Review of Palaeobotany and Palynology 235 (2016) 140–147

Contents lists available at ScienceDirect

Review of Palaeobotany and Palynology

journal homepage: www.elsevier.com/locate/revpalbo

Chemotaxonomy as a tool for interpreting the cryptic diversity of Poaceae pollen

Adele C.M. Julier [a], Phillip E. Jardine [a], Angela L. Coe [a], William D. Gosling [a,b,*], Barry H. Lomax [c], Wesley T. Fraser [a,d]

[a] Department of Environment, Earth and Ecosystems, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK
[b] Palaeoecology & Landscape Ecology, Institute for Biodiversity & Ecosystem Dynamics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
[c] The School of Biosciences, The University of Nottingham, Sutton Bonington Campus, Leicestershire LE12 5RD, UK
[d] Geography, Department of Social Sciences, Oxford Brookes University, Gipsy Lane, Oxford OX3 0BP, UK

ARTICLE INFO

Article history:
Received 8 February 2016
Received in revised form 22 August 2016
Accepted 29 August 2016
Available online 8 October 2016

Keywords:
Fourier Transform Infra-red Spectroscopy
Pollen identification
Poaceae
Sporopollenin
Taxonomy

ABSTRACT

The uniform morphology of different species of Poaceae (grass) pollen means that identification to below family level using light microscopy is extremely challenging. Poor taxonomic resolution reduces recoverable information from the grass pollen record, for example, species diversity and environmental preferences cannot be extracted. Recent research suggests Fourier Transform Infra-red Spectroscopy (FTIR) can be used to identify pollen grains based on their chemical composition. Here, we present a study of twelve species from eight subfamilies of Poaceae, selected from across the phylogeny but from a relatively constrained geographical area (tropical West Africa) to assess the feasibility of using this chemical method for identification within the Poaceae family. We assess several spectral processing methods and use K-nearest neighbour (k-nn) analyses, with a leave-one-out cross-validation, to generate identification success rates at different taxonomic levels. We demonstrate we can identify grass pollen grains to subfamily level with an 80% success rate. Our success in identifying Poaceae to subfamily level using FTIR provides an opportunity to generate high taxonomic resolution datasets in research areas such as palaeoecology, forensics, and melissopalynology quickly and at a relatively low cost.

1. Introduction

The correct identification of pollen grains is an important factor in any research area that uses pollen assemblages to make inferences about vegetation. These research areas can be as diverse as palaeoecology (Germeraad et al., 1968; Mander and Punyasena, 2014), forensics (Horrocks et al., 1998; Mildenhall et al., 2006) and melissopalynology (Herrero et al., 2002; Martin, 2005), as they all share a reliance upon the taxonomic resolution of pollen identification to maximise the accuracy and usefulness of their data. Looking further back into geological time, palynological research has played a fundamental role in understanding plant origination and radiation (e.g. the origin and radiation of vascular plants (Rubinstein et al., 2010), and the radiation of the angiosperms (Lupia et al., 1999)), and shaped our understanding of how the terrestrial biosphere responded to mass extinction events (Looy et al., 2001; Tschudy et al., 1984). This highly diverse group of studies all shares a reliance upon the taxonomic resolution of pollen identification to maximise the accuracy and usefulness of their data. The utility of pollen and spores as an archive becomes reduced, however, when taxonomic resolution leads to a loss of information (Bush, 2002).

The Poaceae (grass) family exemplifies this problem, as it comprises 11,554 currently accepted species in 759 genera (The Plant List, 2013), which exist across a wide climatic gradient, from Antarctica to tropical lowland rainforest. Yet pollen grains from this family are almost indistinguishable below family level using light microscopy, therefore they are generally not classified below 'Poaceae' by the majority of palynologists (Fægri et al., 1989; Holst et al., 2007; Strömberg, 2011). Consequently Poaceae pollens are essentially a rich yet currently underdeveloped archive ripe for palynological research.

Extensive research over the last four decades has used a variety of tools to determine if the identification of Poaceae pollen to below family level is possible. This analysis has been on individual grains using: (i) surface pattern analysis of images of pollen grains obtained through scanning electron microscopy (SEM) (Andersen and Bertelsen, 1972; Mander et al., 2013; Waikhom et al., 2014), (ii) detailed morphometric analysis considering whole grain and pore morphology (Joly et al., 2007; Schüler and Behling, 2010), and (iii) confocal microscopy of pollen exines (Salih et al., 1997). A success rate in identifying Poaceae pollen to species level of 85.8% has been achieved through SEM (Mander et al., 2013), and this technique has even allowed differentiation of cultivars (Datta and Chaturvedi, 2004). These methods, although successful, are

* Corresponding author at: Palaeoecology & Landscape Ecology, Institute for Biodiversity & Ecosystem Dynamics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands.
E-mail address: W.D.Gosling@uva.nl (W.D. Gosling).

http://dx.doi.org/10.1016/j.revpalbo.2016.08.004

time consuming and require considerable sample preparation, laboratory work, and expertise. Therefore, from a practical perspective, the application of these techniques to palaeoecological questions has not yet occurred.

Fourier Transform Infra-Red spectroscopy (FTIR) has recently been used to differentiate pollen taxonomically, demonstrating it is possible to distinguish between plant orders, and in some cases to species level (Dell'Anna et al., 2009; Pappas et al., 2003; Zimmermann, 2016, 2010). FTIR analysis has also been successfully used in characterising pollen surface compounds (Pummer et al., 2013). FTIR analysis generates absorbance spectra, with bands relating to chemical bonds within specific functional groups. The size, shape and position of these bands provides information about the type of bonds present and their chemical environments, which, in the case of biopolymers such as sporopollenin, can be very complex (de Leeuw et al., 2006; Fraser et al., 2014b, 2011; Watson et al., 2012, 2007). Interpretation of FTIR spectra relies upon knowledge of the type of bonds likely to be present in a substance, and how they might vary. In this study, we treat spectra statistically and use classification algorithms to identify pollen, thus removing the need for in-depth biogeochemical analysis.

