

Sporopollenin chemistry and its durability in the geological record: an integration of extant and fossil chemical data across the seed plants

Phillip Jardine, Carina Hoorn, Maxine A.M. Beer, Natasha Barbolini, Amber Woutersen, Giovanni Bogota-angel, William Gosling, Wesley Fraser, Barry Lomax, Huasheng Huang, et al.

▶ **To cite this version:**

Phillip Jardine, Carina Hoorn, Maxine A.M. Beer, Natasha Barbolini, Amber Woutersen, et al.. Sporopollenin chemistry and its durability in the geological record: an integration of extant and fossil chemical data across the seed plants. Palaeontology, 2021, 64 (2), pp.285-305. 10.1111/pala.12523 . insu-03146419

HAL Id: insu-03146419
https://insu.hal.science/insu-03146419

Submitted on 11 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License

Palaeontology
The Palaeontological Association
www.palass.org

[Palaeontology, Vol. 64, Part 2, 2021, pp. 285–305]

SPOROPOLLENIN CHEMISTRY AND ITS DURABILITY IN THE GEOLOGICAL RECORD: AN INTEGRATION OF EXTANT AND FOSSIL CHEMICAL DATA ACROSS THE SEED PLANTS

by PHILLIP E. JARDINE[1,*], CARINA HOORN[2,*], MAXINE A.M. BEER[2], NATASHA BARBOLINI[2,3], AMBER WOUTERSEN[2], GIOVANNI BOGOTA-ANGEL[2,4], WILLIAM D. GOSLING[2], WESLEY T. FRASER[5], BARRY H. LOMAX[6], HUASHENG HUANG[2], MATTEO SCIUMBATA[2], HUAJIE HE[7] *and* GUILLAUME DUPONT-NIVET[8,9]

[1]Institute of Geology & Palaeontology, University of Münster, 48149, Münster, Germany; jardine@uni-muenster.de
[2]Department of Ecosystem & Landscape Dynamics, Institute for Biodiversity & Ecosystem Dynamics (IBED), University of Amsterdam, 1090 GE, Amsterdam, The Netherlands
[3]Department of Ecology, Environment & Plant Sciences, & Bolin Centre for Climate Research, Stockholm University, SE-106 91, Stockholm, Sweden
[4]Facultad del Medio Ambiente y Recursos Naturales, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia
[5]Geography, Department of Social Sciences, Oxford Brookes University, Oxford, OX3 0BP, UK
[6]Agriculture & Environmental Science, University of Nottingham, Sutton Bonington Campus, Leicestershire, LE12 5RD, UK
[7]Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan China
[8]Institute of Geosciences, University of Potsdam, 14476, Potsdam, Germany
[9]Univ Rennes, CNRS, Géosciences Rennes, UMR 6118, 35000, Rennes, France
*These authors contributed equally to this work.

Typescript received 16 July 2020; accepted in revised form 4 December 2020

Abstract: Sporopollenin is a highly resistant biopolymer that forms the outer wall of pollen and spores (sporomorphs). Recent research into sporopollenin chemistry has opened up a range of new avenues for palynological research, including chemotaxonomic classification of morphologically cryptic taxa. However, there have been limited attempts to directly integrate extant and fossil sporopollenin chemical data. Of particular importance is the impact of sample processing to isolate sporopollenin from fresh sporomorphs, and the extent of chemical changes that occur once sporomorphs enter the geological record. Here, we explore these issues using Fourier transform infrared (FTIR) microspectroscopy data from extant and fossil grass, *Nitraria* (a steppe plant), and conifer pollen. We show a 98% classification success rate at subfamily level with extant grass pollen, demonstrating a strong taxonomic signature in isolated sporopollenin. However, we also reveal substantial chemical differences between extant and fossil sporopollenin, which can be tied to both early diagenetic changes acting on the sporomorphs and chemical derivates of sample processing. Our results demonstrate that directly integrating extant and late Quaternary chemical data should be tractable as long as comparable sample processing routines are maintained. Consistent differences between extant and deeper time sporomorphs, however, suggests that classifying fossil specimens using extant training sets will be challenging. Further work is therefore required to understand and simulate the effects of diagenetic processes on sporopollenin chemistry.

Key words: sporopollenin, pollen, chemotaxonomy, seed plants, diagenesis, Fourier transform infrared (FTIR) microspectroscopy.

Pollen and spores (collectively sporomorphs) are a key component of research into plant evolution and palaeoecology (Gray & Boucot 1971; Crane & Lidgard 1989; Wellman *et al.* 2003; Jaramillo *et al.* 2006; Harrington & Jaramillo 2007; Traverse 2007; Slater *et al.* 2016; Jardine *et al.* 2018). They also provide tools to understand past climates and environmental change (Collinson *et al.* 1981; Brinkhuis *et al.* 2006; Hoorn *et al.* 2012; Pross *et al.* 2012) and constitute biostratigraphical indicators to correlate and date sedimentary sequences (Frederiksen 1998; Barbolini *et al.* 2018). In recent years, sporomorph research has expanded beyond purely using morphology to classify specimens and link them to parent plant groups, by incorporating information on sporomorph

© 2021 The Authors.
doi: 10.1111/pala.12523

Palaeontology published by John Wiley & Sons Ltd on behalf of The Palaeontological Association.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

chemical signature. This has yielded information in three key areas: (1) the chemical composition of sporopollenin, the biopolymer that makes up the outer wall (exine) of sporomorphs and gives them an exceptionally high preservation potential in the geological record (Blokker *et al.* 2006; de Leeuw *et al.* 2006; Watson *et al.* 2007; Fraser *et al.* 2012; Watson *et al.* 2012; Fraser *et al.* 2014*a*; Jardine *et al.* 2015; Li *et al.* 2019; Nierop *et al.* 2019); (2) taxonomic controls on sporomorph and sporopollenin chemistry and the possibility of using this information to classify sporomorph specimens (Pappas *et al.* 2003; Schulte *et al.* 2008; Dell'Anna *et al.* 2009; Steemans *et al.* 2010; Zimmermann 2010; Zimmermann & Kohler 2014; Julier *et al.* 2016; Zimmermann *et al.* 2016; Woutersen *et al.* 2018; Jardine *et al.* 2019); and (3) climatic controls on sporomorph chemistry, including the production of UVB absorbing compounds (UACs) which provide the basis for a palaeo-UVB proxy (Rozema *et al.* 1999, 2001*a*, *b*, 2002, 2009; Blokker *et al.* 2005, 2006; Watson *et al.* 2007; Lomax *et al.* 2008, 2012; Fraser *et al.* 2011, 2014*b*; Willis *et al.* 2011; Lomax & Fraser 2015; Jardine *et al.* 2016, 2017, 2020*a*; Seddon *et al.* 2017, 2019; Bell *et al.* 2018).

Despite the high potential of these methods, there are a series of unresolved issues linked to the feasibility of integrating chemical datasets derived from extant and fossil sporomorphs. (Note that here we use the term 'fossil' to denote sporomorphs that have been recovered from the geological record, rather than inferring any fossilization processes *per se*.) These issues relate mostly to how sporopollenin chemistry changes as a function of diagenesis within natural sedimentary systems, and also to the laboratory methods used to isolate sporopollenin from fresh sporomorphs. Resolving these two issues is essential if fossil material is to be classified or analysed in the context of training sets or calibrations based on extant sporomorphs.

Sporopollenin is generally thought to be robust to changes brought about by post-burial diagenetic processes. If this is correct, then sporopollenin that has been isolated from fresh material could be directly related to well-preserved fossil specimens. Fraser *et al.* (2012), for example, demonstrated that the chemical signature from exceptionally preserved Pennsylvanian (~310 Ma) lycopsid megaspores was highly similar to that of extant *Lycopodium* spores. Simulated oxidation and thermal maturation experiments have also suggested a broad taphonomic and diagenetic window for preservation of the sporopollenin chemical signature (Yule *et al.* 2000; Fraser *et al.* 2014*a*; Jardine *et al.* 2015). However, considerable changes in sporopollenin chemistry have been shown to occur with increased thermal maturation once certain temperature thresholds are reached (Watson *et al.* 2012; Fraser *et al.* 2014*a*; Bernard *et al.* 2015). It has been suggested that sporopollenin includes a hydrolysable 'labile' component comprising ester bound *n*-carboxylic acid monomers, which repolymerize *in situ* during diagenesis to produce a recalcitrant polyalkyl 'geopolymer' with a chemical signature that is distinct from the original sporopollenin biopolymer (Watson *et al.* 2012). In addition, the diagenetic products of plant waxes, membrane lipids or sporomorph lipids may also be incorporated during sporopollenin repolymerization (de Leeuw *et al.* 2006; although see also Gupta *et al.* 2007). While such changes to the sporopollenin chemical signature would hamper the direct integration of extant and fossil chemical datasets, relating rates and thresholds of change in experimental settings to those in natural geological systems is difficult, and so it is not known at which combination of temperature, pressure and time repolymerization occurs (Watson *et al.* 2012; Fraser *et al.* 2014*a*; Bernard *et al.* 2015).

Laboratory processing methods may also influence sporopollenin chemistry. Sporomorphs collected from living plants or herbarium specimens require processing to remove the external labile compounds (proteins, lipids and carbohydrates) and isolate the sporopollenin wall (Traverse 2007; Jardine *et al.* 2015). It is currently not well understood how processing procedures may influence sporopollenin chemistry, and therefore whether 'pure' sporopollenin is left over or some modified form of it (Domínguez *et al.* 1998; Loader & Hemming 2000; Nierop *et al.* 2019). Acetolysis is a standard laboratory processing method used in palynological research to isolate sporopollenin. While UAC relative abundances are recoverable from acetolysed sporomorphs (Jardine *et al.* 2016, 2017), the acetolysis procedure may leave an imprint on the chemical signature of the sporomorphs (Domínguez *et al.* 1998; Loader & Hemming 2000; Jardine *et al.* 2015). This imprint would not be present in non-acetolysed material, as is often the case in pre-Quaternary (i.e. Neogene and older) palynological samples where acetolysis is not part of the routine processing procedure (Traverse 2007).

Most previous studies of sporomorph and sporopollenin chemistry have been carried out using a small range of samples, taxa and/or time periods. These studies have provided a wealth of information on the controls on the chemical signature of sporopollenin, but the restricted nature of their scope makes wider generalizations difficult. This is a particular problem for chemotaxonomic analysis of fossil datasets, since most research in this area to date has focused exclusively on freshly harvested or herbarium material, without additional isolation of the sporopollenin (Pappas *et al.* 2003; Schulte *et al.* 2008; Dell'Anna *et al.* 2009; Zimmermann 2010; Zimmermann & Kohler 2014; Julier *et al.* 2016; Zimmermann *et al.* 2016; Jardine *et al.* 2019; although see Woutersen *et al.* 2018; Nierop *et al.* 2019). Therefore, while freshly harvested sporomorphs are known to contain a useable

taxonomic signature in their chemical make-up, it is still not clear how much of this signal is driven by the external labile components and how much taxonomic information is recorded within the sporopollenin biopolymer. If there is only a weak taxonomic signal present in sporopollenin itself, attempts to apply chemotaxonomy to the fossil record will have limited success, even if diagenetic changes in sporopollenin are subtle enough for extant and fossil datasets to be integrated successfully.

Here, we explore these issues with a large dataset incorporating a range of taxa from different time periods, using Fourier transform infrared (FTIR) microspectroscopy to track changes in pollen chemistry. These datasets have been combined to tackle key research questions in the development of chemopalynology. The first question focuses on testing whether chemotaxonomic information is recorded within sporopollenin. In testing this concept, we have focused on the grasses (Poaceae) as they present a classic problem in palynology: despite being highly diverse (> 12 000 species) and ecologically dominant in savannah and steppe environments, grass pollen is morphologically highly similar and challenging to classify below the family level (Bush 2002; Strömberg 2011; Mander *et al.* 2013; Julier *et al.* 2016; Jardine *et al.* 2019). Valuable palaeoecological and palaeoenvironmental information, such as relative contributions from aquatic versus terrestrial grasses, or domesticated versus wild, is therefore often lost in palynological studies (Bush 2002; Jardine *et al.* 2019). Our main dataset comprises extant grass pollen collected from 50 South American species in 9 subfamilies, which we use to quantify classification potential in acetolysed sporopollenin. Additionally, we use this dataset to test whether the photosynthetic pathways in grasses can be determined in the broader FTIR fingerprint. Being able to reliably chemotype photosynthetic pathways of fossil pollen (C_3 vs C_4) would provide much needed information on the evolution of C_3 and C_4 grasses, and their relative dominance across different habitats (Nelson *et al.* 2006, 2008; Strömberg 2011).

The second question explores whether extant and fossil chemical datasets can be integrated to enable chemotaxonomic classification, by determining the impact of diagenetic processes on sporopollenin preservation within natural sedimentary settings. To answer this question, we assembled a dataset of Miocene to Pleistocene (~18 to ~0.5 Ma) grass pollen from western Amazonia (Hoorn 1994; Vink 2012; Bermúdez *et al.* 2017) and the Amazon Fan (Hoorn *et al.* 2017), to provide a direct comparison with the extant grasses examined to deliver on the first question. We also use late Pleistocene (~500 to 13 ka) grass pollen data from Lake Bosumtwi in Ghana (Miller & Gosling 2014; Jardine *et al.* 2016) to examine chemical changes on a 10^4 to 10^5 year timescale, providing a temporal bridge between the extant and fossil grass datasets.

