanatory variables (spectra). 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 10 times to create 10 versions of the dataset (folds) with different training/test splits. A CPPLS model was fitted for each fold using leave-one-out (LOO) cross validation 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 six *Quercus* species reveals clear differences in absorbance between taxonomic sections at specific wavelengths associated with specific chemical functional groups (Figure 2). For example, the lipid peak absorbance at ~1750 cm^{-1} is weaker in Ilex pollen 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 the Ilex sections. 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).

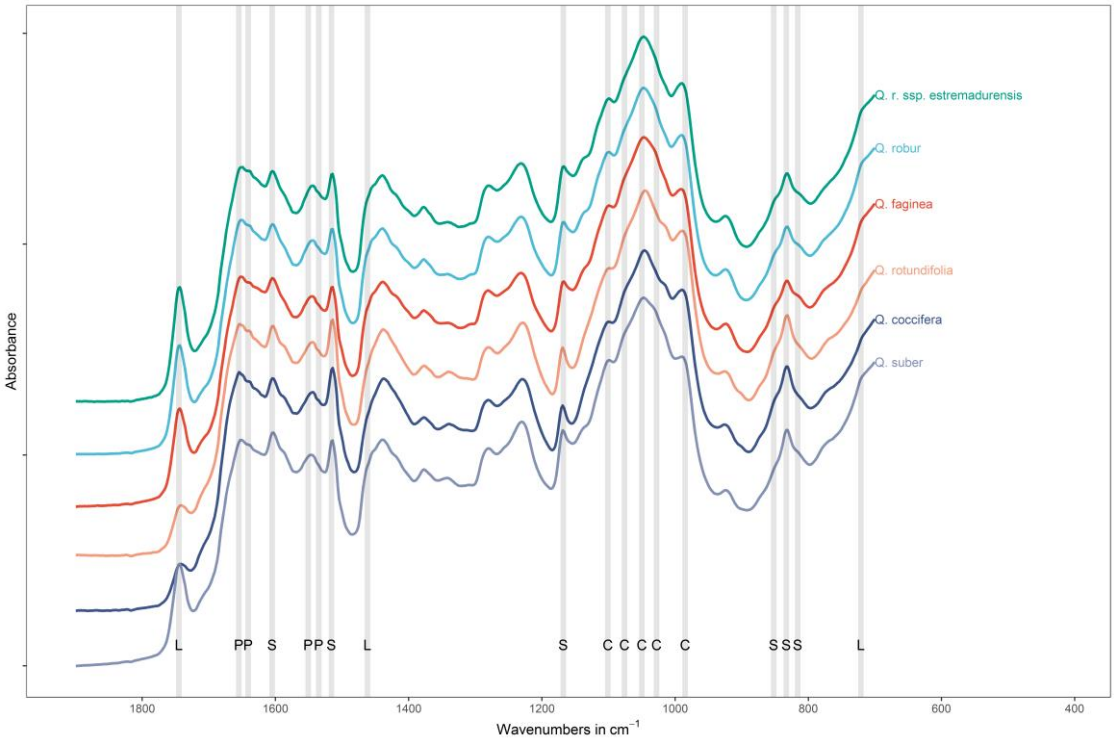

Figure 2 Mean absorbance spectra of Quercus species and notable peak locations. Lipids (L), protein (P), sporopollenin (S), and carbohydrates (C) using wavenumbers compiled in Table 2. Spectra are offset.

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 Ilex section scores are negative along the first component (19.6 % of the variation), whilst individuals from the Quercus and Cerris sections have positive scores along this axis. The Quercus and Cerris sections can mainly be separated along the second principal component (Cerris with negative scores and Quercus with positive scores). However, between-species level variations within different *Quercus* sections species are harder to differentiate, and at the species level in particular there is large overlap within 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, two components explain ~34% of the variation in the dataset and separate the samples into *Quercus* sections, while the next two components explain a further ~8.8% (Figure S1).

