Uncovering a phylogenetic signal in plant biopolymer chemistry: a comparison of sporopollenin isolation approaches for use in palynological research

by PHILLIP E. JARDINE[1]* (iD), MATTHEW S. KENT[2] (iD), WESLEY T. FRASER[3] (iD), KLAUS-HOLGER KNORR[4] (iD) *and* BARRY H. LOMAX[2] (iD)

[1]Palaeobotany Group, Institute of Geology and Palaeontology, University of Münster, 48149 Münster, Germany; jardine@uni-muenster.de
[2]School of Biosciences, University of Nottingham, Sutton Bonington Campus, Leicestershire LE12 5RD, UK
[3]Geography, School of Social Sciences, Oxford Brookes University, Headington Campus, Oxford OX3 0BP, UK
[4]Ecohydrology & Biogeochemistry Group, Institute of Landscape Ecology, University of Münster, 48149 Münster, Germany
*Corresponding author

Abstract: Sporomorphs (pollen and spores) are a mainstay of research into past vegetation, and increasingly sporomorph chemistry is being used as a palaeoecological tool. To make extant sporomorphs directly comparable to fossil specimens, fresh material is processed to remove labile compounds and isolate the sporopollenin wall. A range of processing approaches are currently in use, but the chemistries produced by these different techniques have not yet been compared across a range of taxa. It is therefore not clear how they compare in terms of efficiently isolating sporopollenin without changing its chemical structure, and what impact they have on relative chemical similarities and differences among taxa (i.e. whether more closely related species will always appear chemically more similar, regardless of how they have been processed). Here, we test this by applying five different processing approaches to sporomorphs from 15 taxa from across the vascular plant phylogeny. We show that each approach has its own idiosyncrasies in terms of impacts on sporomorph chemistry. For the most part a common pattern of among-taxon chemical variability is uncovered, and a phylogenetic signal within sporopollenin chemistry is supported. Working with spectral derivatives generally increases agreement among the different processing approaches, but decreases the strength of the phylogenetic signal. No one processing approach is ideal, and the choice of which to use is likely to depend on the goal of the study, the type and quantity of material being processed, and the laboratory facilities available for processing.

Key words: sporopollenin, phylogeny, chemistry, biopolymer, Fourier transform infrared (FTIR) microspectroscopy.

SPOROMORPHS (pollen and spores) are reproductive vectors of land plants (Traverse 2007). They are a cornerstone of research into past vegetation, environmental and climate change, pollination ecology and aerobiology, and are used as tools in biostratigraphy, melissopalynology (for understanding the botanical provenance of honey) and forensics (Moore *et al.* 1991; Jansonius & MacGregor 1996; Traverse 2007). More recently, there has been research from a material science perspective into sporomorphs, with a focus on the physical and chemical structure of sporopollenin, the biopolymer that comprises the outer wall (exine) of sporomorphs, with a view to using sporomorphs as microparticles, microcapsules, and as a target for biomimetic material synthesis (Diego-Taboada *et al.* 2013; Mackenzie *et al.* 2015; Gonzalez-Cruz *et al.* 2018; Li *et al.* 2019; Cojocaru *et al.* 2022).

The analysis of sporopollenin chemistry is also a focus of palaeopalynological research. This is partly stimulated by an interest in the fate of organic matter in the geological record, and the repolymerization pathway of sporopollenin from a biopolymer to a geopolymer (Yule *et al.* 2000; de Leeuw *et al.* 2006; Watson *et al.* 2012; Fraser *et al.* 2014a; Bernard *et al.* 2015; Jardine *et al.* 2021), but also for the use of sporopollenin chemistry as a palaeoclimatic and palaeoecological tool. Specifically, recent research has focused on using sporopollenin chemistry as a proxy for reconstructing ultraviolet-B (UV-B) and total solar irradiance through time (Rozema *et al.* 2001a; 2002; 2009; Blokker *et al.* 2005; Watson *et al.* 2007; Lomax *et al.* 2008; Fraser *et al.* 2011; 2014b; Willis *et al.* 2011; Lomax & Fraser 2015; Jardine *et al.* 2016; 2017; 2020; Seddon *et al.* 2017; 2019; Bell *et al.* 2018; Liu *et al.* 2023), as a chemotaxonomic tool for classifying fossil