Finally, we test the generality of the results obtained with grasses by using extant and Palaeogene chemical data from *Nitraria* (Woutersen *et al.* 2018) and conifer (Pinaceae and Podocarpaceae) pollen. These datasets provide additional extant–fossil comparisons across a broad phylogenetic range.

MATERIAL AND METHOD

Dataset descriptions

To deliver information on these three research topics, we assembled eight distinct yet complementary pollen datasets. These datasets, described below and organized by topic, comprise extant material harvested from field and herbaria collections, and Pleistocene to Paleocene samples retrieved from rock outcrops. An overview of the datasets is provided in Table 1.

Chemotaxonomic signals from sporopollenin. Dataset 1: Extant South American grasses. Pollen from 50 Poaceae species was either taken from the residue collection at the University of Amsterdam or sampled from the National Herbarium of the Netherlands at Naturalis Biodiversity Center, Leiden (Table 2; Jardine *et al.* 2020*b*, fig. S1, table S1). The pollen was extracted and the sporopollenin isolated using standard acetolysis for 6 min.

Diagenesis in non-exceptional settings. Dataset 2: Early Miocene to Pliocene grass pollen (~18 to ~5 Ma) from western Amazonia. The data come from nine samples from six localities in Colombia (Los Chorros and Santa Sofia), Peru (Madre de Dios, Pebas and Santa Teresa) and Venezuela (the Maracaibo Basin) (Jardine *et al.* 2020*b*, fig. S1). The pollen count data from the samples are documented in Hoorn (1994), Vink (2012) and Bermúdez *et al.* (2017), which contain additional stratigraphic and locality information. All sample processing was carried out at the University of Amsterdam. Eight samples are clastics (very fine sands, silts and clays) and were processed using sodium pyrophosphate in a 10% solution with H_2O ($Na_4P_2O_7{\cdot}10(H_20)$), while the one lignite sample (sample 16788 from Los Chorros) was also oxidized with Schulze reagent ($2HNO_3$, 60% : K_2C10_3, 7%). For all samples the organic/inorganic fraction was separated with bromoform with density 2.0 gs/cm^3.

Dataset 3: Late Miocene to Pleistocene (5.85–0.54 Ma) grass pollen from the Amazon Fan. The data come from eight samples from a hydrocarbon exploration well (Well 2) situated on the upper slope of the continental shelf beyond the Amazon River delta (latitude 3°2′59″N, longitude 47°44′46″W; Jardine *et al.* 2020*b*, fig. S1) (Hoorn *et al.* 2017). The samples span a depth of 1644–

TABLE 1. Overview of the sporomorph datasets used in this study.

Dataset	Taxon	Age and type	Locality	Treatment
1	Poaceae	Extant (herbarium specimens)	Amazonia/Andes	Acetolysis
2	Poaceae	Fossil (early Miocene to Pliocene)	Western Amazonia	$Na_4P_2O_7$ and heavy liquid separation (clastics) plus Schultze reagent (lignite)
3	Poaceae	Fossil (late Miocene to Pleistocene)	Amazon Fan	HCl, $Na_4P_2O_7$, acetolysis and heavy liquid separation
4	Poaceae	(Sub)fossil (late Pleistocene)	Lake Bosumtwi, Ghana	HCl, HF, KOH and acetolysis (splits of eight samples processed without acetolysis)
5	*Nitraria*	Extant (herbarium specimens)	Global	Acetolysis
6	*Nitraria*	Fossil (Eocene)	Tarim and Xining basins, China	HCl and HF, then heavy liquid separation for Xining sample only
7	Conifers	Extant (live plants)	Yunnan and Hunan provinces, China	Acetolysis
8	Conifers	Fossil (Paleocene)	Nangqian Basin, China	HCl, HF and heavy liquid separation

TABLE 2. Number of genera and species, and C_3/C_4 species, for each subfamily in the extant grass dataset (Dataset 1).

Subfamily	Number of genera/ species in dataset	Number of C_3/C_4 species in dataset
Anomochlooideae	1/1	1/0
Aristidoideae	1/1	0/1
Bambusoideae	1/3	3/0
Chloridoideae	3/5	0/5
Danthonioideae	1/2	2/0
Oryzoideae	1/1	1/0
Panicoideae	15/24	6/18
Pharoideae	1/2	2/0
Pooideae	8/11	11/0

3793 mbdf (metres below drill floor), which corresponds to a sediment depth of 865–3014 m once water depth (754 m) and the height of the drilling platform (25 m) are taken into account. The count data from the samples were published in Hoorn *et al.* (2017) and Kirschner & Hoorn (2020). The samples were processed at the University of Amsterdam with hydrochloric acid (HCl), sodium pyrophosphate, acetolysis and bromoform separation (Hoorn *et al.* 2017).

Dataset 4: Late Pleistocene (520 to 13 ka) grass pollen from Lake Bosumtwi in Ghana. Count data from the majority of samples were published in Miller & Gosling (2014), and FTIR-derived UAC data from the upper 130 kyr of the Bosumtwi record were published in Jardine *et al.* (2016). The samples were processed at the Open University using HCl, HF, KOH and acetolysis, with eight samples processed both with and without acetolysis to study the impact of this process on sporopollenin chemistry (Jardine *et al.* 2016).

Generality across taxa. Dataset 5: Extant *Nitraria* pollen. Pollen from six different species, collected from various herbaria and collections. The sporopollenin was isolated using standard acetolysis. The FTIR data were previously published in Woutersen *et al.* (2018), which provides further information on the samples, as well as detailed morphological and chemical analyses of these species.

Dataset 6: Middle–late Eocene *Nitraria* pollen from the Tarim and Xining basins north of the Tibetan Plateau, which have been assigned to the form-genera *Nitrariadites* and *Nitraripollis* (Hoorn *et al.* 2012). The Tarim sample (sample MS-BC22) comes from the late Eocene Bashibulake Formation collected at the Mine section, and was processed at Utrecht University with HCl and HF but without heavy liquid separation or oxidation (Houben *et al.* 2011; Bosboom *et al.* 2014). The Xining sample (sample XJ88a) was taken from the middle Eocene Honggou Formation exposed at the Xiejia section, and processed at Palynological Laboratory Services Ltd, Holyhead, UK, with HCl, HF and heavy liquid separation. Further locality details are provided in Meijer *et al.* (2019).

Dataset 7: Extant conifer (Pinaceae and Podocarpaceae) pollen from Yunnan and Hunan provinces, China, collected in April to May 2018 directly from nine individual trees representing seven species (Jardine *et al.* 2020*b*, table S2). The sporopollenin was isolated using standard acetolysis.

Dataset 8: Paleocene fossil conifer pollen from the Nangqian Basin of east-central Tibet (NB unpub. data; see also Horton *et al.* 2002). Samples are from eight form-species in six genera. These samples were processed at Palynological Laboratory Services Ltd with HCl, HF and heavy liquid separation.

Data generation

Chemical data for the South American grasses, *Nitraria* and conifer pollen (Datasets 1–3, 5–8) were generated

using a Thermo Scientific Nicolet iN10 MX Dual detector IR microscope at the University of Amsterdam. Atmospheric CO_2 and H_2O variations were removed by purging the system with a dry nitrogen feed. Individual pollen grains were mounted onto ZnSe windows and analysed using 128 scans and a resolution of 4; in each case the microscope aperture was matched to the size of the pollen grain being analysed. A background spectrum was taken prior to each measurement and automatically removed from the sample spectrum. While we aimed for a target of 30–40 pollen grains per species, in some cases this was not possible due to a lack of material.

Chemical data for the Lake Bosumtwi grass pollen (Dataset 4) were generated at the Open University, UK, using a Thermo Nicolet FTIR bench unit interfaced with a Continuum IR-enabled microscope, which was fitted with a 15× reflachromat objective lens and a liquid nitrogen-cooled MCT-A detector. A purge unit was installed during data generation. Therefore, the uppermost 69 samples from the Bosumtwi sedimentary sequence were analysed with an unpurged system, while for the lower 62 samples the entire system was purged using a Peak Scientific ML85 purge unit. The lower 12 unpurged samples were also reanalysed with the purged system, to check the impact of the purging status on the results. To generate the data individual pollen grains were picked from the sediment samples and mounted on ZnSe windows in groups of 8–10 grains with 3 replicate groups per sample. FTIR spectra were generated using with a microscope aperture of 100 × 100 μm, 256 scans per sample and a resolution of 4. For further details see Jardine *et al.* (2016).

Data analysis

FTIR spectra from individual pollen grains are prone to high levels of Mie-type scattering and other distorting effects (Zimmermann *et al.* 2015*a*, 2016; Zimmermann 2018). This was the case with the individual grain spectra generated in this study, although high quality spectral data were obtained in the majority of cases for the 1800 to 800 cm^{-1} region (Jardine *et al.* 2020*b*, fig. S2). We therefore followed previous studies (Zimmermann & Kohler 2014; Bağcioğlu *et al.* 2015, 2017; Zimmermann *et al.* 2015*b*, 2016) in focusing on the lower end of the spectra, here limiting the spectral range to the 1800 to 800 cm^{-1} region, and discarded any spectra that contained significant noise or distortions in this interval. IR absorbance band assignment was carried out with reference to the published literature (Domínguez *et al.* 1998; Coates 2000; Mayo *et al.* 2003; Fraser *et al.* 2012; Bağcioğlu *et al.* 2015; Jardine *et al.* 2015, 2017; Julier *et al.* 2016; Woutersen *et al.* 2018).

Data analysis was carried out on non-differentiated spectra as well as first and second-derivative spectra. Since taking derivatives of spectra not only enhances smaller spectral features but also adds noise (Varmuza & Filzmoser 2009), the spectra were differentiated using the Savitzky–Golay (SG) smoothing algorithm, using a second degree polynomial and a window size (w, with a larger value leading to more smoothing) determined by leave one out cross validation coupled with k-nn (k-nearest neighbour) classification (described below; see also Julier *et al.* 2016; Jardine *et al.* 2019). All spectra (differentiated and non-differentiated) were processed using extended multiplicative signal correction (EMSC), a model based spectral processing technique for baseline correction, noise removal and normalization (Kohler *et al.* 2009; Zimmermann *et al.* 2015*a*, 2016; Liland *et al.* 2016). Non-differentiated spectra were preprocessed with an EMSC model incorporating third order polynomial effects, whereas differentiated spectra were preprocessed with an EMSC model incorporating first order polynomial effects. EMSC allows for a reference spectrum to be incorporated to guide the baseline correction (Zimmermann *et al.* 2015*a*; Liland *et al.* 2016). Here we followed standard practice and used the mean spectrum for each dataset as the EMSC reference spectrum, except for the *Nitraria* and Lake Bosumtwi spectra (Datasets 4–6) which all showed a pronounced baseline shift. Therefore, the baseline corrected mean spectrum for each of the three datasets was used as the EMSC reference spectra. For the *Nitraria* datasets we used a second-order polynomial baseline, and for the Bosumtwi data a first-order polynomial baseline (i.e. a straight line).

We tested classification potential using k-nn classification on the extant grass spectra (Dataset 1). Unknown objects (here IR spectra) are classified based on the identities of the most similar objects in the training set, with similarity in this case being the lowest Euclidean distance. The parameter k is the number of most similar objects (nearest neighbours) used for classification, and based on past research (Jardine *et al.* 2019) this was kept to $k = 1$ because increasing k beyond this does not improve classification success. To provide a thorough test of classification success we divided the dataset into a training set (2/3) and a validation set (1/3) by holding back every third sample for validation. We then used the training set to select the best combination of processing parameters (i.e. those that gave the highest classification success rate), comparing EMSC corrected non-differentiated, first-derivative and second-derivative spectra, varying the SG smoothing window (w) between 5 and 43. Classification success for the training set was tested using leave one out cross validation, where each spectrum is treated as an unknown and classified in turn, based on the identities of the rest of the spectra in the training set. The combination of processing parameters that produced the highest classification success rate was then used to process both the training set and the validation set;

the spectra in the validation set were classified using the identities of the most similar spectra in the training set. Since most genera are represented by just one species and one individual, we focus primarily on subfamily level classifications, although genus level classification success rate is also given for comparison.

We ordinated the FTIR spectra using principal components analysis (PCA), which enables complex multivariate data to be viewed on a small number of axes (Varmuza & Filzmoser 2009). The PCA plots presented in the main paper are based on non-differentiated spectra to facilitate interpretation of loadings plots; for comparison PCA plots based on second derivative spectra are provided in Jardine *et al.* (2020*b*, figs S3–S8).

All data analysis was carried out using R version 3.5.2 (R Core Team 2018) with the packages 'baseline' version 1.2-1 (Liland & Mevik 2015), 'caret' version 6.0-85 (Kuhn 2020), 'class' version 7.3-15 (Venables & Ripley 2002), 'corrplot' version 0.84 (Wei & Simko 2017), 'EMSC' version 0.9.0 (Liland 2017), 'maps' version 3.3.0 (Becker *et al.* 2018), and 'rworldmap' version 1.3-6 (South 2011). The raw data, R code and supplementary information are available for download from figshare (Jardine *et al.* 2020*b*, tables S1, S2 and figs S1–S9).