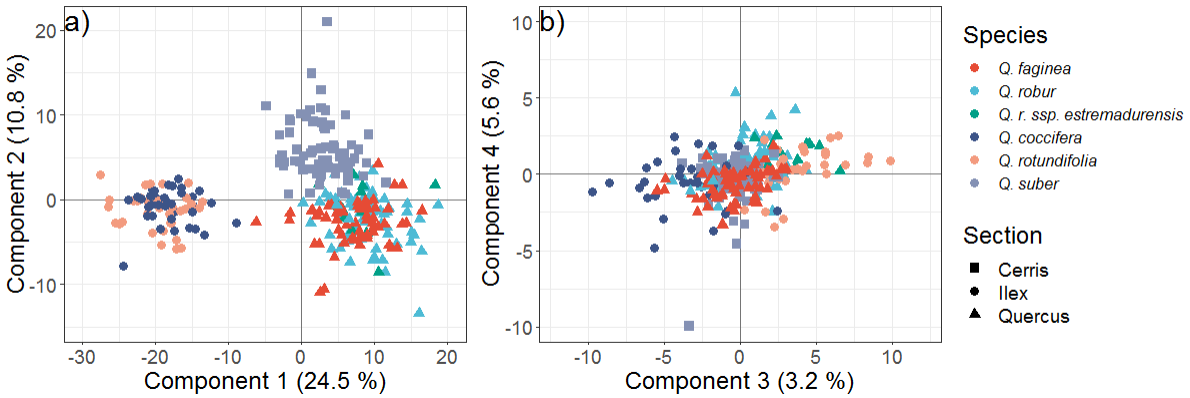

Figure 3 Ordination of components 1-4 of canonical powered partial least squares (CPPLS) models (10-fold). Each point represent a sampled tree. The mean training scores for each sample over the 10-folds was calculated. Proportion of variance explained by each component in parentheses. Colour indicates 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 followed by component 2, while 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 at the *Quercus* section level is achievable based on FTIR.

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 compiled 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 (spectra) and the response (species). Proportion of variance explained by each component in parentheses.

3.2 Discriminant analysis

Using four components (explaining ~40% of the variance) the confusion matrix of the classification CPLS model shows clear differentiation between the *Quercus* sections, with only ~2±2% of *Q. suber* samples misidentified as belonging to the *Quercus* section (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 22% and 31% of the test-set samples (within-section misidentifications: 23% of *Q. robur* as *Q. faginea*; 31% of *Q. faginea* as *Q. robur*; 27% of *Q. coccifera* as *Q. rotundifolia*; 22% of *Q. rotundifolia* as *Q. coccifera*). Overall species accuracy within sections ranges from 64% to 78% in the Ilex and Quercus sections. Increasing the components available to the model to 10 (~50% explained variance) increases species accuracy by 5-10 percentage points in the Ilex and Quercus section (Table S2), with higher uncertainty between Ilex species than Quercus section species. As demonstrated with our ordination plots (Figure 3), differentiation of the three sections of *Quercus* is possible using 40% 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 counts that were predicted as species and sum to 100% columnwise, e.g. for Q. robur 23% of spectra were predicted as Q. faginea, 72% were correctly classified as Q. robur and 5% as estremadurensis. Q. r. estr: Q. robur estremadurensis, Q. rotund.: Q. rotundifolia. Green signifies Quercus section, blue is Ilex section and red is Cerris section

Pred/Ref	*Q. faginea*	*Q. robur*	*Q. r. estr.*	*Q. coccifera*	*Q. rotund.*	*Q. suber*
Q. faginea	64 ±9	23 ±7	28 ±22	0	0	1 ±2
Q. robur	1 ±2	72 ±8	60 ±21	0	0	0
Q. r. estr.	31 ±10	5 ±4	9 ±9	0	0	1 ±2
Q. coccifera	0	0	0	73 ±15	22 ±14	0
Q. rotund.	0	0	0	27 ±15	78 ±14	0
Q. suber	5 ±4	0	2 ± 8	0	0	98 ±3