Spectra produced by FTIR analysis are affected by a number of operational factors, such as intensity of beam, thickness of sample and thickness of slide (if using a microscope enabled FTIR). Spectra may be noisy if the sample to be scanned (and therefore aperture size) is small, or the material is of poor quality, for instance if pollen grains are degraded. Degradation of the samples used in this study is not expected to be significant, although may be present, as some chemical changes have been observed over short time (hours-days) periods (Zimmermann et al., 2015). Changes in spectra driven by degradation can, however, be accounted for by using statistical processing techniques prior to analysis (Zimmermann and Kohler, 2013). For example, use of algorithms such as Savitsky-Golay smoothing can alleviate noisiness, but potentially remove useful information such as subtleties in shape of bands from spectra if their parameters are not calibrated properly, whereas generating first and second derivatives of spectra may result in degradation of the signal-to-noise ratio (Brown et al., 2000; Zimmermann and Kohler, 2013). The chemical structure of sporopollenin, is known to be very stable over geological time (Fraser et al., 2012) and resistant to diagenetic alteration (Watson et al., 2007; Fraser et al., 2014a), meaning that the interpretation of the fossil record may benefit from the application of this technique.

Here we show that analyses of FTIR spectra from a selection of Poaceae taxa can be used to successfully identify pollen grains. Using a simple nearest neighbour classification algorithm our results have very similar levels of success when compared to much more expensive and labour intensive methods currently deployed, such as SEM (Mander et al., 2013). Therefore, FTIR based analyses raise the possibility of a further exploration of the grass pollen record.

2. Methods

2.1. Sample collection and preparation

A total of twelve grass taxa were analysed from eight subfamilies (Table 1) across the grass phylogeny, as outlined in the latest publication by 'The Grass Phylogeny Working Group' (Grass Phylogeny Working Group II, 2012). The sampling strategy employed ensured a wide phylogenetic spread whilst also enabling analysis of lower-order identification by sub-sampling some subfamilies, such as the Ehrhartoideae. Poaceae pollen was obtained from herbarium specimens at the Royal Botanic Gardens, Kew (London, UK) by dissecting out stamen from individual florets. Where possible, two or more specimens for each species were sampled, and specimens from Ghana or neighbouring tropical West African nations were preferentially sampled, to complement current palaeoecological (fossil pollen) investigations at Lake Bosumtwi, Ghana (Miller and Gosling, 2014), and to reduce large-scale environmental variability as much as possible.

Table 1
Subfamily and species of grass sampled for pollen FTIR.

Subfamily	Species
Bambusoideae	*Bambusa vulgaris* Schrad.
Pharoideae	*Leptaspis zeylanica* Nees ex Steud.
Puelioideae	*Puelia olyriformis* (Franch.) Clayton
Ehrhartoideae	*Oryza sativa* L.
Ehrhartoideae	*Oryza longistaminata* A.Chev. & Roehr.
Ehrhartoideae	*Leersia drepanothrix* Stapf.
Arundinoideae	*Phragmites karka* (Retz.) Trin. Ex Steud
Chloridoideae	*Ctenium elegans* Kunth
Chloridoideae	*Enteropogon macrostachys* (A.Rich.) Munro ex Benth.
Panicoideae	*Pennisetum pedicellatum* Trin.
Panicoideae	*Cenchrus setiger* Vahl
Pooideae	*Triticum aestivum* L.

2.2. Chemical analysis

The pollen was washed in acetone and allowed to air-dry on zinc-selenide slides. Groups of two or more pollen grains clustered together were examined using a Continuum IR-enabled microscope with a 15× reflechromat objective lens and nitrogen-cooled MCT-A detector in transmission mode. The microscope was linked to a Thermo Nicolet Nexus (Thermo Fisher Scientific, Waltham, MA, USA) FTIR bench unit at The Open University. Spectra were averaged over 256 scans per sample, and background scans were taken before each sample to alleviate any atmospheric contributions. Visual inspection of spectra and atmospheric suppression correction was conducted using OMNIC software (Thermo Fisher Scientific, Waltham, MA, USA).

2.3. Data processing and analysis

Average spectra were calculated from multiple replicates for every sample (Fig. 1). These average spectra were inspected visually, and comparisons of selected absorbance bands were compiled (after Steemans et al., 2010) to determine potential structural drivers of the statistical patterns observed (Table 2). The absorbance bands chosen were based on those used by other researchers investigating sporopollenin composition. Bands that do not vary between taxa are omitted from visual inspection; for instance, the broad OH band at 3300 cm^{-1} is omitted, as it is present in all taxa in the same form and thus provides no visually quantifiable classification information. The bands included in the visual inspection and references to papers which have used them in investigations of sporopollenin are as follows: C=C band at 3070 cm^{-1}, (Fraser, 2008); vasCH_2 and vsCH_2 at 2925 cm^{-1} and 2850 cm^{-1} respectively, and vC=O at 1710 cm^{-1}, (Fraser et al., 2012; Watson et al., 2007); vasCH_3 at 2960 cm^{-1}, (Steemans et al., 2010); vsCH_3 at 2890 cm^{-1}, (Fraser, 2008); C=C non-conjugated at 1660 cm^{-1}, (Fraser et al., 2014b; Steemans et al., 2010; Zimmermann and Kohler, 2014), OH at 1630 cm^{-1} (Fraser, 2008); C=C (aromatic ring stretch) at 1500 cm^{-1}, (Fraser et al., 2014b; Lomax et al., 2008; Watson et al., 2007); CH_n (asymmetric bending) at 1460 cm^{-1} and CH_3 (symmetric bending) at 1375 cm^{-1}, (Fraser et al., 2012); C=C or CH_n at 720 cm^{-1}, (Fraser et al., 2012; Zimmermann and Kohler, 2014).

All average spectra were z-score standardised (i.e. standardised to zero mean and unit variance) by finding their mean amplitude, subtracting the mean from the actual values, and dividing by the standard deviation. When no other treatments were applied, these z-score standardised spectra are referred to as 'Unprocessed Spectra' (see Fig. 2 for information on processing). These standardised spectra were not subject to variations in signal amplitude due to variable sample thickness (Duarte et al., 2004; Jardine et al., 2015). Standardisation of spectra (and all other statistical manipulations) were performed in R