RESULTS

Chemotaxonomic signals from sporopollenin

The mean FTIR spectrum for each grass subfamily (Dataset 1) shows absorbance bands associated with carboxyl groups at 1710 cm^{-1} (C=O stretching), and aromatic compounds at 1590 cm^{-1} (C=C stretching), 1430 cm^{-1} and 1370 cm^{-1} (C–H bending) (Fig. 1). There are prominent peaks at 1170 and 1030 cm^{-1}, with the 1170 cm^{-1} peak having a shoulder at 1230 cm^{-1}. While peaks at or close to these wavenumbers are present in the phenolic compounds *para*-coumaric and ferulic acid that form part of the sporopollenin biopolymer (Bağcioğlu *et al.* 2015), similar peaks have also been interpreted as new C–O bonds forming during acetolysis (Domínguez *et al.* 1998). The small peak at 1650 cm^{-1} is the amide I peak (C=O stretching), and relates to proteins external to the pollen wall that were not fully removed during acetolysis (Jardine *et al.* 2015, 2017). While the spectra of the different grass subfamilies are consistent in the peaks present, there are subtle variations in both peak height and position.

The maximum subfamily classification success rate with the training dataset was 97%, with SG smoothed second derivative spectra and $w = 25$ (Table 3). The next best combination of parameters was with SG smoothed first derivative spectra (96%), and then unprocessed spectra (90%). The genus level classification success rate was 5–10% below the subfamily rate (Table 3). Processing the validation dataset with second derivatives and classifying the spectra based on the identities of the training dataset gave a classification success rate of 98% at the subfamily level, and 93% at the genus level. Confusion matrices for the training and validation datasets both show a similar pattern of misclassifications at the subfamily level, but without any obvious phylogenetic signal (i.e. specimens are not preferentially misclassified into more closely related subfamilies) (Jardine *et al.* 2020*b*, fig. S9).

The PCA shows clustering at subfamily level, despite large amounts of overlap among the different taxa (Fig. 2A). C_3 and C_4 grasses do not group consistently separately from each other when all taxa are considered (Fig. 2B), although it is challenging to distinguish between photosynthetic pathway and taxonomic patterns with this broad scale analysis, since most subfamilies are represented by either all C_3 or all C_4 species (Table 2). Both *Panicum* and *Paspalum* contain a mix of C_3 and C_4 species, and running a PCA on just these genera reveals both within-genus and within-species clustering (Fig. 3A), as well as a clearer C_3/C_4 separation, with the C_4 species having higher scores on PC 2 (Fig. 3B). While most species in the dataset are represented by only one individual plant, both *Panicum laxum* and *Panicum pilosum* are represented by two individuals. In both cases the individuals within each species cluster together in the ordination space, demonstrating a broadly consistent chemical signature among individuals of the same species (Fig. 3A).

Diagenesis in non-exceptional settings

There are substantial chemical differences between the extant (Dataset 1) and fossil (Datasets 2 and 3) South American grass pollen grains (Fig. 4). A PCA of these data divides the extant and fossil data on PC 1, with some additional separation between the Amazon Fan and western Amazonia spectra (Fig. 4A). Both the mean spectrum from each dataset (Fig. 4B) and the PC 1 loadings (Fig. 4C) show that the fossil spectra have relatively higher 1600 and 1370 cm^{-1} aromatic peaks, and a much reduced 1710 cm^{-1} carboxyl peak. Additional differences occur in the lower wavenumbers and probably relate at least in part to laboratory processing procedures: while all spectra have a peak at ~1030 cm^{-1}, this is more pronounced in the extant grains and also to some extent in the Amazon Fan spectra, all of which were acetolysed. The peak at 1170 cm^{-1} with a shoulder at 1230 cm^{-1} is strongly expressed in the extant spectra but very weakly in the fossil ones; conversely a peak at 1115 cm^{-1} is present in the fossil spectra but not those from the extant grasses. As with the peaks at ~1370, 1230, 1170 and 1030 cm^{-1}, a peak at ~1115 cm^{-1} is present in both

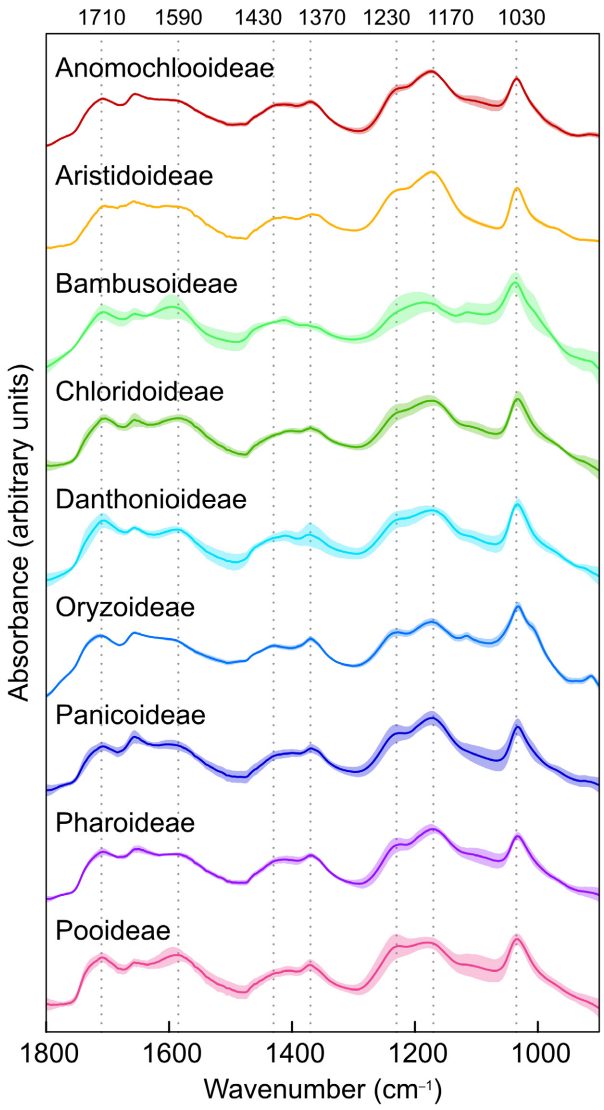


FIG. 1. Stacked mean FTIR spectra for each subfamily in the extant South American grass dataset. Subfamily colours are the same as those in Figure 2A, and shaded regions show ±1 standard deviation about the mean. Key peaks are marked with dashed grey lines.

para-coumaric acid and ferulic acid (Bağcioğlu *et al.* 2015).

Thermal maturation of palynomorphs causes them to darken in colour, which forms the basis of thermal maturation indices (Traverse 2007; Fraser *et al.* 2014*a*). Comparison of light microscope (LM) and scanning electron microscope (SEM) images of extant and early Miocene grass pollen, however, demonstrates that the early Miocene pollen is well-preserved and not visibly darkened (Fig. 5). (Since the processing protocol used for the western Amazonian samples (see Material and Method, above) will not have caused bleaching of the material this is taken to be the primary sporomorph colour.) This suggests that thermal maturation of the host rock is minimal, and the chemical differences between the extant and fossil pollen (Fig. 4) are not related to particularly poor preservational conditions.

TABLE 3. Maximum subfamily and genus-level classification success rates for the training dataset, for different data treatments, where *w* is the Savitzky–Golay smoothing window.

Processing	Subfamily		Genus	
	Maximum classification success	*w*	Maximum classification success	*w*
Unprocessed	90%	NA	80%	NA
Savitzky-Golay smoothing, 1st derivative	96%	9	89%	7
Savitzky-Golay smoothing, 2nd derivative	97%	25	92%	15

There are some differences detectable among localities in the western Amazonia fossil spectra (Dataset 2), with samples from Santa Sofia having a more pronounced 1370 cm^{-1} peak and samples from Los Chorros having higher peaks at ~1040 and 1115 cm^{-1} (Fig. 6A). Multiple samples from the same locality have highly similar spectra, which is also the case for Los Chorros despite these samples representing different lithologies and sample processing methods (sample 16835 is clastic that was processed with sodium pyrophosphate, while sample 16788 is a lignite that was processed with both sodium pyrophosphate and Schultze reagent; see Material and Method, above, for further details). A PCA shows grouping of spectra within localities, but no consistent changes with time (Fig. 6B). Similarly, there are differences among samples in the Amazon Fan fossil spectra (Dataset 3) that leads to within-sample grouping in a PCA, but no obvious changes through time (Fig. 7).

The Lake Bosumtwi grasses (Dataset 4) show limited evidence for consistent chemical changes over the last 500 kyr. The first two axes of a PCA, which together account for 56% of the variance in the dataset, show no obvious temporal trends, apart from a possible slight increase in PC 2 scores at ~350 ka BP (Fig. 8A). Higher principal components also show no clear change through time (Jardine *et al.* 2020*b*, fig. S8). The main chemical shift occurs ~50 to 90 ka BP with a transient increase in PC 1 scores, driven by increases and decreases in the relative heights of peaks and troughs across the spectra (Fig. 8B). There are no clear differences associated with whether the FTIR system was purged or unpurged

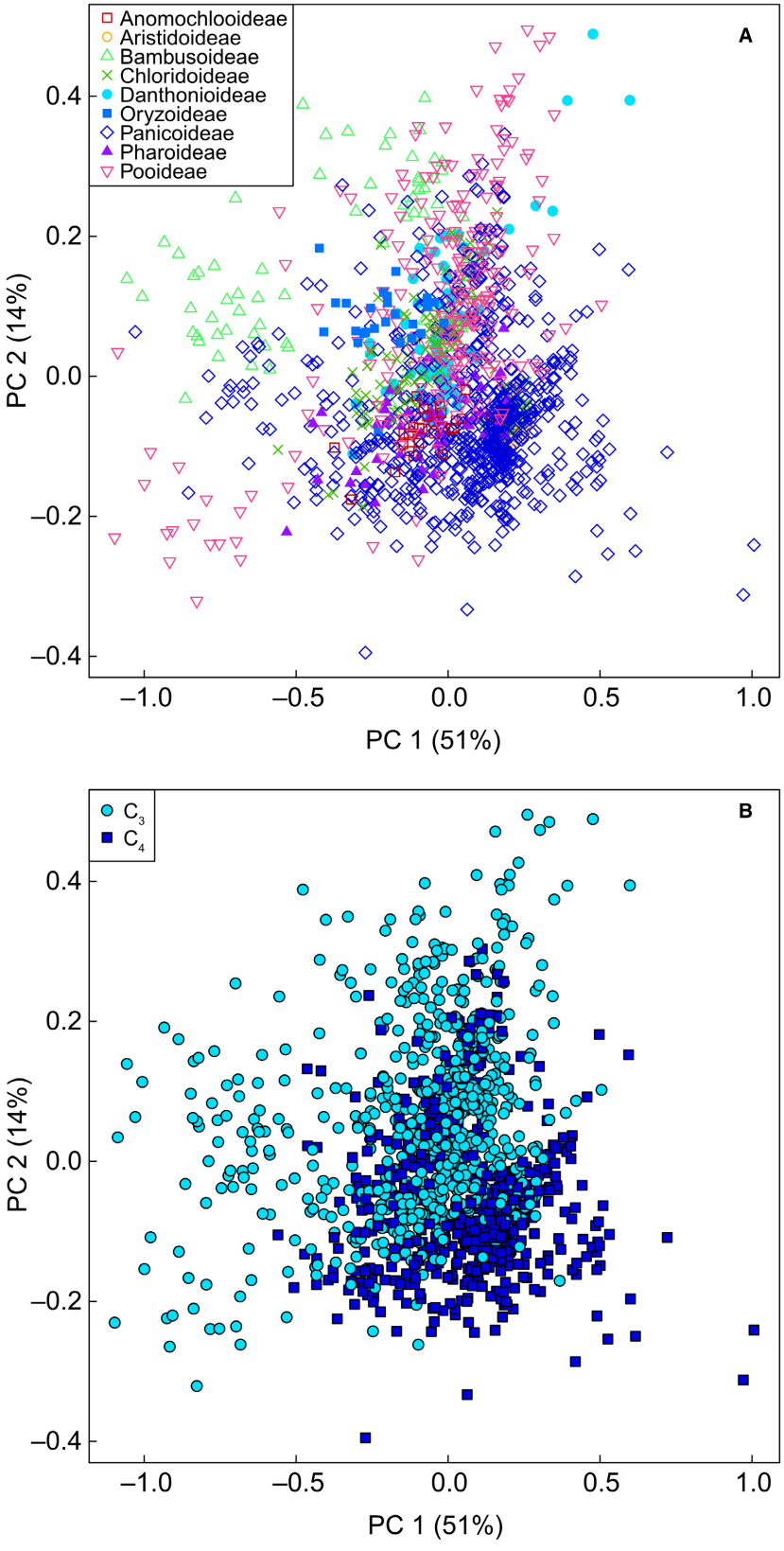


FIG. 2. Principal components analysis (PCA) of extant South American grass pollen FTIR spectra. A, symbols coloured by subfamily. B, symbols coloured by photosynthetic pathway. Values in parentheses provide the percentage of variance explained by each principal component. Subfamily colours in A are the same as those in Figure 1.