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, Bağcıoğlu, Sandt, & Kohler, 2015; Zimmermann, Tafintseva, Bağcıoğlu, Berdahl, & Kohler, 2016; Julier et al., 2016; Zimmermann et al., 2017; Woutersen et al., 2018; Jardine et al., 2019; Zimmermann & Kohler, 2014). Our results build on these previous studies to reveal the potential for chemical variations in pollen to distinguish sub-generic variation between species belonging to three different *Quercus* sections in pollen samples from 297 individuals from Portugal. We identify a clear separation of *Quercus* into the three taxonomic sections of Ilex, Cerris, and Quercus (Fig. 3 and Table 3), using two components of a PLS model and only 30% of the explained variance in the spectral data. One component (component 1) can be used to differentiate Ilex from Quercus, while component 2 can be used to separate Cerris. Combined, these two components achieve the performance equivalent to traditional palynological methods using light microscopy, where *Quercus* pollen can be 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 ~50% of the variance (compared to ~40% 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 the Quercus section) are misclassified as the other species. Thus, while our results indicate that sub-generic 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 individual 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 separation of different Quercus species 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 separations between the three sections of *Quercus* under FTIR? In our study, we find that lipid peaks are the most important factors in separating Ilex samples from the taxa in the datasets. 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 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* species. Indeed, our results extend the inferences made in that previous study by demonstrating a sub-generic 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 *Q. suber* from other taxa. 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*. These 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. The chemical structure of sporopollenin is comprised of phenylpropanoid building blocks and hydroxylated fatty acids (Schulze Osthoff & Wiermann, 1987), and it is thought that the composition of these phenylpropanoid units is species and environment dependent (Vogt, 2010). Thus, it is possible that a species-specific diagnostic tool based on sporopollenin exists for at least one taxon between the different *Quercus* sections using FTIR. Given that we analysed bulk samples, it is unlikely that these differences are a result of scattering effects related to the specific sporopollenin structures. Nevertheless, 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 phylogenetically important information we used to distinguish between the species is stored in the lipids and sporopollenin components of the pollen chemistry.

4.3 Prospects for developments using fossil pollen

Long-term ecological (i.e. fossil pollen) data are an important tool in macroecological and biogeographic research. When fossil pollen data are integrated from multiple sites at continental scales, they can be used to (i) track relative-niche shifts in association with the emergence of no-analogue climates since the Last Glacial Maximum (Veloz et al., 2012); (ii) forecast future range shifts as a result of climate change through integration with statistical species-distribution modelling approaches (e.g. Nogués-Bravo et al., 2016, 2018), and (iii) test hypotheses surrounding the factors relating to the emergence of disequilibrium of ecological communities (Webb, 1986; Gaüzère, Iversen, Barnagaud, Svenning, & Blonder, 2018). A number of recent palaeoecological studies have also attempted to link fossil pollen data to species-based functional-trait databases via a functional-biogeographic approach (Violle, Reich, Pacala, Enquist, & Kattge, 2014). However, a common limitation of the preceding studies has been the taxonomic mismatch between the pollen data and the macroecological tools and databases that have been used, with the fossil pollen data often resolved only to the genus or even family level. This taxonomic mismatch is likely to create large uncertainties in any model inferences. As a result, any technique that can be used to increase the potential taxonomic resolution of fossil pollen records has the potential to have major biogeographic implications.

The results presented in this paper build on previous research which has proposed that FTIR may be a tool to improve classical, morphology based classification of pollen in long-term ecological research based on palynological records (Julier et al., 2016; Woutersen et al., 2018; Zimmermann, 2018; Jardine et al., 2019). Given the high biodiversity of *Quercus* and the potential sensitivity of Mediterranean ecosystems to future climate changes (Guiot & Cramer, 2016), we were motivated to investigate the potential for using FTIR to distinguish variations in *Quercus* spp. What, then, do our results tell us about the applicability of these FTIR approaches as a tool for taxonomic identification of *Quercus* in pollen records?

Our study reveals clear differences in FTIR spectra across the dataset, enabling sub-generic level differentiation of *Quercus* species belonging to three different sections. In contrast, *within*-section variability is more difficult to differentiate using these approaches. As a result, our findings demonstrate that FTIR methods may work at a comparable level to traditional pollen-morphological approaches, within which skilled pollen analysts are often able to differentiate between the three different *Quercus* sections. However, despite the potential, we note a number of caveats that should be considered before the application can be applied to the fossil record.