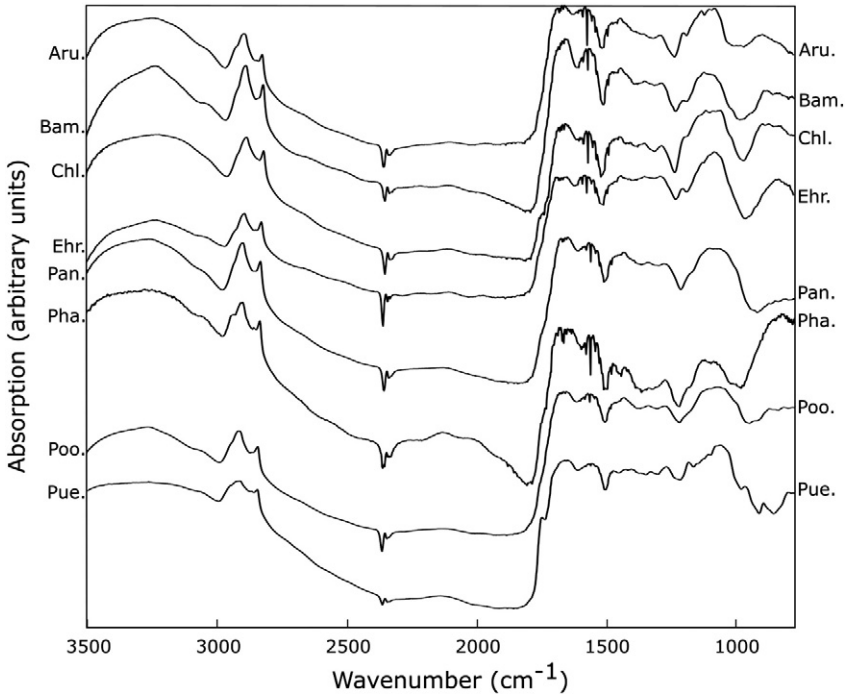

Fig. 1. Average standardised spectra of one representative individual from each subfamily. Aru. = Arundinoideae, Bam. = Bambusoideae, Chl. = Chloridoideae, Ehr. = Ehrhartoideae, Pan. = Panicoideae, Pha. = Pharoideae, Poo. = Pooideae, Pue. = Puelioideae.

v. 3.1.2, using R Studio (RStudio Team, 2012) (see electronic supplementary material for full details of code used).

Further processing of raw data was conducted to investigate various extraneous factors that may have impacted upon the analyses. Such processing involved one or more of the following: (i) truncation, (ii) baseline correction, and/or (iii) atmospheric suppression of the region from 1800 to 2700 cm^{-1} (Fig. 2) to remove the effects of CO_2 and scattering from that region ('Truncated Spectra'). Baseline correction using the R package baseline (Liland and Mevik, 2015) was conducted because some unprocessed spectra exhibited climbing baselines on visual inspection ('Baseline Corrected Spectra'). Atmospheric suppression using the OMNIC atmospheric suppression algorithm and z-score standardisation was performed to assess the impact of water vapour across the spectra ('Atmospheric Suppressed Spectra'). Collectively, this approach allowed us to compensate for possible noise and/or atmospheric effects and to directly compare against the 'unprocessed spectra'. Fig. 2 shows a visual summary of the stages performed during spectral processing.

The final stage of processing was the application of 'spectral pre-processing' *sensu* Zimmermann and Kohler (2013) to both the raw spectra and all of the modified spectra. This final stage of the processing involved generating first and second derivatives of spectra, and applying Savitsky-Golay spectral smoothing (Zimmermann and Kohler, 2013). The Savitsky-Golay smoothing technique applies an algorithm which approximates the spectrum using polynomial least-square fitting to a moving window. Both polynomial order, and window size used affect the resultant spectrum. As different window sizes may be optimal for different regions of the spectrum, and this study aimed to provide a simple tool for pollen identification and thus used the whole spectrum, a window size of 11 was chosen; this falls within the range of optimal values defined by Zimmermann and Kohler (2013). The R package Prospectr was used to perform Savitsky-Golay smoothing and derivation of spectra (Stevens and Ramirez-Lopez, 2013). Principle Component Analysis (PCA) was used to visualise all resultant data from the processing steps detailed above.

Using the R-package, Class (Venables and Ripley, 2002), k-nearest neighbour (k-nn) analyses were performed, using leave-one-out cross-validation to generate identification success rates. This analysis uses a Euclidean distance matrix to classify individual spectra based upon their nearest neighbours in the matrix, and then removing one

Table 2
Chemical bond, wavenumber of absorbance band and presence in difference Poaceae subfamilies, after Steemans et al., 2010. 'sh' indicates a shoulder, '+' a weak absorbance band, '++' a strong absorbance band and '+++' very strong. * indicates where an absorbance band differs in strength within subfamily and '—' indicates an absence of that absorbance band. '~' before wavenumber indicates the position of a band may vary. In style of Steemans et al. (2010).

Group	Wavenumber (cm^{-1})	Arundinoideae	Bambusoideae	Chloridoideae	Ehrhartoideae	Pharoideae	Pooideae	Puelidoideae
C=C	3070	sh	sh	—	sh	sh	sh	sh
vasCH_3	2960	sh	sh	*	sh	+	—	*
vasCH_2	2925	++	++	++	++	++	++	+++
vsCH_3	2890	*	*	*	sh	sh	—	*
vsCH_2	2850	++	+	+	+	+	+	+
vC=O	~1710	sh	sh	sh	sh	sh	—	sh
C=C (non-conjugated)	1660	+	*	*	*	+	+	+
OH	1630	*	*	*	*	+	sh	+
C=C (aromatic ring stretch)	~1500	sh	sh	sh	sh	sh	sh	sh
$CH_{2/3}$	1460	*	+	sh	*	+	sh	*
CH3 (symmetric)	1375	sh	*	+	*	sh	sh	*
C=C (cis-substituted) or CH_n (rocking)	720	*	*	+	+	—	+	—

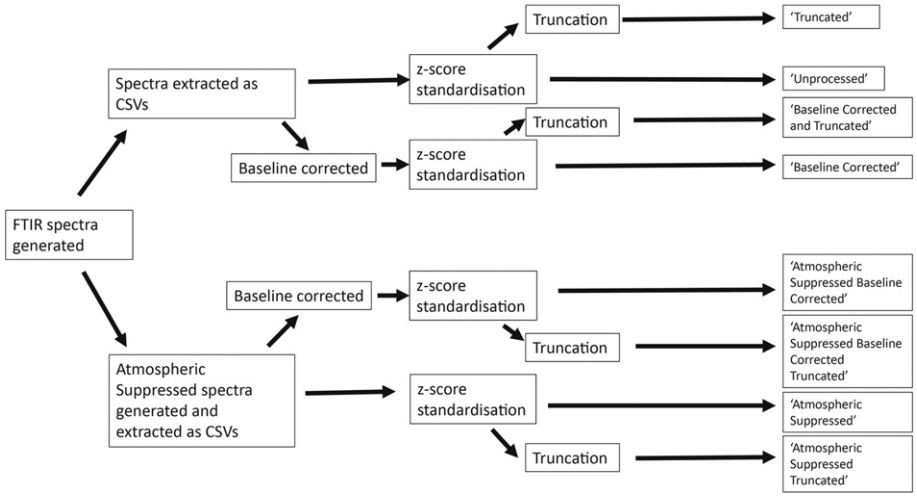

Fig. 2. Flowchart showing the processing steps taken to produce the different pre-processed spectra upon which we based further analyses. Results for all of these treatments may be found in the SOM.