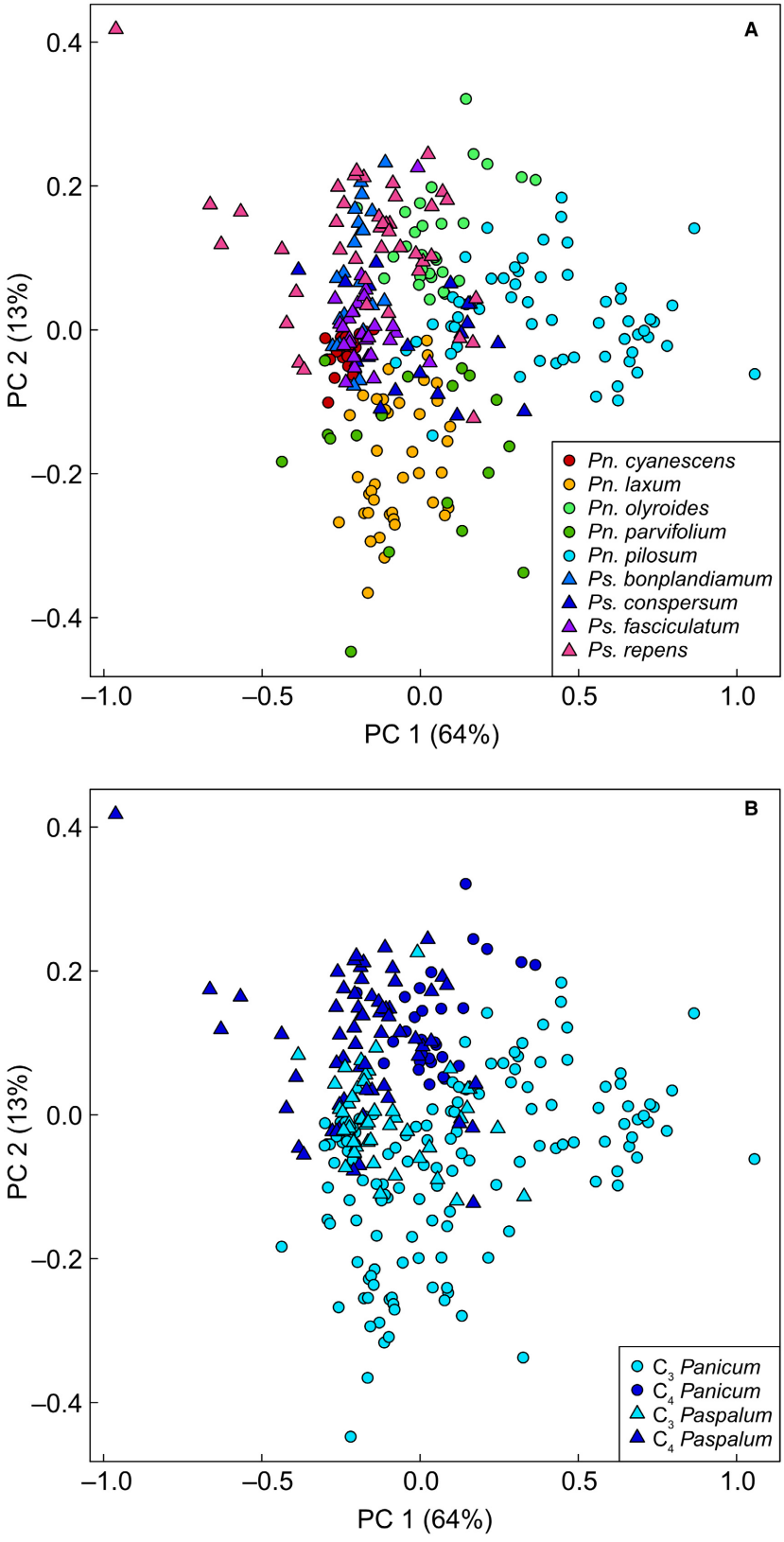


FIG. 3. PCA of *Panicum* and *Paspalum* FTIR spectra. A, symbols coloured by species; B symbols coloured by photosynthetic pathway. Circles = *Panicum*; triangles = *Paspalum*. Values in parentheses provide the percentage of variance explained by each principal component.

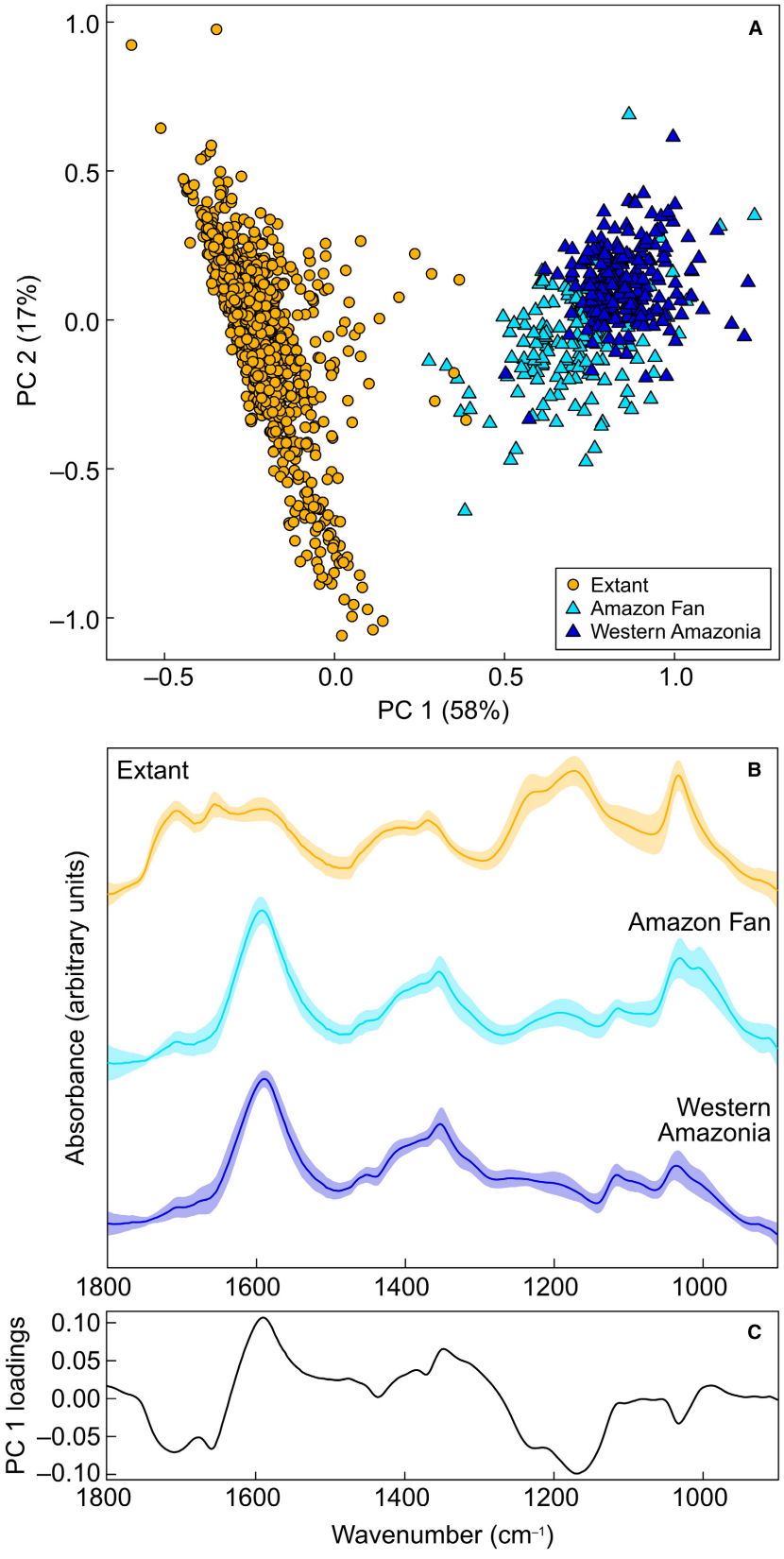


FIG. 4. Comparison of extant and fossil South American grass pollen. A, PCA of extant and fossil South American grasses, with extant grass pollen grains shown as orange circles, Amazon Fan fossil grains shown as light blue triangles, and western Amazonia fossil grains shown as dark blue triangles; values in parentheses provide the percentage of variance explained by each principal component. B, stacked mean spectra for the three datasets, with colours as in A and shaded regions showing ±1 standard deviation about the mean. C, loadings plot for PC 1.

(Jardine *et al.* 2020*b*, fig. S8). Comparing acetolysed and non-acetolysed spectra reveals the presence of peaks at 1170 and 1030 cm^{-1} in the acetolysed spectra, and a more pronounced 1710 cm^{-1} peak in the non-acetolysed spectra, as well as a range of smaller peaks in the 1250 to 1450 cm^{-1} range (Fig. 8C). The acetolysed Bosumtwi spectra are similar to those of the extant South American grasses (Dataset 1; Figs 1, 4B), with the major differences being the lack of protein-related peaks in the Bosumtwi spectra, which suggests the full removal of these during burial and/or processing. The non-acetolysed Bosumtwi spectra are distinct from the (also non-acetolysed) western Amazonia fossil spectra (Dataset 2; Fig. 4B) in having a much more pronounced 1710 cm^{-1} peak and a greater number of clearly identifiable peaks in the < 1450 cm^{-1} region, although both contain 1600 and 1370 cm^{-1} aromatic peaks.

Generality across taxa

Both the *Nitraria* (Datasets 5 and 6) and conifer (Datasets 7 and 8) pollen show large differences between the extant and fossil spectra (Fig. 9), with similar changes to those in the South American grasses (Datasets 1–3; Fig. 4B). While the spectra for the extant taxa are broadly similar, the Palaeogene *Nitraria* and conifer pollen spectra are simpler than the Miocene to Pleistocene South American grass spectra, with less clearly defined peaks. The fossil *Nitraria* spectra are broadly similar across the two samples and basins (Tarim vs Xining), but show some differences in the 1000 to 1100 cm^{-1} range (Fig. 9A).

DISCUSSION

Our results demonstrate a mix of promising aspects and challenges for sporopollenin chemistry-based palynology. The high classification success rate demonstrated by the extant grass taxa (Dataset 1, 98% at subfamily level for the validation dataset) indicates that sporomorph chemotaxonomy can be applied effectively with isolated sporopollenin, including with acetolysed grains and larger numbers of taxa. Even reliably classifying grass pollen grains to subfamily level represents a major advance from what is possible with standard light microscopy, and our genus level results suggest that classifying to finer taxonomic levels may also be achievable (Julier *et al.* 2016). Our results also suggest that the photosynthetic pathways of plants may be identifiable via FTIR analysis of sporomorphs (Fig. 3B), and this would provide a new and efficient means of quantifying the relative abundance of C_3 versus C_4 grasses through time (Amundson *et al.* 1997; Loader & Hemming 2000; Nelson *et al.* 2006, 2008). Nevertheless, we stress that further replication is needed across taxonomic scales, including multiple individuals within species, to gain a fuller understanding of likely classification success at genus and species level, and better assess the potential for distinguishing among C_3 and C_4 grasses.

The consistency of the signal through time in the Lake Bosumtwi grass pollen spectra (Dataset 4, Fig. 8), and the broad chemical similarity with the extant South American grasses (Fig. 4C), suggests that chemotaxonomy using extant calibration datasets has potential in late Quaternary settings. The late Quaternary timeframe encompasses questions regarding vegetation changes across glacial–interglacial cycles (Torres *et al.* 2005, 2013; Melles *et al.* 2012; Miller & Gosling 2014; Miller *et al.* 2016), human impacts on natural systems (Bush 2002; Bush *et al.* 2015; Loughlin *et al.* 2018), and the spread of agriculture (Piperno *et al.* 2004; Weiss *et al.* 2004; Fuller 2007; Crowther *et al.* 2016; Jardine *et al.* 2019), which could be addressed using a chemotaxonomic approach. In addition to increasing the taxonomic resolution of morphologically cryptic taxa (such as grasses), chemotaxonomy may allow for automated pollen counting, which is a long-standing goal in palynology (Holt *et al.* 2011; Holt & Bennett 2014).

In spite of the promising results demonstrated by the extant and late Quaternary datasets, considerable differences occur between the spectra of extant and fossil pollen. These differences are caused partly by the laboratory processing methods used, and partly by post-burial processes that result in the transformation of biological material into geological/palaeobotanical samples. Comparison of the acetolysed and non-acetolysed grass pollen from Lake Bosumtwi (Dataset 4), in which sample age and processing are otherwise identical, reveals a number of differences across the spectra (Fig. 8C). Encouragingly, the non-acetolysed spectra are highly similar to published FTIR spectra of sporopollenin that has been chemically isolated for microencapsulation, for example the 'high temperature switched treatment' of Gonzalez-Cruz *et al.* (2018; see their fig. 7). This method employs treatments with acetone (overnight), orthophosphoric acid (7 days) and KOH (12 h), with interim washes with water, acetone, HCl, NaOH and ethanol, to isolate sporopollenin. This similarity suggests that the non-acetolysed Bosumtwi spectra are a reasonable measurement of 'naturally' isolated sporopollenin before diagenetic processes have caused further changes, which provides a benchmark to compare the extant and fossil FTIR spectra against.