First, these analyses are based on fresh pollen sampled from modern taxa. Using these samples, successful differentiation between taxa from different sections is primarily driven by lipids, followed by sporopollenins (Figure 4). Although sporopollenin is stable enough to allow the identification of morphological features for pollen classification under favourable (i.e. anoxic) preservation conditions (Friis, Pedersen, & Crane, 2001), the extent to which these lipids are preserved in chemical sequences in sub-fossil pollen sequences remains uncertain. Thus, although lipids are identified as a primary driver of chemical variation between Ilex and the Quercus and Cerris sections, it may be more beneficial in future studies to focus primarily on variations within the sporopollenins, whose preservation and stability in fossil sequences are well established (Fraser et al., 2012).

In this study we find some evidence to suggest that sporopollenin peaks may be a useful classification scheme to differentiate between the *Quercus* sections (Figure 3). In particular, the sporopollenin peak at 852 cm^{-1} is influential in the differentiation of *Q. suber* from other taxa (Figure 4). Furthermore, although FTIR is very useful for characterising a broad range of chemical functional

groups within a pollen grain and is therefore being used widely in palaeoecological research based on pollen chemistry, it may not actually be the most optimal method for resolving finer-scale differences within sporopollenins. This is because of the nature of sporopollenin, which is abundant in non-polar bonds of long-chain fatty acids crosslinked with phenylpropanoid building blocks. Whilst FTIR relies largely on vibrational absorbance by polar chemical functional groups and to a lesser extent on non-polar bonds, Raman uses induced energy changes, which preferably target the non-polar bonds characterised in sporopollenins. In this regard, Raman spectroscopy may be a better candidate for fossil-pollen taxonomy in long-term ecological research, since it may resolve the structure of sporopollenin in more detail through detection of caretonoid building blocks (Merlin, 2009). Future research, which aims to resolve finer scale variations in the sporopollenin component, may be useful to determine whether pollen-chemical variations can be used to distinguish phylogenetic variation in fossil pollen.

Second, while this study infers chemical differences using FTIR analysis on bulk pollen samples, 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.

Third, 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 related to phylogenetic structure (Figure 3 and Table 3), we also find that ~60 % of the total variation remains unexplained. We suggest 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.

One possible 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; Bağcıoğlu et al., 2017; Zimmermann et al., 2017). For example, temperature has been connected to changes in protein as well as lipid content in both controlled and field experimental conditions (Lahlali et al., 2014; Zimmermann & Kohler, 2014; Jiang et al., 2015; Zimmermann et al., 2017), whilst the effects of exposure to UV-B radiation on pollen chemistry have gained increasing attention (Rozema et al., 2001; Blokker et al., 2006; Fraser et al., 2011; Willis et al., 2011; Jardine et al., 2017; Bell et al., 2018). Additionally, there may be confounding effects related to local adaptation and hybridisation. For example, Bell et al. (2018) investigated changes in *Cedrus* pollen chemistry to UV-B radiation and their results suggest a heritable component in the pollen chemical response to UV-B, from analyses of FTIR spectra of pollen samples from botanic gardens closely resembling their source of origin. *Quercus* as a genus also shows propensity for hybridisation and complex phylogenetic patterns (Rushton, 1993), and together these factors might be expected to increase the amount of variation we observe in pollen-chemical spectra derived from FTIR. In summary, the larger number of replicates used in this study provide us with a greater understanding of the amount of variation FTIR observed in *Quercus* pollen grains, and separating the effects of environment, local adaption, and sub-species variation is an important challenge for future studies of pollen chemistry if we are to fully

characterise chemical variation within any given pollen type and use FTIR approaches on pollen grains from sediments.