data point (a single spectrum) before re-introducing it for classification. The number of nearest neighbours is instrumental in classification success, as increasing k expands the analysis to include the next closest points, with classification decided by majority vote (the sample is classified based on the most common taxon among its nearest neighbours). So for $k = 1$, only the closest point is included in the analysis, whereas in $k = 5$, the 5 nearest neighbour points are included, with the classification being decided by majority rule, or the most common taxon among those points. To test whether or not the success of the classification algorithm was due to sample size, we used a null model which assigns names to points randomly within the Euclidian distance matrix generated from the original data, and then applies the classification test to the new matrix, repeating the process 999 times. This process allows the distribution of random successful classifications to be compared to the actual classification success.

3. Results

FTIR spectra of pollen from different grass subfamilies display broadly similar characteristics (Fig. 1). Nevertheless, there are small variations in the spectra between subfamilies and lower orders of classification (Table 2) (de Leeuw et al., 2006; Watson et al., 2007). Some absorbance bands are present in, and are similar between, all taxa analysed, such as the C=C stretching band of C=C—H group at 3070 cm^{-1}, the $vasCH_2$ and $vsCH_2$ bands at 2925 cm^{-1} and 2850 cm^{-1} respectively, and the C=C aromatic ring stretch at around 1500 cm^{-1} (Table 2). Other bands are variable both between and within subfamilies, such as the $vasCH_3$ and $vsCH_3$ bands at 2960 cm^{-1} and 2890 cm^{-1}, respectively (Table 2). Absorbance bands that vary between and within subfamilies include the non-conjugated C=C band at 1600 cm^{-1}, the asymmetric bending CH_n at 1460 cm^{-1}, the symmetric bending CH_3 band at 1375 cm^{-1} and the cis-substituted C=C/CH_n rocking band at 720 cm^{-1}. Some absorbance bands exhibit more within-subfamily variation (* in Table 2), such as the band at 1630 cm^{-1}, whereas others display less within-subfamily variation, but differ in presence and shape between subfamilies, as exemplified by the band at 720 cm^{-1}.

PCA of the pre-processed spectra (Fig. 3) shows that groupings are present but complex. Most subfamilies appear broadly grouped together but with significant spread over one or both PCA1 or PCA2. The Ehrhartoideae, for instance, are distributed along PCA1, but more closely grouped along PCA2. The Bambusoideae, however, are less clearly grouped, with a wide spread across both axes. No subfamily is clearly separate from the others, with all exhibiting some degree of overlap with other subfamilies. Neither are there clear phylogenetic groupings, as subfamilies do not cluster according to degree of relatedness (Grass Phylogeny Working Group II, 2012). Loadings plots of the PCA show significant contributions from CO_2 (at around 2400 cm^{-1} on the x-axis). This contribution is not as prevalent in treated spectra (see SOM for details).

The combined treatment of first derivative with Savitsky-Golay smoothing returns the most accurate classification, with successful classifications at subfamily level reaching 80%. Other treatments, including no pre-processing, first and second derivatives, and second derivative with Savistsky-Golay smoothing did not result in as high a proportion of successful classifications (Fig. 4) at any taxonomic level. The maximum success achieved by the algorithm to genus level was 77%, to species level 77%, and individual level 69%. A random permutation test confirmed that the chances of achieving the successful classification proportions observed if no pattern was present were 0.01%.

The success rates at subfamily level for the different treatments outlined in Fig. 2 are presented in Table 3. This shows that unprocessed and truncated spectra have the highest success rates, at 80% and 82% respectively, while all Atmospheric Suppressed Spectra show lower identification success rates. Subfamily level results have been presented here for the sake of brevity and comparison, but for full results, including k values, see SOM.

4. Discussion

From visual (Fig. 1) and PCA (Fig. 3) analyses of the FTIR spectra it was possible to discern differences between subfamilies via analysis of absorbance band strength and nature although, as is demonstrated by Table 2, chemical differences between taxa are complex and subtle. The exact composition of sporopollenin remains enigmatic, due to its inert nature and resistance to decay or chemical degradation, even over very long periods of geological time (Fraser et al., 2012). Sporopollenin is, however, known to comprise aromatic components such as *para*-coumaric acid and ferulic acid, linked by esther linkages and aliphatic components (Boom, 2004; de Leeuw et al., 2006; Fraser et al., 2014b; Watson et al., 2012). Based upon Table 2, and a knowledge of the broad structure of sporopollenin, it is likely that chemical variation between taxa arises from variation in aliphatic chain length (shown by the variation in shape and presence of CH_n bands in Table 2) and degree of saturation (shown by variation in C=C bands). Further chemical analyses via gas chromatography/mass spectrometry (GC/MS) should be able to confirm these tentative conclusions, but as we are primarily concerned with the development of an identification tool, not the chemical structure of sporopollenin, this analysis is not pertinent here.