The spectra of acetolysed extant and late Quaternary pollen (Figs 1, 4B, 8C, 9) contain prominent peaks at 1170 and 1030 cm^{-1} that are only weakly expressed in the non-acetolysed material. A similar effect was demonstrated by Domínguez *et al.* (1998), and the presence of additional carbon compounds following acetolysis was

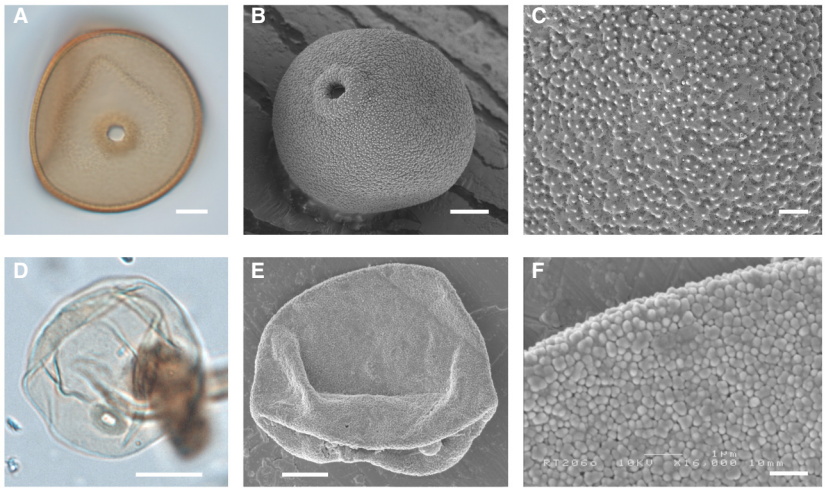


FIG. 5. Extant and fossil grass pollen under LM and SEM. A–C, extant grains from *Paspalum fasciculatum* (subfamily Panicoideae). D–F, fossil grains from the early Miocene Santa Teresa locality in Peru. Scale bars represent: 10 μm (A, D); 5 μm (B, E); 1 μm (C, F).

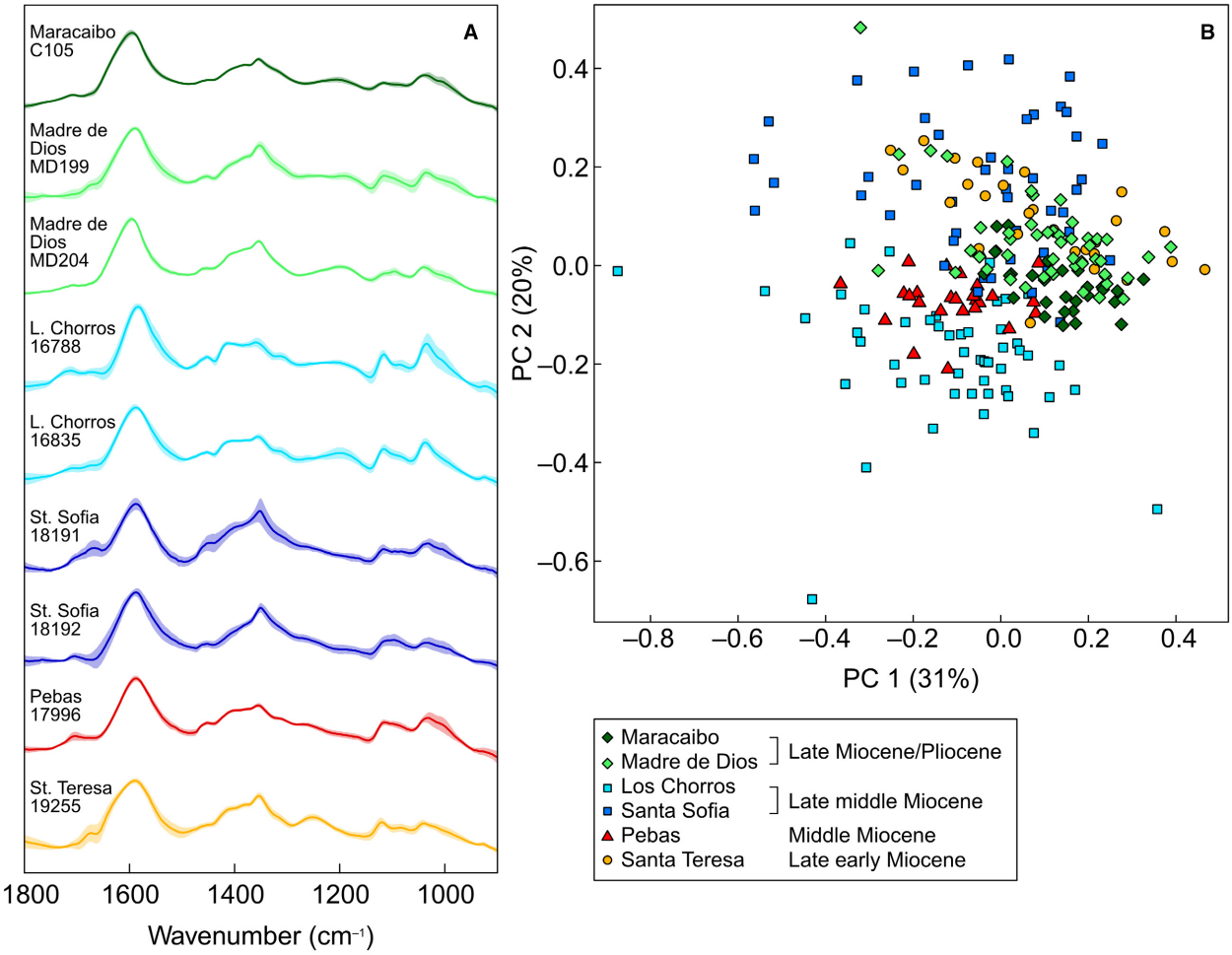


FIG. 6. Western Amazonia fossil grasses. A, mean FTIR spectrum for each sample, ordered by time interval and with shaded regions showing ±1 standard deviation about the mean. B, PCA of individual spectra, with symbol style representing time period, and colours representing locality; values in parentheses provide the percentage of variance explained by each principal component.

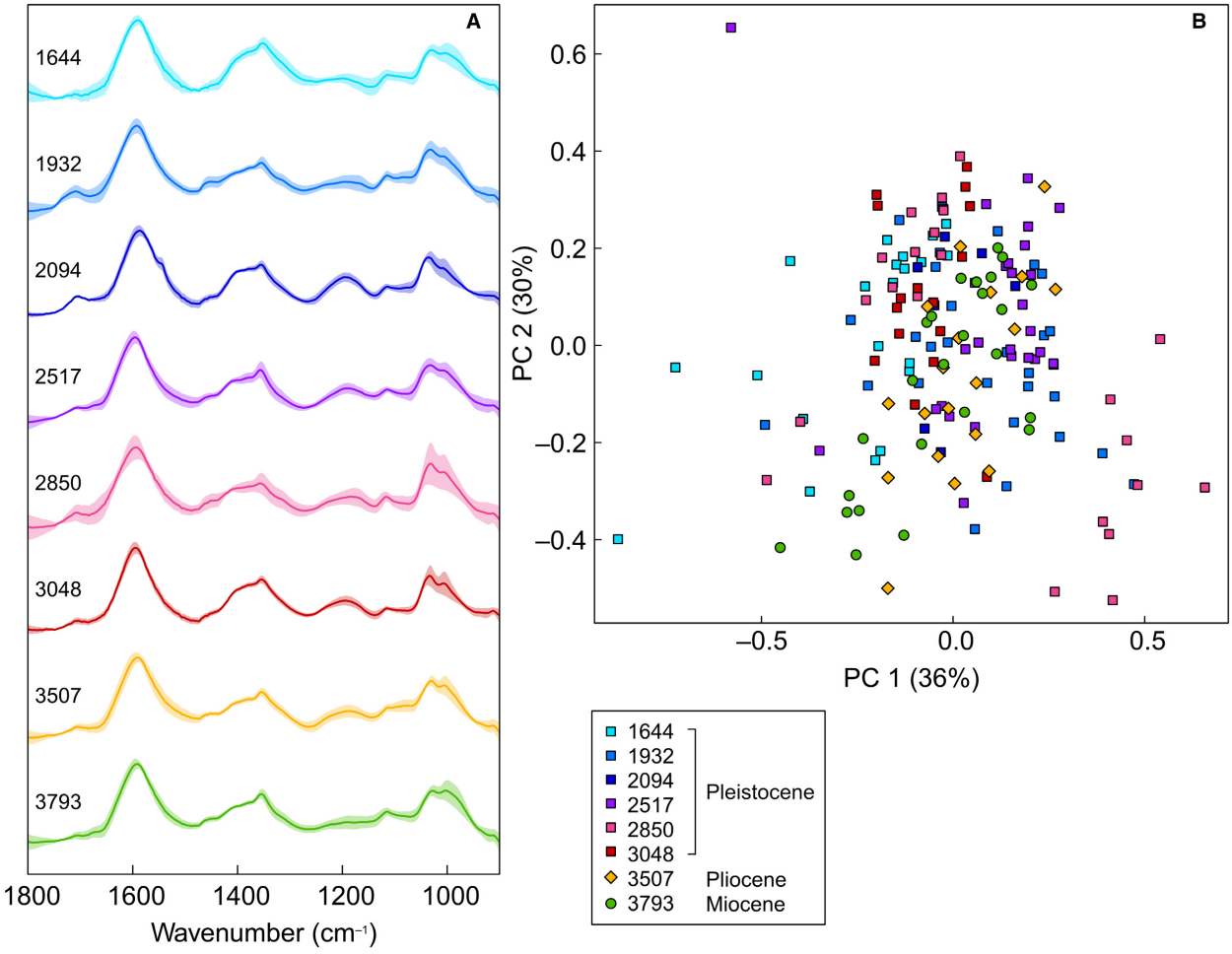


FIG. 7. Amazon Fan fossil grasses. A, mean FTIR spectrum for each sample, ordered by core depth and with shaded regions showing ±1 standard deviation about the mean. B, PCA of individual spectra, with symbol style representing time interval, and colours representing sample; values in parentheses provide the percentage of variance explained by each principal component. Sample numbers are metres below drill floor (mbdf) sample depth.

highlighted by Amundson *et al.* (1997) and Loader & Hemming (2000). Amundson *et al.* (1997) put their presence down to the incomplete removal of glacial acetic acid or cellulose acetate following acetolysis. Domínguez *et al.* (1998) found that refluxing the acetolysed material in methanol removed the additional peaks, so this should be explored as an extra step in palynological processing, although it is not clear if this would remove all differences between acetolysed and non-acetolysed material (Fig. 8C). Otherwise, if the samples to be compared are all extant or late Quaternary, it will be necessary to ensure that all have undergone the same chemical treatments to make them fully comparable.

Interestingly, the effect of acetolysis was less pronounced in the fossil material. Comparing the Miocene to Pliocene dataset from western Amazonia (Dataset 2, not acetolysed) with the Miocene to Pleistocene Amazon Fan dataset (Dataset 3, acetolysed) reveals broadly similar spectra (Fig. 4C). The additional peaks that are present in the extant and late Quaternary acetolysed material are only subtly expressed in the Amazon Fan spectra. As yet, it is not clear whether this is down to different base materials (i.e. extant sporopollenin vs sporopollenin that has undergone a degree of burial across longer geological timescales) or to additional steps in the processing of the Amazon Fan samples that could have removed the impact of acetolysis. While the extant pollen grains were only acetolysed, and the Lake Bosumtwi samples were acetolysed as the last step in their processing, the Amazon Fan samples underwent heavy liquid separation after acetolysis. It is not clear how this may have impacted upon the sporopollenin chemistry, but even further additional washing steps may have removed the chemical imprint of the acetolysis

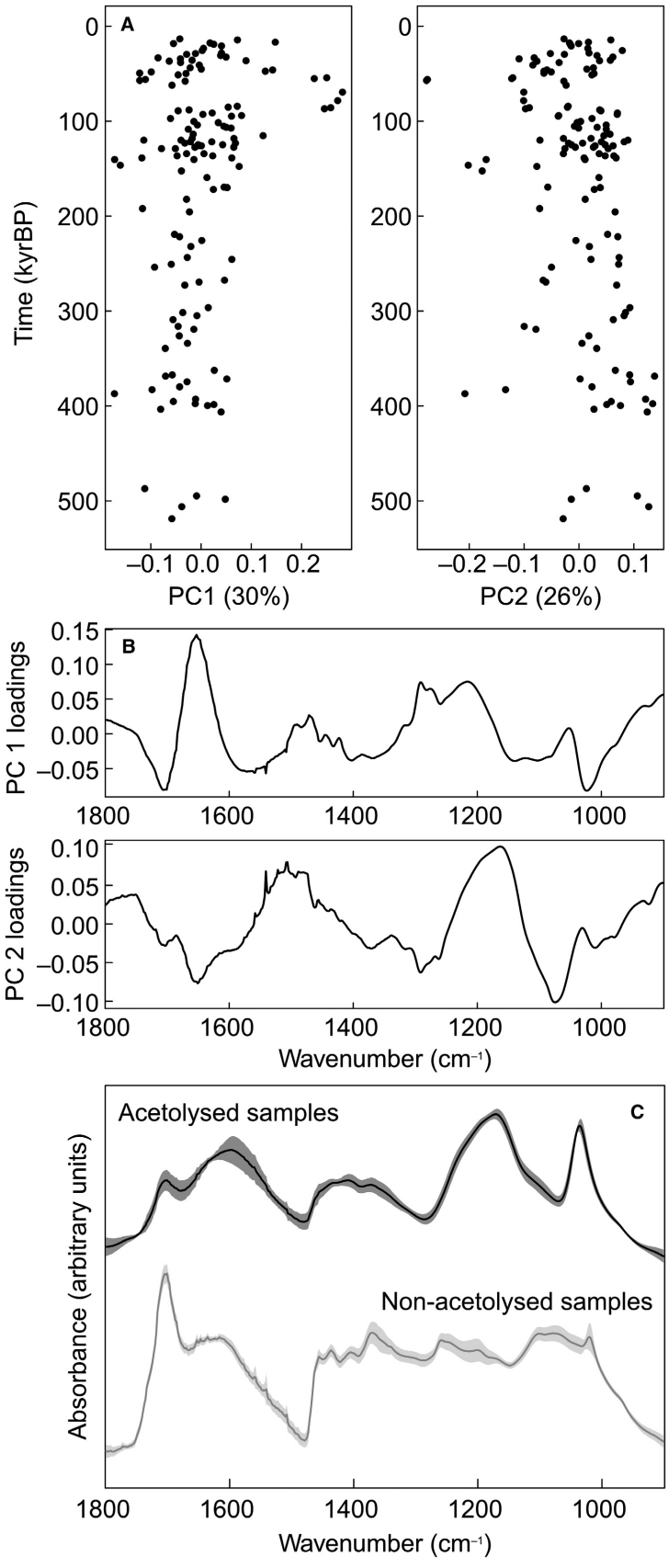


FIG. 8. FTIR data from Lake Bosumtwi, Ghana. A, PCA sample scores from PC 1 and PC 2 plotted against time; values in parentheses provide the percentage of variance explained by each principal component. B, loadings from PC 1 and PC 2. C, mean spectra for subset of eight samples, both acetolysed and non-acetolysed; shaded regions show ±1 standard deviation about the mean.