5. Conclusions

We investigated the chemical variation in pollen sampled from 297 individuals of *Quercus* using FTIR to see whether this technique could enable taxonomic discrimination of pollen, with the goal of using this technique in future long-term ecological and biogeographic analyses. Our results achieved excellent (~97%) recall to section level, showing that sub-genus 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 some *Quercus* taxa, the importance of lipids in determining the whole signal presents challenges for future biogeographic research using FTIR because lipids may be less likely to be preserved in sediment sequences. 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 highly skilled palynologists using traditional morphological 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 upon acceptance of manuscript.

xi. References

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

Bell, B. A., Fletcher, W. J., Ryan, P., Seddon, A. W. R., Wogelius, R. A., & Ilmen, R. (2018). UV-B-absorbing compounds in modern *Cedrus atlantica* pollen: The potential for a summer UV-B proxy for Northwest Africa. *The Holocene*, *28*(9), 1382–1394. https://doi.org/10.1177/0959683618777072

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 Ecology*. https://doi.org/10.1007/s11258-005-9026-y

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

Carrión, J. S., Parra, I., Navarro, C., & Munuera, M. (2000). Past distribution and ecology of the cork oak (*Quercus suber*) in the Iberian Peninsula: A pollen-analytical approach. *Diversity and Distributions*, *6*(1), 29–44. https://doi.org/10.1046/j.1472-4642.2000.00070.x

Colombo, P. M., Lorenzoni, F. C., & Grigoletto, F. (1983). Pollen grain morphology supports the taxonomical discrimination of Mediterranean Oaks (Quercus, Fagaceae). *Plant Systematics and Evolution*, *141*(3), 273–284.

Davis, M. B., & Shaw, R. G. (2001). Range Shifts and Adaptive Responses to Quaternary Climate Change. *Science*, *292*(5517), 673–679. https://doi.org/10.1126/science.292.5517.673

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

Delcourt, H. R., Delcourt, P. A., & Webb, T. (1982). Dynamic plant ecology: The spectrum of vegetational change in space and time. *Quaternary Science Reviews*, *1*(3), 153–175. https://doi.org/10.1016/0277-3791(82)90008-7

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., 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

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

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

Elenga, H., Peyron, O., Bonnefille, R., Jolly, D., Cheddadi, R., Guiot, J., … Jonson, H. (2000). Pollen-based biome reconstruction for southern Europe and Africa 18,000 yr BP. *Journal of Biogeography*, *27*(3), 621–634. https://doi.org/10.1046/j.1365-2699.2000.00430.x

Erdtman, G., & Praglowski, J. R. (1959). Six notes on pollen morphology and pollenmorphological techniques. *Botaniske Notiser*, *112*, 175–184.

Fægri, K., & Iversen, J. (1989). *Textbook of Pollen Analysis* (4th ed.; K. Faegri, P. E. Kaland, & K. Krzywinski, Eds.). John Wiley & Sons, New York.

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

Fraser, Wesley T., Sephton, M. A., Watson, J. S., Self, S., Lomax, B. H., James, D. I., … 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 Research*, *30*(00). https://doi.org/10.3402/polar.v30i0.8312

Friis, E. M., Pedersen, K. R., & Crane, P. R. (2001). Fossil evidence of water lilies (Nymphaeales) in the Early Cretaceous. *Nature*, *410*(6826), 357. https://doi.org/10.1038/35066557

Gaüzère, P., Iversen, L. L., Barnagaud, J.-Y., Svenning, J.-C., & Blonder, B. (2018). Empirical predictability of community responses to Climate Change. *Frontiers in Ecology and Evolution*, *6*. https://doi.org/10.3389/fevo.2018.00186

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

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

Haskell, J. (2001). The latitudinal gradient of diversity through the Holocene as recorded by fossil pollen in Europe. *Evolutionary Ecology Research*, (3), 345–360.

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

Jardine, P. E., Abernethy, F. A. J., Lomax, B. H., Gosling, W. D., & Fraser, W. T. (2017). Shedding light on sporopollenin chemistry, with reference to UV reconstructions. *Review of Palaeobotany and Palynology*, *238*, 1–6. https://doi.org/10.1016/j.revpalbo.2016.11.014

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

Jiang, Y., Lahlali, R., Karunakaran, C., Kumar, S., Davis, A. R., & Bueckert, R. A. (2015). Seed set, pollen morphology and pollen surface composition response to heat stress in field pea: Heat stress on reproductive development in pea. *Plant, Cell & Environment*, *38*(11), 2387–2397. https://doi.org/10.1111/pce.12589

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

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

Manos, P. S., Doyle, J. J., & C. Nixon, K. (1999). Phylogeny, biogeography, and processes of molecular differentiation in Quercus Subgenus Quercus (Fagaceae). *Molecular Phylogenetics and Evolution*, *12*(3), 333.349.