sporomorphs (Julier *et al.* 2016; Woutersen *et al.* 2018; Jardine *et al.* 2019; 2021; Muthreich *et al.* 2020) and, via stable carbon isotope analysis, reconstructing C_3 vs C_4 grass abundances (Nelson *et al.* 2006; 2014) and moisture availability (Griener *et al.* 2013; Bell *et al.* 2017; 2019), as well as identifying reworked sporomorphs in fossil samples (Korasidis *et al.* 2022). Key to these endeavours is understanding the evolution of sporopollenin, including its chemical disparity across the land plant phylogeny and through time, and the strength of a phylogenetic signal in sporopollenin chemistry vs evolutionary convergence (de Leeuw *et al.* 2006; Fraser *et al.* 2012; Woutersen *et al.* 2018; Nierop *et al.* 2019; Jardine *et al.* 2021).

Fresh sporomorphs (i.e. those collected from plants, rather than fossil sporomorphs from the geological record) comprise a mix of proteins, lipids and carbohydrates in addition to the sporopollenin-based exine (Zimmermann & Kohler 2014; Bağcıoğlu *et al.* 2015; Jardine *et al.* 2015; 2017; Zimmermann 2018; Muthreich *et al.* 2020). Chemical analysis of sporopollenin, or the use of sporopollenin 'shells' as microparticles or microcapsules, therefore requires the removal of these labile compounds to isolate the sporopollenin for subsequent investigation or use (Domínguez *et al.* 1998; 1999; Ahlers *et al.* 1999; Loader & Hemming 2000; Jardine *et al.* 2015; Mundargi *et al.* 2016; Gonzalez-Cruz *et al.* 2018; Li *et al.* 2019; Lutzke *et al.* 2020). Consequently, a range of processing methods have been developed for sporopollenin isolation, with

differences in the reagents used, the durations and number of steps required, and the equipment needed (Table 1).

For example, chemical analysis of sporomorphs with a view to palaeoproxy development and chemotaxonomic applications have often used simple and quick methods developed for routine morphological palynological research, such as acetolysis and, to a lesser extent, treatment with potassium hydroxide (KOH) (Rozema *et al.* 2001a, 2001b; Blokker *et al.* 2005; Jardine *et al.* 2015; 2016; 2017; 2020; 2021; Woutersen *et al.* 2018). For carbon isotope analysis of sporomorphs, sulfuric acid (H_2SO_4) rather than acetolysis has been used to isolate sporopollenin, to avoid contamination of the $\delta^{13}C$ signal by carbon-containing chemicals (Loader & Hemming 2000; Bell *et al.* 2017). Longer procedures involving multiple reagents and steps have typically been used for microencapsulation applications and research into the chemical composition of sporopollenin, where clean and unaltered sporopollenin is needed and bulk quantities (typically 100s of mg to 10s of g) of sporomorphs are processed (Domínguez *et al.* 1998, 1999; Gonzalez-Cruz *et al.* 2018; Li *et al.* 2019; Uddin *et al.* 2019; Lutzke *et al.* 2020) (Table 1).

While these various methods have the same overall aim (that of removing the non-sporopollenin components of sporomorphs and leaving isolated sporopollenin behind) they have different chemical effects, with the potential to either not fully remove the labile compounds, or to alter the chemical signature of the isolated sporopollenin.

TABLE 1. Overview of sporopollenin isolation approaches, with examples of use in the published literature.