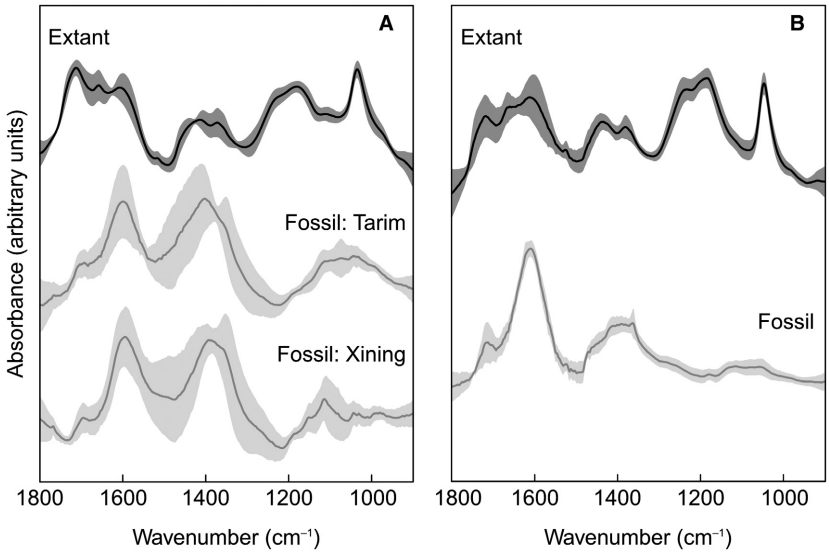


FIG. 9. Mean FTIR spectra for extant and fossil *Nitraria* and conifer pollen, with shaded regions showing ±1 standard deviation about the mean. A, *Nitraria*. B, conifers.

procedure. Clearly, more work needs to be done in this area.

Beyond acetolysis, the other processing reagents used in this study may have led to inter-dataset differences. While the extant datasets were simply acetolysed, the fossil datasets were variously treated with KOH, HF, HCl, sodium pyrophosphate, Schulze reagent and heavy liquid separation, as well as acetolysis in some cases (Table 1). KOH has been shown to have a minimal impact on sporopollenin chemistry (Lutzke *et al.* 2020), and the same is thought to be true for HF (Domínguez *et al.* 1998) but the effect of the other reagents has not been explicitly tested. The Bosumtwi samples (Dataset 4) were treated with HF, HCl and KOH, and the overall similarity of the extant (Figs 1, 4) and acetolysed Bosumtwi (Fig. 8C) spectra suggests that the impact of these reagents on sporopollenin chemistry is likely to be limited. This is supported by the similarity between the non-acetolysed Bosumtwi spectra (Fig. 8C) and published FTIR spectra from sporopollenin isolated for pharmaceutical microencapsulation applications (see previous discussion and Gonzalez-Cruz *et al.* 2018). If correct, this suggests that the HF and HCl used in processing the fossil *Nitraria* (Dataset 6) and conifer (Dataset 8) pollen should also have caused minimal changes to the sporopollenin chemistry, although we cannot rule out subtle responses to these reagents.

The two samples from the Miocene Los Chorros locality in western Amazonia (Dataset 2) represent two different lithologies (clastic and lignite) and processing approaches: while both samples were processed with sodium pyrophosphate and bromoform heavy liquid separation, the lignite sample (sample 16788) had an additional processing step with the use of Schulze reagent. The pollen from both samples are chemically highly similar and plot together in ordination space (Fig. 6), which suggests that (fossil) sporopollenin chemistry is robust to Schulze reagent, although the impact of length of treatment with Schulze reagent requires further study. The effect of heavy liquid separation is harder to evaluate, but some of the subtle differences between the fossil *Nitraria* samples (Fig. 9A) may be the result of this procedure being used in processing the Xining sample but not the Tarim sample. However, given the short exposure time to the reagent, we view this scenario as being plausible but unlikely. The two fossil *Nitraria* samples are also much more similar to each other than they are to the fossil conifer sample (Fig. 9B), despite the use of heavy liquid separation with this sample as well. In summary, while more research needs to be carried out in this area, our results indicate that acetolysis leaves a relatively stronger imprint on sporopollenin chemistry, while other reagents appear to alter the sporopollenin chemical signature much less.

There are also changes in pollen chemistry that are not attributable to the processing methodology. In particular, the loss of the 1710 cm^{-1} carboxyl peak, the relative dominance of the 1600 cm^{-1} aromatic peak, and the general simplification of the spectra demonstrated in the fossil samples is more likely to be down to diagenetic processes acting on the sporopollenin chemistry. This is consistent with the finding that sporopollenin comprises a more labile component comprising carboxylic acid monomers (Watson *et al.* 2012), with the major differences between the extant/late Quaternary and fossil material being due to the loss or repolymerization of this labile fraction during early diagenesis. This loss is not recreated with acetolysis or burial over short-timescales (i.e. 10^5 years) and/or at shallow depths (150 m for the

oldest Lake Bosumtwi samples, vs 865 m for the sample of a similar age from the Amazon Fan). Our results suggest that this repolymerization can happen relatively early once sporomorphs enter the geological record, but do not rule out the possibility that the resultant geopolymer can be altered by further diagenetic processes. Such an effect may account for the continued simplification of the spectra shown by the Palaeogene *Nitraria* and conifer pollen (Datasets 6 and 8; Fig. 9) compared to the Neogene grasses (Datasets 2 and 3; Fig. 4). These results also suggest that exceptional preservation of the kind reported by Fraser *et al.* (2012) is likely to be rare in the fossil record, with some level of chemical change to be anticipated even over short geological timescales.

Similar changes (i.e. the loss carboxylic acids and a concomitant reduction in the 1710 cm^{-1} C=O peak) have been recreated in thermal maturation experiments (Yule *et al.* 2000; Watson *et al.* 2012; Bernard *et al.* 2015), suggesting that such experiments are successfully simulating early diagenetic processes and the repolymerization of the labile sporopollenin fraction. It is not clear, however, how well they can recreate longer term changes following the formation of the sporopollenin geopolymer. Nevertheless, a similar approach with extant material is likely to provide a better match for deep time fossil specimens than is possible with acetolysis (or other laboratory processing) alone, which may enable the development of calibration datasets from extant samples that can be used with fossil material. Using the chemical signature of fossil specimens with known affinities may also be viable as a means of classifying problematic fossil material. Such as approach has been used with Silurian cryptospores to demonstrate their relationship with land plants rather than marine palynomorph groups (Steemans *et al.* 2010), and with fossil cuticle fragments with ambiguous phylogenetic relationships (Vajda *et al.* 2017). Therefore, as long as a taxonomic signature is still present in deep time sporopollenin (or cuticular material), there is potential to use it for classification purposes even if it is modified from its original state.

We have taken a broad-scale exploratory approach with this study, analysing different datasets that incorporate a range of taxa, time periods and laboratory processing procedures. This has enabled us to consider a number of variables that may impact upon sporopollenin chemistry, using samples that are representative of the majority of routine palynological investigations and existing sample sets. However, it also means that separating out the impact of these different variables, and understanding how they act individually and in combination with each other, is more challenging than in a manipulative experimental setting. Further work is therefore needed to test and validate the results presented in this paper. This research should be targeted towards dedicated experimental approaches where processing procedures can be harmonized across extant and fossil datasets, and the potential controlling factors can be manipulated and tested independently of each other. This will be particularly useful for disentangling the impact(s) of different processing reagents, how they may interact in processing protocols, and whether and how they act differently on sporopollenin from extant sporomorphs vs fossil sporopollenin of different diagenetic grades. A further area of uncertainty in our analysis is to what extent ambient growing conditions, aside from variations in UVB (Lomax *et al.* 2008), exert a control on sporopollenin chemistry, and whether that could lead to differences among datasets. A better understanding of other potential environmental controls on sporopollenin chemistry is therefore needed, ideally both from controlled growth experiments as well as collecting from natural environmental gradients.

CONCLUSION

We have shown that that extant sporomorphs can be successfully classified using the chemical signature of sporopollenin. Diagenetic repolymerization of sporopollenin happens relatively soon after burial, however, meaning that chemical datasets from extant and fossil sporomorphs are challenging to integrate directly. Further differences between extant and fossil material can occur via the processing methods used to isolate sporopollenin from fresh sporomorphs. Taken together, these results show that linking extant and fossil sporomorph chemistry datasets is not a trivial task. In late Quaternary settings sporopollenin appears chemically highly similar to extant material, suggesting great potential to integrate across datasets if the laboratory processing methods are similar. This supports the idea of using FTIR-based chemotaxonomy to increase the taxonomic resolution in morphologically cryptic groups (Julier *et al.* 2016; Jardine *et al.* 2019), and to develop automated classification systems for generating pollen counts (Zimmermann *et al.* 2016). For deeper time material, however, changes in sporopollenin chemistry are likely to have occurred leading to systematic differences between the extant and fossil chemical signature. Further experiments in laboratory processing and thermal maturation (Watson *et al.* 2012; Fraser *et al.* 2014*a*; Bernard *et al.* 2015; Jardine *et al.* 2015), applied to a range of taxa, will therefore be necessary to study the impact of geological processes on classification ability and the phylogenetic and environmental signals in sporopollenin. It remains to be seen whether spectra close to those from pre-Quaternary specimens can be recreated from extant sporomorphs, and therefore whether classification from extant training sets is feasible for deep time fossil

material. Where this is not possible, generating training sets based on fossil morphotaxa will still be of great value for assigning specimens of unknown affinity to plant lineages, which in itself may help unravel some of the key developments in land plant evolution and diversification.

Acknowledgements. We are deeply indebted to Antoine Cleef for collecting the grass sample material from Naturalis Biodiversity Center in Leiden. We also thank Annemarie Philip and Malcolm Jones for sample processing, and Caixia Wei for pollen photography. Sally Thomas and four anonymous reviewers are thanked for detailed reviews that greatly increased the quality of this paper. This research was funded by ERC grant MAGIC 649081 (GDN), and NERC grant NE/K005294/1 (WDG, WTF, BHL). BHL and WTF's ongoing work on sporopollenin chemistry is currently supported by NERC grants NE/P013724/1 and NE/R001324/1. PEJ is currently funded by DFG grant 443701866.

DATA ARCHIVING STATEMENT

Research Data and R code associated with this article, as well as an appendix that contains the supplementary figures and tables, can be accessed at: https://doi.org/10.6084/m9.figshare.11382102

Editor. Evelyn Kustatscher

REFERENCES

AMUNDSON, R., EVETT, R. R., JAHREN, A. H. and BARTOLOME, J. 1997. Stable carbon isotope composition of Poaceae pollen and its potential in paleovegetational reconstructions. *Review of Palaeobotany & Palynology*, **99**, 17–24.

BAĞCIOĞLU, M., ZIMMERMANN, B. and KOHLER, A. 2015. A multiscale vibrational spectroscopic approach for identification and biochemical characterization of pollen. *PLoS One*, **10**, 1–19.

—— KOHLER, A., SEIFERT, S., KNEIPP, J., ZIMMERMANN, B. and McMAHON, S. 2017. Monitoring of plant–environment interactions by high-throughput FTIR spectroscopy of pollen. *Methods in Ecology & Evolution*, **8**, 870–880.

BARBOLINI, N., RUBIDGE, B. S. and BAMFORD, M. K. 2018. A new approach to biostratigraphy in the Karoo foreland system: utilising restricted-range palynomorphs and their first appearance datums for correlation. *Journal of African Earth Sciences*, **140**, 114–133.