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

Mitchell, F., & Cole, E. (1998). Reconstruction of long!term successional dynamics of temperate woodland in Bialowieza Forest, Poland. *Journal of Ecology*, (86), 1042–1059.

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 Ecol Evol*, *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

Odgaard, B. V. (1999). Fossil pollen as a record of past biodiversity. *Journal of Biogeography*, *26*(1), 7–17. https://doi.org/10.1046/j.1365-2699.1999.00280.x

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., Bodénès, C., Ducousso, A., Roussel, G., & Kremer, A. (2003). Hybridization as a mechanism of invasion in oaks. *New Phytologist*, *161*(1), 151–164. https://doi.org/10.1046/j.1469-8137.2003.00944.x

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

Prentice, I. C., Guiot, J., Huntley, B., Jolly, D., & Cheddadi, R. (1996). Reconstructing biomes from palaeoecological data: A general method and its application to European pollen data at 0 and 6 ka. *Climate Dynamics*, *12*(3), 185–194. https://doi.org/10.1007/BF00211617

Prentice, I. C., & Webb, T. (1998). BIOME 6000: Reconstructing global mid-Holocene vegetation patterns from palaeoecological records. *Journal of Biogeography*, *25*(6), 997–1005. https://doi.org/10.1046/j.1365-2699.1998.00235.x

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

Ritchie, J. C. (1995). Current trends in studies of long-term plant community dynamics. *New Phytologist*, *130*(4), 469–494. https://doi.org/10.1111/j.1469-8137.1995.tb04325.x

Rivas-Martinez, S., Penas, A., del Rio, S., Gonzales, T. E. D., & Rivas-Saenz, S. (2017). Bioclimatology of the Iberian Peninsula and the Balearic Islands. In J. Loidi (Ed.), *The Vegetation of the Iberian Peninsula: Volume 1* (pp. 29–80). Springer International Publishing.

Rozema, J., Broekman, R. A., Blokker, P., Meijkamp, B. B., de Bakker, N., van de Staaij, J., … Kars, S. M. (2001). UV-B absorbance and UV-B absorbing compounds (para-coumaric acid) in pollen and sporopollenin: The perspective to track historic UV-B levels. *Journal of Photochemistry and Photobiology B: Biology*, *62*(1–2), 108–117.

Rushton, B. S. (1993). Natural hybridization within the genus Quercus L. *Annals of Forest Science*, *50*, 73–90.

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

Schulze Osthoff, K., & Wiermann, R. (1987). Phenols as integrated compounds of sporopollenin from Pinus pollen. *Journal of Plant Physiology*, *131*(1), 5–15. https://doi.org/10.1016/S0176-1617(87)80262-6

Scott, R. J. (1994). Pollen exine – the sporopollenin enigma and the physics of pattern. In A. D. Stead & R. J. Scott (Eds.), *Molecular and Cellular Aspects of Plant Reproduction* (pp. 49–82). https://doi.org/10.1017/CBO9780511752339.006

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

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

Violle, C., Reich, P. B., Pacala, S. W., Enquist, B. J., & Kattge, J. (2014). The emergence and promise of functional biogeography. *Proceedings of the National Academy of Sciences*, *111*(38), 13690–13696. https://doi.org/10.1073/pnas.1415442111

Vogt, T. (2010). Phenylpropanoid Biosynthesis. *Molecular Plant*, *3*(1), 2–20. https://doi.org/10.1093/mp/ssp106

Webb, T. (1986). Is vegetation in equilibrium with climate? How to interpret late-Quaternary pollen data. *Vegetatio*, *67*(2), 75–91. https://doi.org/10.1007/BF00037359

Willis, K. J., Feurdean, A., Birks, H. J. B., Bjune, A. E., Breman, E., Broekman, R., … Rozema, J. (2011). Quantification of UV-B flux through time using UV-B-absorbing compounds contained in fossil Pinus sporopollenin. *New Phytologist*, *192*(2), 553–560. https://doi.org/10.1111/j.1469-8137.2011.03815.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. *Ecol Evol*, *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