Treatment	Steps involved	Applications and references
Acetolysis	Treatment with 9:1 mix of acetic anhydride ([CH$_3$CO]$_2$O) and sulfuric acid (H$_2$SO$_4$), typically 3–10 min at >80°C	Palaeoproxy development, chemotaxonomic applications (Rozema *et al.* 2001a; Blokker *et al.* 2005; Jardine *et al.* 2015; 2021)
Potassium hydroxide (KOH)	Treatment with 10% KOH, typically *c.* 1 h at >80°C	Palaeoproxy development (Blokker *et al.* 2005)
Sulfuric acid (H$_2$SO$_4$)	Treatment with 96% H$_2$SO$_4$, typically *c.* 1 h at room temperature	Carbon isotope analysis of sporomorphs (Loader & Hemming 2000; Bell *et al.* 2017)
Enzymes and solvents	Treatment with enzymes to remove carbohydrates, and a sequence of solvents of increasing polarity to remove proteins and lipids	Sporopollenin chemical composition and structure (Ahlers *et al.* 1999; Li *et al.* 2019; Lutzke *et al.* 2020)
Solvents, alkalis and acids	Treatment with acetone ([CH3]$_2$CO) to remove lipids, potassium hydroxide (KOH) to remove proteins, and phosphoric acid (H$_3$PO$_4$) to remove carbohydrates	Microencapsulation (Gonzalez-Cruz *et al.* 2018; Uddin *et al.* 2019). It has been suggested that the alkali and acid steps should be switched (Gonzalez-Cruz *et al.* 2018) or the alkali step left out altogether (Mundargi *et al.* 2016)
Dichloromethane (DCM) and methanol (MeOH)	Soxhlet extraction with DCM/MeOH mixture	Sporopollenin chemical composition, isolation as prelude to thermal maturation experiments (Watson *et al.* 2012; Fraser *et al.* 2014a; Nierop *et al.* 2019)
Anhydrous hydrogen fluoride	Treatment with hydrogen fluoride in pyridine (HF-Py) for 5 h at 40°C	Sporopollenin chemical composition and structure (Domínguez *et al.* 1998, 1999)

Acetolysis, for example, is known to leave a chemical trace in processed samples by adding carbon-containing compounds, as well as removing some phenolic components from sporopollenin (Amundson *et al.* 1997; Domínguez *et al.* 1998; Loader & Hemming 2000; Blokker *et al.* 2005; Jardine *et al.* 2015; 2021; Lutzke *et al.* 2020). Lutzke *et al.* (2020) found that treatment with enzymes and solvents failed to remove all proteins during processing, with samples needing an additional treatment with KOH to fully isolate the sporopollenin. Lutzke *et al.* (2020) also found that treatment with phosphoric acid, which is commonly used as part of a sequence of treatments to isolate sporopollenin for microencapsulation (Gonzalez-Cruz *et al.* 2018; Uddin *et al.* 2019) altered sporopollenin chemistry to some extent, by partially removing aromatic components.

For palaeopalynological applications investigating either the chemical evolution of sporopollenin, or seeking to classify fossil sporomorphs based on their chemical signature, a key question is what impact sample processing choices have on the results obtained. In particular, it is essential to understand how relative inter-taxon chemical differences vary among processing approaches, whether the magnitude of these taxon-to-taxon chemical differences changes substantially, and how they compare with phylogenetic inter-taxon distances (i.e. the total length of the branches connecting species in a phylogeny through their most recent common ancestor as a measure of relatedness; Faith 1992; Verdú *et al.* 2012). The strength of a phylogenetic signal in sporopollenin, and its maintenance with processing, is important not only for understanding sporopollenin chemical evolution, but also for classification problems. A strong phylogenetic signal can be leveraged to classify sporomorphs at higher and more inclusive taxonomic levels, removing the need to focus on species-level classifications that are likely to be impractical in highly diverse taxa such as grasses (Julier *et al.* 2016; Jardine *et al.* 2019), or in deep time investigations (Jardine *et al.* 2021).