BECKER, R. A., WILKS, A. R., BROWNRIGG, R., MINKA, T. P. and DECKMYN, A. 2018. maps: Draw Geographical Maps. R package version 3.3.0. https://CRAN.R-project.org/package=maps

BELL, B. A., FLETCHER, W. J., RYAN, P., SEDDON, A. W. R., WOGELIUS, R. A. and 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**, 1382–1394.

BERMÚDEZ, M. A., HOORN, C., BERNET, M., CARRILLO, E., VAN DER BEEK, P. A., GARVER, J. I., MORA, J. L. and MEHRKIAN, K. 2017. The detrital record of late-Miocene to Pliocene surface uplift and exhumation of the Venezuelan Andes in the Maracaibo and Barinas foreland basins. *Basin Research*, **29**, 370–395.

BERNARD, S., BENZERARA, K., BEYSSAC, O., BALAN, E. and BROWN, G. E. Jr 2015. Evolution of the macromolecular structure of sporopollenin during thermal degradation. *Heliyon*, **1**, e00034.

BLOKKER, P., YELOFF, D., BOELEN, P., BROEKMAN, R. A. and ROZEMA, J. 2005. Development of a proxy for past surface UV-B irradiation: a thermally assisted hydrolysis and methylation py-GC/MS method for the analysis of pollen and spores. *Analytical Chemistry*, **77**, 6026–6031.

—— BOELEN, P., BROEKMAN, R. and ROZEMA, J. 2006. The occurrence of *p*-coumaric acid and ferulic acid in fossil plant materials and their use as UV-proxy. *Plant Ecology*, **182**, 197–207.

BOSBOOM, R., DUPONT-NIVET, G., GROTHE, A., BRINKHUIS, H., VILLA, G., MANDIC, O., STOICA, M., KOUWENHOVEN, T., HUANG, W., YANG, W. and GUO, Z. 2014. Timing, cause and impact of the late Eocene stepwise sea retreat from the Tarim Basin (west China). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **403**, 101–118.

BRINKHUIS, H., SCHOUTEN, S., COLLINSON, M. E., SLUIJS, A., SINNINGHE DAMSTE, J. S., DICKENS, G. R., HUBER, M., CRONIN, T. M., ONODERA, J., TAKAHASHI, K., BUJAK, J. P., STEIN, R., VAN DER BURGH, J., ELDRETT, J. S., HARDING, I. C., LOTTER, A. F., SANGIORGI, F., VAN KONIJNENBURG-VAN CITTERT, H., DE LEEUW, J. W., MATTHIESSEN, J., BACKMAN, J. and MORAN, K. 2006. Episodic fresh surface waters in the Eocene Arctic Ocean. *Nature*, **441**, 606–609.

BUSH, M. B. 2002. On the interpretation of fossil Poaceae pollen in the lowland humid neotropics. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **177**, 5–17.

—— McMICHAEL, C. H., PIPERNO, D. R., SILMAN, M. R., BARLOW, J., PERES, C. A., POWER, M. and PALACE, M. W. 2015. Anthropogenic influence on Amazonian forests in pre-history: an ecological perspective. *Journal of Biogeography*, **42**, 2277–2288.

COATES, J. 2000. Interpretation of infrared spectra, a practical approach. 10815–10837. *In* MEYERS, R. A. (ed.) *Encyclopedia of analytical chemistry*. John Wiley & Sons.

COLLINSON, M. E., FOWLER, K. and BOULTER, M. C. 1981. Floristic changes indicate a cooling climate in the Eocene of southern England. *Nature*, **291**, 315–317.

CRANE, P. R. and LIDGARD, S. 1989. Angiosperm diversification and paleolatitudinal gradients in cretaceous floristic diversity. *Science*, **246**, 675–678.

CROWTHER, A., LUCAS, L., HELM, R., HORTON, M., SHIPTON, C., WRIGHT, H. T., WALSHAW, S., PAWLOWICZ, M., RADIMILAHY, C., DOUKA, K., PICORNELL-GELABERT, L., FULLER, D. Q. and BOIVIN, N. L. 2016. Ancient crops provide first

archaeological signature of the westward Austronesian expansion. *Proceedings of the National Academy of Sciences*, **113**, 6635–6640.

DELL'ANNA, R., LAZZERI, P., FRISANCO, M., MONTI, F., MALVEZZI CAMPEGGI, F., GOTTARDINI, E. and BERSANI, M. 2009. Pollen discrimination and classification by Fourier transform infrared (FT-IR) microspectroscopy and machine learning. *Analytical & Bioanalytical Chemistry*, **394**, 1443–1452.

DOMÍNGUEZ, E., MERCADO, J. A., QUESADA, M. A. and HEREDIA, A. 1998. Isolation of intact pollen exine using anhydrous hydrogen fluoride. *Grana*, **37**, 93–96.

FRASER, W. T., SEPHTON, M. A., WATSON, J. S., SELF, S., LOMAX, B. H., JAMES, D. I., WELLMAN, C. H., CALLAGHAN, T. V. and BEERLING, D. J. 2011. UV-B absorbing pigments in spores: biochemical responses to shade in a high-latitude birch forest and implications for sporopollenin-based proxies of past environmental change. *Polar Research*, **30**, 8312.

—— SCOTT, A. C., FORBES, A. E. S., GLASSPOOL, I. J., PLOTNICK, R. E., KENIG, F. and LOMAX, B. H. 2012. Evolutionary stasis of sporopollenin biochemistry revealed by unaltered Pennsylvanian spores. *New Phytologist*, **196**, 397–401.

—— WATSON, J. S., SEPHTON, M. A., LOMAX, B. H., HARRINGTON, G. J., GOSLING, W. D. and SELF, S. 2014*a*. Changes in spore chemistry and appearance with increasing maturity. *Review of Palaeobotany & Palynology*, **201**, 41–46.

—— LOMAX, B. H., JARDINE, P. E., GOSLING, W. D. and SEPHTON, M. A. 2014*b*. Pollen and spores as a passive monitor of ultraviolet radiation. *Frontiers in Ecology & Evolution*, **2**.

FREDERIKSEN, N. O. 1998. Upper Paleocene and lowermost Eocene angiosperm pollen biostratigraphy of the eastern Gulf Coast and Virginia. *Micropaleontology*, **44**, 45–68.

FULLER, D. Q. 2007. Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. *Annals of Botany*, **100**, 903–924.

GONZALEZ-CRUZ, P., UDDIN, M. J., ATWE, S. U., ABIDI, N. and GILL, H. S. 2018. Chemical treatment method for obtaining clean and intact pollen shells of different species. *ACS Biomaterials Science & Engineering*, **4**, 2319–2329.

GRAY, J. and BOUCOT, A. J. 1971. Early Silurian spore tetrads from New York: earliest New World evidence for vascular plants? *Science*, **173**, 918–921.

GUPTA, N. S., BRIGGS, D. E. G., COLLINSON, M. E., EVERSHED, R. P., MICHELS, R., JACK, K. S. and PANCOST, R. D. 2007. Evidence for the in situ polymerisation of labile aliphatic organic compounds during the preservation of fossil leaves: implications for organic matter preservation. *Organic Geochemistry*, **38**, 499–522.

HARRINGTON, G. J. and JARAMILLO, C. A. 2007. Paratropical floral extinction in the Late Palaeocene-Early Eocene. *Journal of the Geological Society, London*, **164**, 323–332.

HOLT, K. A. and BENNETT, K. D. 2014. Principles and methods for automated palynology. *New Phytologist*, **203**, 735–742.

HOLT, K., ALLEN, G., HODGSON, R., MARSLAND, S. and FLENLEY, J. 2011. Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory. *Review of Palaeobotany & Palynology*, **167**, 175–183.

HOORN, C. 1994. An environmental reconstruction of the palaeo-Amazon River system (Middle–Late Miocene, NW Amazonia). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **112**, 187–238.

—— STRAATHOF, J., ABELS, H. A., XU, Y., UTESCHER, T. and DUPONT-NIVET, G. 2012. A late Eocene palynological record of climate change and Tibetan Plateau uplift (Xining Basin, China). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **344–345**, 16–38.

—— BOGOTÁ-A, G. R., ROMERO-BAEZ, M., LAMMERTSMA, E. I., FLANTUA, S. G. A., DANTAS, E. L., DINO, R., DO CARMO, D. A. and CHEMALE, F. 2017. The Amazon at sea: onset and stages of the Amazon River from a marine record, with special reference to Neogene plant turnover in the drainage basin. *Global & Planetary Change*, **153**, 51–65.

HORTON, B. K., YIN, A., SPURLIN, M. S., ZHOU, J. and WANG, J. 2002. Paleocene-Eocene syncontractional sedimentation in narrow, lacustrine-dominated basins of east-central Tibet. *Geological Society of America Bulletin*, **114**, 771–786.

HOUBEN, A. J. P., BIJL, P. K., GUERSTEIN, G. R., SLUIJS, A. and BRINKHUIS, H. 2011. Malvinia escutiana, a new biostratigraphically important Oligocene dinoflagellate cyst from the Southern Ocean. *Review of Palaeobotany & Palynology*, **165**, 175–182.

JARAMILLO, C. A., RUEDA, M. J. and MORA, G. 2006. Cenozoic plant diversity in the Neotropics. *Science*, **311**, 1893–1896.

JARDINE, P. E., FRASER, W. T., LOMAX, B. H. and GOSLING, W. D. 2015. The impact of oxidation on spore and pollen chemistry. *Journal of Micropalaeontology*, **34**, 139–149.

—— —— —— SEPHTON, M. A., SHANAHAN, T. M., MILLER, C. S. and GOSLING, W. D. 2016. Pollen and spores as biological recorders of past ultraviolet irradiance. *Scientific Reports*, **6**, 1–8.

—— ABERNETHY, F. A. J., LOMAX, B. H., GOSLING, W. D. and FRASER, W. T. 2017. Shedding light on sporopollenin chemistry, with reference to UV reconstructions. *Review of Palaeobotany & Palynology*, **238**, 1–6.

—— HARRINGTON, G. J., SESSA, J. A. and DAŠKOVÁ, J. 2018. Drivers and constraints on floral latitudinal diversification gradients. *Journal of Biogeography*, **45**, 1408–1419.

—— GOSLING, W. D., LOMAX, B. H., JULIER, A. C. M. and FRASER, W. T. 2019. Chemotaxonomy of domesticated grasses: a pathway to understanding the origins of agriculture. *Journal of Micropalaeontology*, **38**, 83–95.

—— FRASER, W. T., GOSLING, W. D., ROBERTS, C. N., EASTWOOD, W. J. and LOMAX, B. H. 2020*a*. Proxy reconstruction of ultraviolet-B irradiance at the Earth's surface, and its relationship with solar activity and ozone thickness. *The Holocene*, **30**, 155–161.

—— HOORN, C., BEER, M. A. M., BARBOLINI, N., WOUTERSEN, A., BOGOTA-ANGEL, G., GOSLING, W. D., FRASER, W. T., LOMAX, B. H., HUANG, H., SCIUMBATA, M., HE, H. and DUPONT-NIVET, G. 2020*b*. Data and code from "Sporopollenin chemistry and its durability in the geological record: an integration of extant and fossil chemical data across the seed plants". *Figshare*. https://doi.org/10.6084/m9.figshare.11382102

JULIER, A. C. M., JARDINE, P. E., COE, A. L., GOSLING, W. D., LOMAX, B. H. and FRASER, W. T. 2016. Chemotaxonomy as a tool for interpreting the cryptic diversity of Poaceae pollen. *Review of Palaeobotany & Palynology*, **235**, 140–147.

KIRSCHNER, J. A. and HOORN, C. 2020. The onset of grasses in the Amazon drainage basin, evidence from the fossil record. *Frontiers of Biogeography*, **12**(2).

KOHLER, A., BÖCKER, U., WARRINGER, J., BLOMBERG, A., OMHOLT, S. W., STARK, E. and MARTENS, H. 2009. Reducing inter-replicate variation in Fourier transform infrared spectroscopy by extended multiplicative signal correction. *Applied Spectroscopy*, **63**, 296–305.

KUHN, M. 2020. caret: Classification and Regression Training. R package version 6.0-85. https://CRAN.R-project.org/package=caret

LI, F. S., PHYO, P., JACOBOWITZ, J., HONG, M. and WENG, J. K. 2019. The molecular structure of plant sporopollenin. *Nature Plants*, **5**, 41–46.

LEEUW, J. W. DE, VERSTEEGH, G. J. M. and VAN BERGEN, P. F. 2006. Biomacromolecules of algae and plants and their fossil analogues. *Plant Ecology*, **182**, 209–233.

LILAND, K. H. 2017. EMSC: Extended Multiplicative Signal Correction. R package version 0.9.0. https://CRAN.R-project.org/package=EMSC

—— and MEVIK, B.-H. 2015. baseline: Baseline Correction of Spectra. R package version 1.2-1. https://CRAN.R-project.org/package=baseline

—— KOHLER, A. and AFSETH, N. K. 2016. Model-based pre-processing in Raman spectroscopy of biological samples. *Journal of Raman Spectroscopy*, **47**, 643–650.

LOADER, N. J. and HEMMING, D. L. 2000. Preparation of pollen for stable carbon isotope analyses. *Chemical Geology*, **165**, 339–344.

LOMAX, B. H. and FRASER, W. T. 2015. Palaeoproxies: botanical monitors and recorders of atmospheric change. *Palaeontology*, **58**, 759–768.