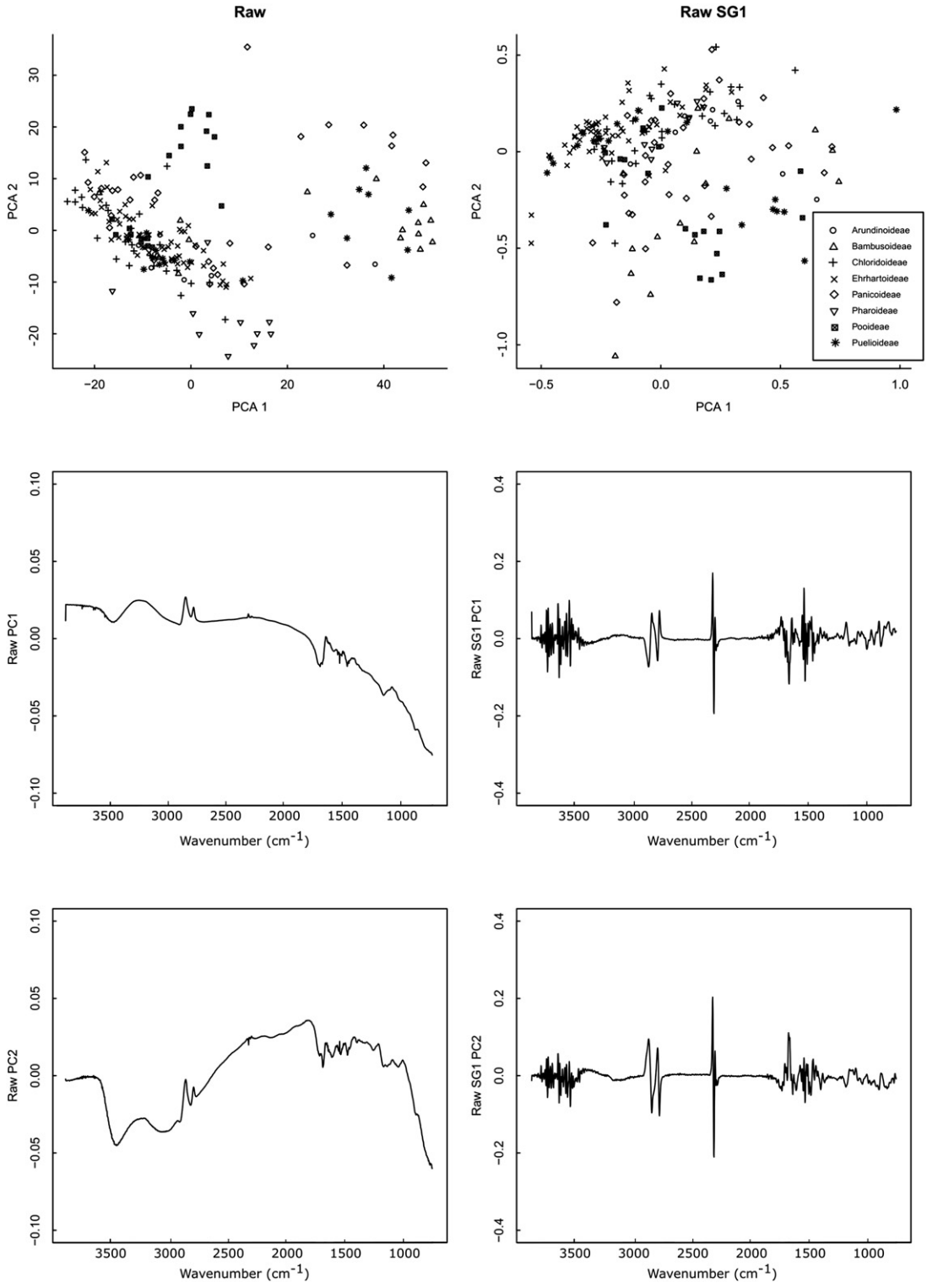

Fig. 3. PCA plots of Raw and Raw SG1 (Savitsky-Golay smoothed, first derivative) spectra and, below them, their loadings plots, illustrating areas of the spectrum which contribute most strongly to the signal.

Some scattering artefacts and noise are evident in the spectra, likely the result of either: (i) the absence of physical atmospheric suppression during sample analysis, and/or (ii) for 8% of samples, sufficiently large groups of pollen grains were not present on the slide, so that some single-grain measurements had to be taken (see SOM for details of numbers of grains scanned for each sample). We consider, however, that these artefacts and noise are random in their nature and would affect all samples equally. A full analysis of the impact of pollen grain number of identification success rate can be found in the supplementary information (SOM).

The first and second derivatives, Savitsky-Golay smoothing and combinations thereof, comprise some of the most commonly used processing techniques used in analysis of FTIR generated spectra (Zimmermann and Kohler, 2013). Using first and second derivatives without Savistky-Golay smoothing on unprocessed spectra shows a decrease in classification success across all taxonomic levels compared to

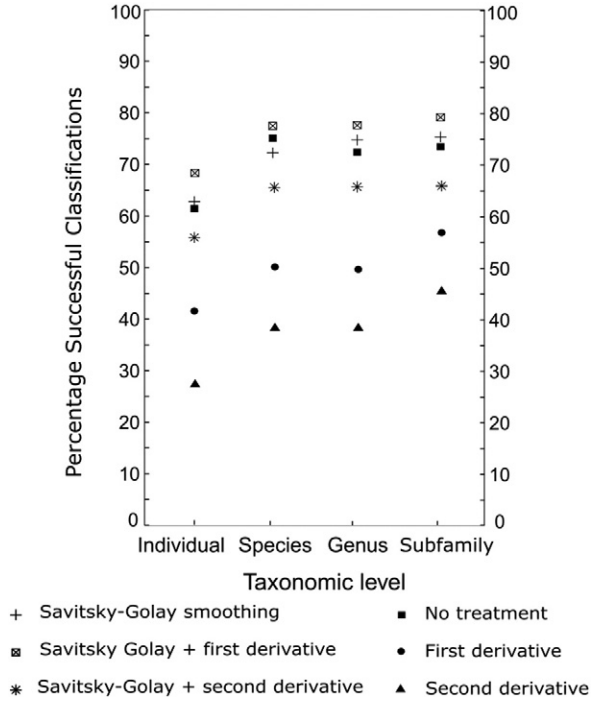

Fig. 4. Plot showing maximum classification success of processing techniques for spectra at a range of taxonomic levels, with values representing most successful number of nearest neighbours for each pre-processing treatment and level. For full k-nn results see electronic supplementary materials (SOM).

raw spectra, from a raw data success rate at species level of 74% to 49% success for first derivative spectra and 37% for second derivative data (see Fig. 3). This is likely due to the exaggeration of the contribution of noise to the analysis of un-smoothed data, which is eliminated by the point-averaging smoothing effect of the Savitsky-Golay smoothing algorithm (Zimmermann and Kohler, 2013).