Here, we investigate these issues using sporomorphs from a broad phylogenetic range of plant taxa, whilst subjecting them to several different processing techniques. We document the chemical changes that occur with each processing approach, and the relative similarities and differences among taxa following processing, using Fourier transform infrared (FTIR) spectroscopy. Fourier Transform infrared spectroscopy is a vibrational spectroscopic technique that allows for the rapid and non-destructive characterization of extant and fossil sporomorph chemistry, and as such it has become a standard tool for biogeochemical measurement in (palaeo)palynological studies (e.g. Steemans *et al.* 2010; Zimmermann 2010; Fraser *et al.* 2012; Zimmermann & Kohler 2014; Jardine *et al.* 2017; 2020; Muthreich *et al.* 2020; Liu *et al.* 2023). Since FTIR provides information on the different components of sporomorphs (i.e. the sporopollenin wall and labile compounds such as proteins, lipids and carbohydrates) (Zimmermann & Kohler 2014; Bağcıoğlu *et al.* 2015; Jardine *et al.* 2015), it is appropriate for tracking chemical changes associated with different sporopollenin isolation techniques. We also study the impact of spectral processing measures (smoothing and taking derivatives), and the relationship between sporopollenin chemistry and phylogeny.

MATERIAL AND METHOD

The sample set comprises 15 taxa, which were selected to provide a broad phylogenetic range, incorporating angiosperms (10 species), gymnosperms (4 species) and a lycopod (1 species) (Fig. 1; Table 2). For the angiosperms, four species are from the superasterid clade, four are superrosids, one (*Platanus hispanica*) is a basal eudicot, and one (*Secale cereale*) is a monocot (Table 2). The angiosperm and gymnosperm samples were purchased from Pharmallerga (Lisov, Czechia), and the *Lycopodium clavatum* spores were purchased from Sigma-Aldrich (Gillingham, UK). Samples were purchased rather than collected from nature to ensure both uniformity and sufficient material to perform all processing techniques.

We selected a range of sporomorph processing approaches to provide a broad comparison of methods and techniques. We used a combined enzyme and solvent approach (hereafter enzymes/solvents) because while this is not routinely applied in palaeopalynological research, it has been used for researching the chemical composition of sporopollenin (Ahlers *et al.* 1999; Li *et al.* 2019; Lutzke *et al.* 2020) and should remove all non-sporopollenin compounds without changing the sporopollenin chemical signature (Lutzke *et al.* 2020). Acetolysis and KOH were used because they are common in palaeopalynological research (Traverse 2007), and H_2SO_4 because it is routinely used in pollen carbon isotope studies (Loader & Hemming 2000; Bell *et al.* 2017). We also applied the anhydrous hydrogen fluoride treatment introduced by Domínguez *et al.* (1998), which uses hydrogen fluoride in pyridine (HF-Py), because it has previously been used to study the chemical composition of sporopollenin (Domínguez *et al.* 1998, 1999), and hydrofluoric acid (HF) is routinely handled in many palynology labs. Palynologists are likely to be able to make use of this approach, despite it not yet being widely adopted by palaeopalynologists.

