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