Truncated Spectra show the highest success rate, at 82%, with Unprocessed Spectra at 80% and Baseline Corrected Spectra at 81% (Table 3). Spectra which were subjected to atmospheric suppression showed large reductions in identification success rates (Table 3). As the Atmospheric Suppression was performed using an algorithm provided by the OMNIC software, we consider that it is therefore likely the suppression algorithm may remove useful information from the grass pollen spectra. This is because FTIR, and the associated OMNIC software are most commonly used to analyse pure chemical samples, whereas the identification of grass pollen relies on subtle chemical variations in sporopollenin. Further, we believe the OMNIC atmospheric suppression may remove useful information from spectra because Truncated Spectra return higher identification success rates, whilst having wavenumbers removed which do not contain useful information, only CO_2 contributions and scattering effects (Mohlenhoff et al., 2005).

Table 3
Table showing maximum success rate at subfamily level for different treatments, and their corresponding pre-processing treatment.

Treatment	Maximum identification success (% at subfamily level)	Pre-processing
Truncated	82	Savitsky Golay Smoothing, first derivative
Unprocessed	80	Savitsky Golay Smoothing, first derivative
Baseline corrected truncated	81	Savitsky Golay Smoothing, first derivative
Baseline corrected	79	Savitsky Golay Smoothing, first derivative
Atmospheric suppressed baseline corrected	61	None
Atmospheric suppressed baseline corrected truncated	60	Savitsky Golay Smoothing
Atmospheric suppressed	46	Savitsky Golay Smoothing, first derivative
Atmospheric suppressed truncated	64	None

Pollen grains used in this study were washed in acetone, which would have removed surface chemicals and traces of insecticide or other contaminants. It is likely, however, that internal material such as intine and possibly cytoplasm remained present in the grains. This means that the signal observed from grass pollen is not that of 'pure' sporopollenin, but, rather, represents the pollen grain as a whole entity (Blokker et al., 2006). For the purposes of modern pollen identification, this does not pose an immediate issue, as it is likely that cytoplasmic and pollen-wall associated compounds contribute to the taxonomic signal detected in this study, and others which have focussed on other taxonomic groups such as the Pinales (Zimmermann, 2010). For fossil pollen samples, or pollen that has been acetolysed, however, some of this information may be lost, either due to chemical alteration during processing (Jardine et al., 2015), or lack of non-sporopollenin pollen components in preserved grains. In these cases, it may be beneficial to use FTIR analysis in conjunction with other, structural methods of distinguishing between similar taxa, such as those described by Sivaguru and Mander (Mander et al., 2013; Sivaguru et al., 2012).

Environmental factors could affect success of classification, as it has been demonstrated that factors such as UVB irradiance (Fraser et al., 2014a; Lomax et al., 2008), year-on-year environmental factors (Zimmermann and Kohler, 2014) and heat stress (Jiang et al., 2015; Lahlali et al., 2014) have an impact upon the chemical composition of pollen grains and spores. Our samples were also taken from a relatively narrow geographical and temporal range (tropical West Africa, and within the 20th Century), compared to the time ranges and geographical ranges dealt with in the fossil record. Therefore, we suggest that the effects of any UVB fluxes are likely to be minimal in this sample set, as UVB changes occur over longer time scales and wider geographical areas (Magri, 2011). The fact that classification success improves from individual to species level, where individuals from the same species have been collected in different areas and different years (SOM) demonstrates a clear taxonomic signal. The absence of a significant improvement between species and subfamily level suggests that although this is a successful classification technique, it is not a phylogenetic one. Although the number of samples varies between taxa, the randomisation algorithm demonstrates that successful classification is not due purely to chance.

These results show that differentiation between Poaceae taxa below family level is possible using the relatively fast and inexpensive method of FTIR microspectroscopy. At subfamily level, it is possible to achieve an 80% classification success rate, and at species level, a 77% classification success rate. The ability to identify grass pollen to subfamily level, or below, allows for a more detailed interrogation of grass pollen record. One specific benefit of increased taxonomic resolution could be the recovery of a more complete ecological picture from the fossil pollen record and the determination of previously hidden plant-climate relationships. For instance, the identification of pollen from the Puelioideae subfamily would indicate a forest-origin for the grass pollen observed, whereas Ehrartoideae might suggest a more open habitat (Grass Phylogeny Working Group II, 2012).

5. Conclusions

We have shown that FTIR analysis followed by spectral and statistical processing has the potential to significantly improve pollen identification within the Poaceae. Our data demonstrate that it is possible to achieve an 80% successful classification rate to subfamily level for Poaceae pollen, which, when applied, will allow new insights into

taxonomic resolution in fossil pollen records. The rapidity and relative low costs of FTIR analyses make this a potentially very useful method for subfamily identification of Poaceae pollen.

Acknowledgements

ACMJ is supported via a Natural Environmental Research Council (NERC) studentship awarded to WDG BHL and WTF (NE/K005294/1). We also thank the Royal Society (RG120535) for funding and the Herbarium of the Royal Botanic Gardens at Kew for access to samples. We would like to thank Luke Mander for discussions about figures and taxonomy, and two reviewers for their detailed and constructive feedback.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.revpalbo.2016.08.004.