For each treatment 1 mL of sporomorphs was processed, except for the enzymes/solvents treatment where 125 mg of sporomorphs were used. In each case the samples were processed in 15 mL centrifuge tubes, and deionized water was used throughout.
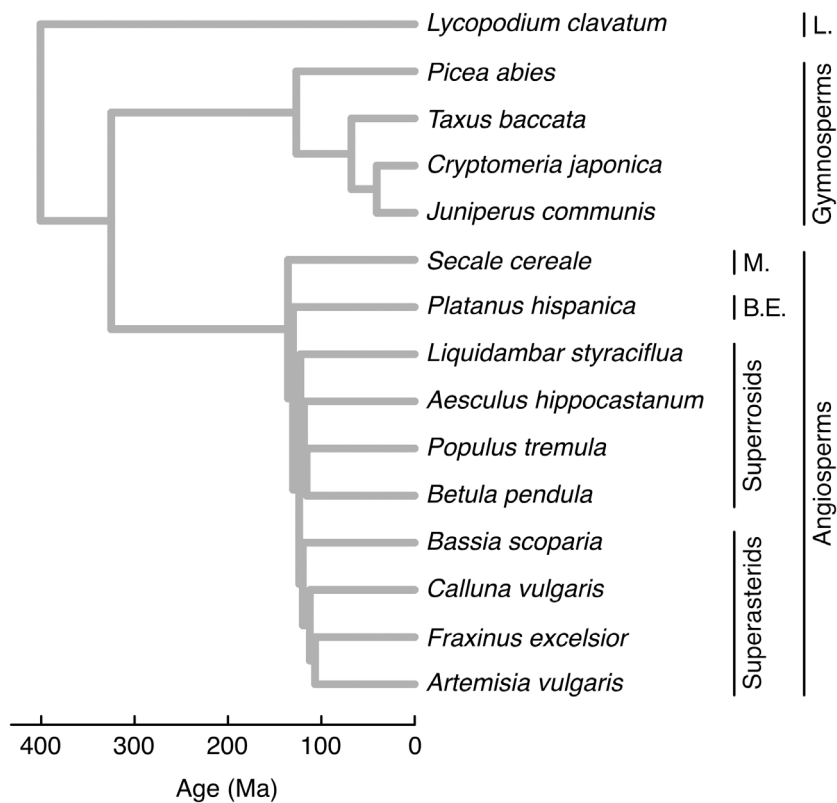
FIG. 1. Molecular phylogeny of the taxa included in this study, showing the main clades (see also Table 2). *Abbreviations*: L., Lycopods; M., Monocots, B.E., basal Eudicots.

For the enzymes/solvents treatment we adapted the protocols outlined in Ahlers *et al.* (1999), Li *et al.* (2019) and Lutzke *et al.* (2020). The sporomorph samples were first suspended in 5 mL of 0.1 M sodium acetate buffer (pH 4.5) containing 1% w/v Cellulase Onozuka R-10 and Macerozyme R-10 for 72 h at 30°C, with regular stirring throughout. The samples were then centrifuged and washed five times with 5 mL of water, then left in water for 24 h at 30°C. Following this, the samples were again centrifuged and washed five times with 5 mL of water; thereafter they were washed three times with 2 mL of 100% methanol, and three times with 2 mL of 100%

TABLE 2. Species used in this study; Clade 1 and Clade 2 are suprafamilial taxa that incorporate the focal species.

Species	Family	Clade 1	Clade 2
Aesculus hippocastanum	Sapindaceae	Superrosids	Angiosperms
Artemisia vulgaris	Asteraceae	Superasterids	Angiosperms
Bassia scoparia	Amaranthaceae	Superasterids	Angiosperms
Betula pendula	Betulaceae	Superrosids	Angiosperms
Calluna vulgaris	Ericaceae	Superasterids	Angiosperms
Cryptomeria japonica	Cupressaceae	Gymnosperms	Gymnosperms
Fraxinus excelsior	Oleaceae	Superasterids	Angiosperms
Juniperus communis	Cupressaceae	Gymnosperms	Gymnosperms
Liquidambar styraciflua	Altingiaceae	Superrosids	Angiosperms
Lycopodium clavatum	Lycopodiaceae	Lycopods	Lycopods
Picea abies	Pinaceae	Gymnosperms	Gymnosperms
Platanus hispanica	Platanaceae	Basal eudicots	Angiosperms
Populus tremula	Salicaceae	Superrosids	Angiosperms
Secale cereale	Poaceae	Monocots	Angiosperms
Taxus baccata	Taxaceae	Gymnosperms	Gymnosperms

diethyl ether. This was followed by six 24-hour treatments with solvents of increasing polarity: 100% diethyl ether, 100% chloroform, ≥99.5% acetone, 100% methanol, 100% 2-ethoxyethanol, and water (5 mL in each case). Between treatments, the sporomorph suspension was centrifuged and washed three times with 2 mL of the extraction solvent, followed by two washes with 2 mL of the following solvent in the sequence. Following the extraction steps, the samples were washed fives time with 5 mL of water, incubated in water at 40°C for 72 h, centrifuged, and finally washed three times with 5 mL of water.