—— —— SEPHTON, M. A., CALLAGHAN, T. V., SELF, S., HARFOOT, M., PYLE, J. A., WELLMAN, C. H. and BEERLING, D. J. 2008. Plant spore walls as a record of long-term changes in ultraviolet-B radiation. *Nature Geoscience*, **1**, 592–596.

—— —— HARRINGTON, G., BLACKMORE, S., SEPHTON, M. A. and HARRIS, N. B. W. 2012. A novel palaeoaltimetry proxy based on spore and pollen wall chemistry. *Earth & Planetary Science Letters*, **353–354**, 22–28.

LOUGHLIN, N. J. D., GOSLING, W. D., MOTHES, P. and MONTOYA, E. 2018. Ecological consequences of post-Columbian indigenous depopulation in the Andean-Amazonian corridor. *Nature Ecology & Evolution*, **2**, 1233–1236.

LUTZKE, A., MOREY, K. J., MEDFORD, J. I. and KIPPER, M. J. 2020. Detailed characterization of *Pinus ponderosa* sporopollenin by infrared spectroscopy. *Phytochemistry*, **170**, 112195.

MANDER, L., LI, M., MIO, W., FOWLKES, C. C. and PUNYASENA, S. W. 2013. Classification of grass pollen through the quantitative analysis of surface ornamentation and texture. *Proceedings of the Royal Society B*, **280**, 20131905.

MAYO, D. W., MILLER, F. A. and HANNAH, R. W. 2003. *Course notes on the interpretation of infrared and Raman spectra*. John Wiley & Sons, 583 pp.

MEIJER, N., DUPONT-NIVET, G., ABELS, H. A., KAYA, M. Y., LICHT, A., XIAO, M., ZHANG, Y., ROPERCH, P., POUJOL, M., LAI, Z. and GUO, Z. 2019. Central Asian moisture modulated by proto-Paratethys Sea incursions since the early Eocene. *Earth & Planetary Science Letters*, **510**, 73–84.

MELLES, M., BRIGHAM-GRETTE, J., MINYUK, P. S., NOWACZY, N. R., WENNRICH, V., DECONTO, R. M., ANDERSON, P. M., ANDREEV, A. A., COLETTI, A., COOK, T. L., HALTIA-HOVI, E., KUKKONEN, M., LOZHKIN, A. V., ROSÉN, P., TARASOV, P., VOGEL, H. and WAGNER, B. 2012. 2.8 million years of arctic climate change from Lake El'gygytgyn, NE Russia. *Science*, **337**, 315–320.

MILLER, C. S. and GOSLING, W. D. 2014. Quaternary forest associations in lowland tropical West Africa. *Quaternary Science Reviews*, **84**, 7–25.

—— —— KEMP, D. B., COE, A. L. and GILMOUR, I. 2016. Drivers of ecosystem and climate change in tropical West Africa over the past ~540 000 years. *Journal of Quaternary Science*, **31**, 671–677.

NELSON, D. M., HU, F. S. and MICHENER, R. H. 2006. Stable-carbon isotope composition of Poaceae pollen: an assessment for reconstructing C_3 and C_4 grass abundance. *The Holocene*, **16**, 819–825.

—— —— SCHOLES, D. R., JOSHI, N. and PEARSON, A. 2008. Using SPIRAL (Single Pollen Isotope Ratio AnaLysis) to estimate C_3- and C_4-grass abundance in the paleorecord. *Earth & Planetary Science Letters*, **269**, 11–16.

NIEROP, K. G. J., VERSTEEGH, G. J. M., FILLEY, T. R. and DE LEEUW, J. W. 2019. Quantitative analysis of diverse sporomorph-derived sporopollenins. *Phytochemistry*, **162**, 207–215.

PAPPAS, C. S., TARANTILIS, P. A., HARIZANIS, P. C. and POLISSIOU, M. G. 2003. New method for pollen identification by FT-IR spectroscopy. *Applied Spectroscopy*, **57**, 23–27.

PIPERNO, D. R., WEISS, E., HOLST, I. and NADEL, D. 2004. Processing of wild cereal grains in the Upper Palaeolithic revealed by starch grain analysis. *Nature*, **430**, 670–673.

PROSS, J., CONTRERAS, L., BIJL, P. K., GREENWOOD, D. R., BOHATY, S. M., SCHOUTEN, S., BENDLE, J. A., ROHL, U., TAUXE, L., RAINE, J. I., HUCK, C. E., VAN DE FLIERDT, T., JAMIESON, S. S. R., STICKLEY, C. E., VAN DE SCHOOTBRUGGE, B., ESCUTIA, C., BRINKHUIS, H. and THE INTEGRATED OCEAN DRILLING PROGRAM

EXPEDITION 318 SCIENTISTS. 2012. Persistent near-tropical warmth on the Antarctic continent during the early Eocene epoch. *Nature*, **488**, 73–77.

R CORE TEAM. 2018. R: A language and environment for statistical computing. Version 3.5.2. http://www.R-project.org

ROZEMA, J., VAN DE STAAIJ, J., BJÖRN, L.-O. and DE BAKKER, N. 1999. Depletion of stratospheric ozone and solar UV-B radiation: evolution of land plants, UV-screens and functions of polyphenolics. 1–19. *In* ROZEMA, J. (ed.) *Stratospheric ozone depletion: The effects of enhanced UV-B radiation on terrestrial ecosystems*. Backhuys, 344 pp.

—— BROEKMAN, R. A., BLOKKER, P., MEIJKAMP, B., DE BAKKER, N., VAN DE STAAIJ, J., VAN BEEM, A., ARIESE, F. and KARS, S. M. 2001*a*. 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 & Photobiology B*, **62**, 108–117.

—— NOORDJIK, A. J., BROEKMAN, R. A., VAN BEEM, A., MEIJKAMP, B. M., DE BAKKER, N. V. J., VAN DE STAAIJ, J. W. M., STROETENGA, M., BOHNCKE, S. J. P., KONERT, M., KARS, S., PEAT, H., SMITH, R. I. L. and CONVEY, P. 2001*b*. (Poly)phenolic compounds in pollen and spores of Antarctic plants as indicators of UV-B: a new proxy for the reconstruction of past solar UV-B? *Plant Ecology*, **154**, 11–26.

—— VAN GEEL, B., BJÖRN, L.-O., LEAN, J. and MADRONICH, S. 2002. Toward solving the UV puzzle. *Science*, **296**, 1621–1622.

—— BLOKKER, P., MAYORAL FUERTES, M. A. and BROEKMAN, R. 2009. UV-B absorbing compounds in present-day and fossil pollen, spores, cuticles, seed coats and wood: evaluation of a proxy for solar UV radiation. *Photochemical & Photobiological Sciences*, **8**, 1233–1243.

SCHULTE, F., LINGOTT, J., PANNE, U. and KNEIPP, J. 2008. Chemical characterization and classification of pollen. *Analytical Chemistry*, **80**, 9551–9556.

SEDDON, A. W. R., JOKERUD, M., BARTH, T., BIRKS, H. J. B., KRÜGER, L. C., VANDVIK, V. and WILLIS, K. J. 2017. Improved quantification of UV-B absorbing compounds in *Pinus sylvestris* L. pollen grains using an internal standard methodology. *Review of Palaeobotany & Palynology*, **247**, 97–104.

—— FESTI, D., ROBSON, T. M. and ZIMMERMANN, B. 2019. Fossil pollen and spores as a tool for reconstructing ancient solar-ultraviolet irradiance received by plants: an assessment of prospects and challenges using proxy-system modelling. *Photochemical & Photobiological Sciences*, **18**, 275–294.

SLATER, S. M., WELLMAN, C. H. and LOMAX, B. 2016. Middle Jurassic vegetation dynamics based on quantitative analysis of spore/pollen assemblages from the Ravenscar Group, North Yorkshire, UK. *Palaeontology*, **59**, 305–328.

SOUTH, A. 2011. rworldmap: a new R package for mapping global data. *The R Journal*, **3**, 35–43.

STEEMANS, P., LEPOT, K., MARSHALL, C. P., LE HERISSE, A. and JAVAUX, E. J. 2010. FTIR characterisation of the chemical composition of Silurian miospores (cryptospore and trilete spores) from Gotland, Sweden. *Review of Palaeobotany & Palynology*, **162**, 577–590.

STRÖMBERG, C. A. E. 2011. Evolution of grasses and grassland ecosystems. *Annual Review of Earth & Planetary Sciences*, **39**, 517–544.

TORRES, V., VANDENBERGHE, J. and HOOGHIEMSTRA, H. 2005. An environmental reconstruction of the sediment infill of the Bogotá basin (Colombia) during the last 3 million years from abiotic and biotic proxies. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **226**, 127–148.

—— HOOGHIEMSTRA, H., LOURENS, L. and TZEDAKIS, P. C. 2013. Astronomical tuning of long pollen records reveals the dynamic history of montane biomes and lake levels in the tropical high Andes during the Quaternary. *Quaternary Science Reviews*, **63**, 59–72.

TRAVERSE, A. 2007. *Paleopalynology*. Springer, 813 pp.

VAJDA, V., PUCETAITE, M., MCLOUGHLIN, S., ENGDAHL, A., HEIMDAL, J. and UVDAL, P. 2017. Molecular signatures of fossil leaves provide unexpected new evidence for extinct plant relationships. *Nature Ecology & Evolution*, **1**, 1093–1099.

VARMUZA, K. and FILZMOSER, P. 2009. *Introduction to multivariate statistical analysis in chemometrics*. CRC Press, 336 pp.

VENABLES, W. N. and RIPLEY, B. D. 2002. *Modern applied statistics with S*. Springer, 498 pp.

VINK, J. 2012. A palynological study of Neogene fluvial sediments of SW Amazonia (Madre de Dios foreland basin, Peru) with inferences for their use in paleo-biodiversity studies. Unpublished BSc thesis, University of Amsterdam.

WATSON, J. S., SEPTHON, M. A., SEPHTON, S. V., SELF, S., FRASER, W. T., LOMAX, B. H., GILMOUR, I., WELLMAN, C. H. and BEERLING, D. J. 2007. Rapid determination of spore chemistry using thermochemolysis gas chromatography-mass spectrometry and micro-Fourier transform infrared spectroscopy. *Photochemical & Photobiological Sciences*, **6**, 689–694.

—— FRASER, W. T. and SEPHTON, M. A. 2012. Formation of a polyalkyl macromolecule from the hydrolysable component within sporopollenin during heating/pyrolysis experiments with Lycopodium spores. *Journal of Analytical & Applied Pyrolysis*, **95**, 138–144.

WEI, T. and SIMKO, V. 2017. corrplot: Visualisation of a Correlation Matrix. R package version 0.84. https://CRAN.R-project.org/package=corrplot

WEISS, E., WETTERSTROM, W., NADEL, D. and BAR-YOSEF, O. 2004. The broad spectrum revisited: evidence from plant remains. *Proceedings of the National Academy of Sciences*, **101**, 9551–9555.

WELLMAN, C. H., OSTERLOFF, P. L. and MOHIUDDIN, U. 2003. Fragments of the earliest land plants. *Nature*, **425**, 282–285.

WILLIS, K. J., FEURDEAN, A., BIRKS, H. J. B., BJUNE, A. E., BREMAN, E., BROEKMAN, R., GRYTNES, J. A., NEW, M., SINGARAYER, J. S. and ROZEMA, J. 2011. Quantification of UV-B flux through time using UV-B-absorbing compounds contained in fossil Pinus sporopollenin. *New Phytologist*, **192**, 553–560.

WOUTERSEN, A., JARDINE, P. E., BOGOTÁ-ANGEL, G., ZHANG, H.-X., SILVESTRO, D., ANTONELLI, A., GOGNA, E., ERKENS, R. H. J., GOSLING, W. D., DUPONT-NIVET, G. and HOORN, C. 2018. A novel approach to study the morphology and chemistry of pollen in a phylogenetic context, applied to the steppe-desert taxon *Nitraria* L. (Nitrariaceae). *PeerJ*, **6**, e5055.

YULE, B. L., ROBERTS, S. and MARSHALL, J. E. A. 2000. The thermal evolution of sporopollenin. *Organic Geochemistry*, **31**, 859–870.

ZIMMERMANN, B. 2010. Characterization of pollen by vibrational spectroscopy. *Applied Spectroscopy*, **64**, 1364–1373.

—— 2018. Chemical characterization and identification of Pinaceae pollen by infrared microspectroscopy. *Planta*, **247**, 171–180.

—— and KOHLER, A. 2014. Infrared spectroscopy of pollen identifies plant species and genus as well as environmental conditions. *PLoS One*, **9**, 1–12.

—— BAGCIOGLU, M., SANDT, C. and KOHLER, A. 2015*a*. Vibrational microspectroscopy enables chemical characterization of single pollen grains as well as comparative analysis of plant species based on pollen ultrastructure. *Planta*, **242**, 1237–1250.

—— TKALČEC, Z., MEŠIĆ, A. and KOHLER, A. 2015*b*. Characterizing aeroallergens by infrared spectroscopy of fungal spores and pollen. *PLoS One*, **10**, 1–22.

—— TAFINTSEVA, V., BAĞCIOĞLU, M., HØEGH BERDAHL, M. and KOHLER, A. 2016. Analysis of allergenic pollen by FTIR microspectroscopy. *Analytical Chemistry*, **88**, 803–811.