References

Andersen, T.S., Bertelsen, F., 1972. Scanning electron microscope studies of pollen of cereals and other grasses. Grana 12, 79–86.
Blokker, P., Boelen, P., Broekman, R., Rozema, J., 2006. The occurrence of p-coumaric acid and ferulic acid in fossil plant materials and their use as UV-proxy. Plant Ecol. 182, 197–207. http://dx.doi.org/10.1007/s11258-005-9026-y.
Boom, A., 2004. A Geochemical Study of Lacustrine Sediments: Towards Palaeo-climatic Reconstructions of High Andean Biomes in Colombia (PhD Thesis). University of Amsterdam, The Netherlands http://hdl.handle.net/11245/1.225858.
Brown, C.D., Vega-Montoto, L., Wentzell, P.D., 2000. Derivative preprocessing and optimal corrections for baseline drift in multivariate calibration. Appl. Spectrosc. 54, 1055–1068. http://dx.doi.org/10.1366/0003702001950571.
Bush, M.B., 2002. On the interpretation of fossil Poaceae pollen in the lowland humid neotropics. Palaeogeogr. Palaeoclimatol. Palaeoecol. 177, 5–17. http://dx.doi.org/10.1016/S0031-0182(01)00348-0.
Datta, K., Chaturvedi, M., 2004. Pollen morphology of Basmati cultivars (*Oryza sativa* race Indica) – exine surface ultrastructure. Grana 43, 89–93. http://dx.doi.org/10.1080/00173130310017391.
de Leeuw, J.W., Versteegh, G.J.M., van Bergen, P.F., 2006. Biomacromolecules of algae and plants and their fossil analogues. Plant Ecol. 182, 209–233. http://dx.doi.org/10.1007/s11258-005-9027-x.
Dell'Anna, R., Lazzeri, P., Frisanco, M., Monti, F., Campeggi, F.M., Gottardini, E., Bersani, M., 2009. Pollen discrimination and classification by Fourier transform infrared (FT-IR) microspectroscopy and machine learning. Anal. Bioanal. Chem. 394, 1443–1452. http://dx.doi.org/10.1007/s00216-009-2794-9.
Duarte, I.F., Barros, A., Almeida, C., Spraul, M., Gil, A.M., 2004. Multivariate analysis of NMR and FTIR data as a potential tool for the quality control of beer. J. Agric. Food Chem. 52, 1031–1038. http://dx.doi.org/10.1021/jf030659z.
Fægri, K., Iversen, J., Krzywinski, K., Kaland, P.E., 1989. Textbook of Pollen Analysis. 4th ed. Wiley, Chichester.
Fraser, W.T., 2008. Evaluation of Spore Wall Chemistry as a Novel Biochemical Proxy for UV-B Radiation (PhD Thesis). The Open University, UK http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.494651.
Fraser, W.T., Sephton, M.A., Watson, J.S., Self, S., Lomax, B.H., James, D.I., Wellman, C.H., Callaghan, T.V., Beerling, D.J., 2011. UV-B absorbing pigments in spores: biochemical responses to shade in a high-latitude birch forest and implications for sporopollenin-based proxies of past environmental change. Polar Res. 30, 8312. http://dx.doi.org/10.3402/polar.v30i0.8312.
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 Phytol. 196, 397–401. http://dx.doi.org/10.1111/j.1469-8137.2012.04301.x.
Fraser, W.T., Lomax, B.H., Jardine, P.E., Gosling, W.D., Sephton, M.A., 2014a. Pollen and spores as a passive monitor of ultraviolet radiation. Paleoecology 2, 12. http://dx.doi.org/10.3389/fevo.2014.00012.
Fraser, W.T., Watson, J.S., Sephton, M.A., Lomax, B.H., Harrington, G., Gosling, W.D., Self, S., 2014b. Changes in spore chemistry and appearance with increasing maturity. Rev. Palaeobot. Palynol. 201, 41–46. http://dx.doi.org/10.1016/j.revpalbo.2013.11.001.
Germeraad, J.H., Hopping, C.A., Muller, J., 1968. Palynology of Tertiary sediments from tropical areas. Palinol. Tert. Sediments Trop. Areas 6, 189–348. http://dx.doi.org/10.1016/0034-6667(68)90051-1.
Grass Phylogeny Working Group II, 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. New Phytol. 193, 304–312. http://dx.doi.org/10.1111/j.1469-8137.2011.03972.x.
Herrero, B., Valencia-Barrera, R.M., San Martin, R., Pando, V., 2002. Characterization of honeys by melissopalynology and statistical analysis. Can. J. Plant Sci. 82, 75–82. http://dx.doi.org/10.4141/P00-187.
Holst, I., Moreno, J.E., Piperno, D.R., 2007. Identification of teosinte, maize, and *Tripsacum* in Mesoamerica by using pollen, starch grains, and phytoliths. Proc. Natl. Acad. Sci. 104, 17608–17613. http://dx.doi.org/10.1073/pnas.0708736104.
Horrocks, M., Coulson, S.A., Walsh, K.A.J., 1998. Forensic palynology: variation in the pollen content of soil surface samples. J. Forensic Sci. 43, 320–323. http://dx.doi.org/10.1520/JFS16139J.
Jardine, P.E., Fraser, W.T., Lomax, B.H., Gosling, W.D., 2015. The impact of oxidation on spore and pollen chemistry. J. Micropalaeontol. 34, 139–149. http://dx.doi.org/10.1144/jmpaleo2014-022.
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. Plant Cell Environ. 38, 2387–2397. http://dx.doi.org/10.1111/pce.12589.
Joly, C., Barillé, L., Barreau, M., Mancheron, A., Visset, L., 2007. Grain and annulus diameter as criteria for distinguishing pollen grains of cereals from wild grasses. Rev. Palaeobot. Palynol. 146, 221–233. http://dx.doi.org/10.1016/j.revpalbo.2007.04.003.
Lahlali, R., Jiang, Y., Kumar, S., Karunakaran, C., Liu, X., Borondics, F., Hallin, E., Bueckert, R., 2014. ATR–FTIR spectroscopy reveals involvement of lipids and proteins of intact pea pollen grains to heat stress tolerance. Front. Plant Sci. 5, 747. http://dx.doi.org/10.3389/fpls.2014.00747.
Liland, K.H., Mevik, B.-H., 2015. Baseline: Baseline Correction of Spectra. R Package Version 1.2-1. https://CRAN.R-project.org/package=baseline.
Lomax, B.H., Fraser, W.T., Sephton, M.A., Callaghan, T.V., Self, S., Harfoot, M., Pyle, J.A., Wellman, C.H., Beerling, D.J., 2008. Plant spore walls as a record of long-term changes in ultraviolet-B radiation. Nat. Geosci. 1, 592–596. http://dx.doi.org/10.1038/ngeo278.
Looy, C.V., Twitchett, R.J., Dilcher, D.L., Van Konijnenburg-Van Cittert, J.H.A., Visscher, H., 2001. Life in the end-Permian dead zone. Proc. Natl. Acad. Sci. U. S. A. 98, 7879–7883. http://dx.doi.org/10.1073/pnas.131218098.
Lupia, R., Lidgard, S., Crane, P.R., 1999. Comparing palynological abundance and diversity: implications for biotic replacement during the Cretaceous angiosperm radiation. Paleobiology 25, 305–340. http://dx.doi.org/10.1017/S009483730002131X.
Magri, D., 2011. Past UV-B flux from fossil pollen: prospects for climate, environment and evolution. New Phytol. 192, 310–312. http://dx.doi.org/10.1111/j.1469-8137.2011.03864.x.
Mander, L., Punyasena, S.W., 2014. On the taxonomic resolution of pollen and spore records of Earth's vegetation. Int. J. Plant Sci. 175, 931–945. http://dx.doi.org/10.1086/677680.
Mander, L., Li, M., Mio, W., Fowlkes, C.C., Punyasena, S.W., 2013. Classification of grass pollen through the quantitative analysis of surface ornamentation and texture. Proc. R. Soc. B Biol. Sci. 280, 20131905. http://dx.doi.org/10.1098/rspb.2013.1905.
Martin, P., 2005. Importance of melissopalynology for beekeeping and trade. Bee World 86, 75–76. http://dx.doi.org/10.1080/0005772X.2005.11417317.
Mildenhall, D.C., Wiltshire, P.E.J., Bryant, V.M., 2006. Forensic palynology: why do it and how it works. Forensic Sci. Int. 163, 163–172. http://dx.doi.org/10.1016/j.forsciint.2006.07.012.
Miller, C.S., Gosling, W.D., 2014. Quaternary forest associations in lowland tropical West Africa. Quat. Sci. Rev. 84, 7–25. http://dx.doi.org/10.1016/j.quascirev.2013.10.027.
Mohlenhoff, B., Romeo, M., Diem, M., Wood, B.R., 2005. Mie-type scattering and non-beer-Lambert absorption behavior of human cells in infrared microspectroscopy. Biophys. J. 88, 3635–3640. http://dx.doi.org/10.1529/biophysj.104.057950.
Pappas, C.S., Tarantilis, P.A., Harizanis, P.C., Polissiou, M.G., 2003. New method for pollen identification by FT-IR spectroscopy. Appl. Spectrosc. 57, 23–27. http://dx.doi.org/10.1366/000370203321165160.
Pummer, B.G., Bauer, H., Bernardi, J., Chazallon, B., Facq, S., Lendl, B., Whitmore, K., Grothe, H., 2013. Chemistry and morphology of dried-up pollen suspension residues. J. Raman Spectrosc. 44, 1654–1658. http://dx.doi.org/10.1002/jrs.4395.
RStudio Team, 2012. RStudio: Intergrated development for R. RStudio, Inc., Boston, MA. http://www.rstudio.com.
Rubinstein, C.V., Gerrienne, P., de la Puente, G.S., Astini, R.A., Steemans, P., 2010. Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana). New Phytol. 188, 365–369. http://dx.doi.org/10.1111/j.1469-8137.2010.03433.x.
Salih, A., Jones, A.S., Bass, D., Cox, G., 1997. Confocal imaging of exine as a tool for grass pollen analysis. Grana 36, 215–224. http://dx.doi.org/10.1080/00173139709362610.
Schüler, L., Behling, H., 2010. Poaceae pollen grain size as a tool to distinguish past grasslands in South America: a new methodological approach. Veg. Hist. Archaeobotany 20, 83–96. http://dx.doi.org/10.1007/s00334-010-0265-z.
Sivaguru, M., Mander, L., Fried, G., Punyasena, S.W., 2012. Capturing the surface texture and shape of pollen: a comparison of microscopy techniques. PLoS One 7, e39129. http://dx.doi.org/10.1371/journal.pone.0039129.
Steemans, P., Lepot, K., Marshall, C.P., Le Hérissé, A., Javaux, E.J., 2010. FTIR characterisation of the chemical composition of Silurian miospores (cryptospores and trilete spores) from Gotland, Sweden. Rev. Palaeobot. Palynol. 162, 577–590. http://dx.doi.org/10.1016/j.revpalbo.2010.07.006.
Stevens, A., Ramirez-Lopez, L., 2013. An Introduction to the Prospectr Package: R Package Vignette R Package Version 0.1.3. https://CRAN.R-project.org/package=prospectr.
Strömberg, C.A.E., 2011. Evolution of grasses and grassland ecosystems. Annu. Rev. Earth Planet. Sci. 39, 517–544. http://dx.doi.org/10.1146/annurev-earth-040809-152402.
The Plant List, 2013. Version 1.1. Published on the Internet. http://www.theplantlist.org/ (accessed 14 February 2015).
Tschudy, R.H., Pillmore, C.L., Orth, C.J., Gilmore, J.S., Knight, J.D., 1984. Disruption of the terrestrial plant ecosystem at the Cretaceous-Tertiary boundary, Western Interior. Science 225, 1030–1032. http://dx.doi.org/10.1126/science.225.4666.1030.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Statistics and Computing. Springer New York, New York, NY https://CRAN.R-project.org/package=class.
Waikhom, S., Louis, B., Roy, P., Singh, W., Bharwaj, P., Talukdar, N., 2014. Scanning electron microscopy of pollen structure throws light on resolving Bambusa- Dendrocalamus complex: bamboo flowering evidence. Plant Syst. Evol. 300, 1261–1268. http://dx.doi.org/10.1007/s00606-013-0959-7.