For the acetolysis treatment the samples were acetolysed in 3 mL of a 9:1 mixture of ≥99% acetic anhydride ($[CH_3CO]_2O$) and 96% sulfuric acid (H_2SO_4) for 10 min at 90°C in a water bath, followed by one wash with 100% glacial acetic acid (CH_3COOH) and two washes with water, with centrifugation and decanting in between washes.

For the H_2SO_4 treatment (Loader & Hemming 2000) the samples were processed in 3 mL 96% H_2SO_4 for 1 h at room temperature. A complicating factor when working with H_2SO_4 is its higher density compared to most other reagents, which makes centrifugation challenging. Therefore, following the treatment the samples were transferred to 50 mL tubes before being topped up with water, mixed, centrifuged and decanted, followed by repeated washing and centrifugation until the samples were pH neutral (*c.* 8 washes needed).

For the anhydrous hydrogen fluoride treatment (Domínguez *et al.* 1998), samples were treated in 3 mL HF-Py (70% hydrogen fluoride, 30% pyridine) for 5 h at 40°C in a water bath. The samples were stirred several times during treatment. Following treatment, the sample tubes were topped up with water, centrifuged and decanted, followed by repeated washing and centrifugation until the samples were pH neutral.

For the KOH treatment, the samples were processed in 3 mL 10% KOH for 1 h at 90°C in a water bath, followed by repeated washing and centrifugation until the samples were pH neutral. Following all treatments, the samples were washed twice with ethanol, centrifuged, decanted and left to air dry.

IR spectra were generated using ATR (attenuated total reflectance) FTIR, because this approach is suitable for generating bulk scans from sporomorph samples, and provides high quality spectra without the distortions and scattering effects that are often produced in transmission-based IR analyses of a limited number of grains (Zimmermann *et al.* 2016; Jardine *et al.* 2021). We generated the data using an Agilent Cary 670 FTIR spectrometer, with a KBr beamsplitter and a thermo-electrically cooled DLaTGS detector, fitted with a Pike Technologies (Madison WI, USA) MIRacle universal ATR accessory with a single reflection ZnSe ATR crystal, housed in the Institute

of Landscape Ecology at the University of Münster. The IR and ATR units were purged with compressed air to limit the impact of variations in atmospheric CO_2 and H_2O on the measurements. FTIR spectra were generated using 32 scans per sample and a resolution of 4 cm^{-1}, with a resultant spectral data interval (i.e. the spacing between data points in the FTIR spectra) of 1.93 cm^{-1}. We generated six replicate measurements per taxon and treatment, resulting in 15 taxa × 6 treatments × 6 replicate measurements = 540 measurements in total. Before each set of six replicate measurements a background spectrum was recorded using the empty ATR crystal, and
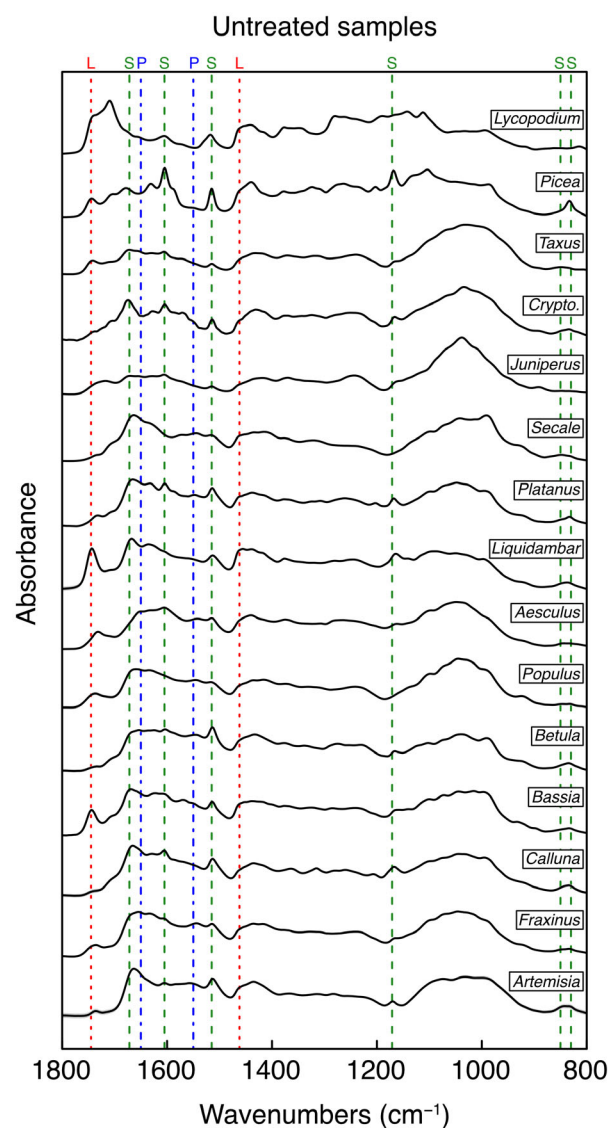