Watson, J.S., Sephton, M.A., Sephton, S.V., Self, S., Fraser, W.T., Lomax, B.H., Gilmour, I., Wellman, C.H., Beerling, D.J., 2007. Rapid determination of spore chemistry using thermochemolysis gas chromatography–mass spectrometry and micro-Fourier transform infrared spectroscopy. Photochem. Photobiol. Sci. 6, 689–694. http://dx.doi.org/10.1039/B617794H.

Watson, J.S., Fraser, W.T., Sephton, M.A., 2012. Formation of a polyalkyl macromolecule from the hydrolysable component within sporopollenin during heating/pyrolysis experiments with *Lycopodium* spores. J. Anal. Appl. Pyrolysis 95, 138–144. http://dx.doi.org/10.1016/j.jaap.2012.01.019.

Zimmermann, B., 2010. Characterization of pollen by vibrational spectroscopy. Appl. Spectrosc. 64, 1364–1373. http://dx.doi.org/10.1366/000370210793561664.

Zimmermann, B., 2016. Analysis of allergenic pollen by FTIR microspectroscopy. Anal. Chem. 88, 803–811. http://dx.doi.org/10.1021/acs.analchem.5b03208.

Zimmermann, B., Kohler, A., 2013. Optimizing Savitzky-Golay parameters for improving spectral resolution and quantification in infrared spectroscopy. Appl. Spectrosc. 67, 892–902. http://dx.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, e95417. http://dx.doi.org/10.1371/journal.pone.0095417.

Zimmermann, B., Tkalčec, Z., Mešić, A., Kohler, A., 2015. Characterizing aeroallergens by infrared spectroscopy of fungal spores and pollen. PLoS One 10, e0124240. http://dx.doi.org/10.1371/journal.pone.0124240.