FIG. 2. Stacked mean FTIR spectra for each taxon in the untreated sample set. Grey shaded regions show ±1 standard deviation about the mean. Key peaks are marked with dotted/dashed lines. *Abbreviations*: L, lipids; S, sporopollenin; P, proteins (see also Table 3); *Crypto.*, *Cryptomeria*.

automatically removed from the sample measurements during data generation.

Prior to analysis the FTIR spectra were baseline corrected with a second-order polynomial baseline, and truncated to the 1800 to 800 cm^{-1} region to focus on the main among-taxon chemical variations (Jardine et al. 2021). To remove variations in absolute absorbance values the truncated spectra were standardized using standard normal variates (SNV), also known as z-score standardization, by subtracting the spectrum mean and dividing by the standard deviation (Varmuza & Filzmoser 2009). This results in spectra with a mean of zero and variance of one. Peaks were identified and linked to the primary components of sporomorphs (sporopollenin, proteins, lipids, carbohydrates) using the published literature (Coates 2000; Mayo 2003; Watson et al. 2007; Zimmermann & Kohler 2014; Fraser et al. 2014a; Bağcıoğlu et al. 2015; Zimmermann 2018; Lutzke et al. 2020).

For each taxon × treatment combination, the mean FTIR spectrum was calculated and used in further analyses. We used principal components analysis (PCA) to visualize the inter-taxon chemical relationships for each treatment, because this approach reduces the number of dimensions of complex multivariate data, allowing the main variation in the dataset to be visualized on a limited number of axes (Varmuza & Filzmoser 2009). The inter-taxon Euclidean distances were compared across treatments, and with the phylogenetic distances between taxa, using pairs plots with correlation coefficients, and box plots (Zuur et al. 2009). Transforming FTIR spectra to derivatives via the Savitzky–Golay smoothing algorithm is a common step prior to multivariate analysis and classification (e.g. Julier et al. 2016; Zimmermann et al. 2016; Muthreich et al. 2020): the use of derivatized spectra emphasizes smaller-scale spectral features that can be useful for classification, but also increases the influence of random noise in the spectra, which is counteracted by smoothing as part of the Savitzky–Golay algorithm (Varmuza & Filzmoser 2009). Since this approach has the potential to alter the relative pairwise chemical distances between taxa and their arrangement in ordination space (Jardine et al. 2019), we re-ran all analyses on second derivative spectra calculated using the Savitzky–Golay algorithm with a second order polynomial and a smoothing window size of 21. This window size was chosen as an intermediate value from previous analyses of sporomorph chemistry where the spectral data interval is the same as that used here (i.e. 1.93 cm^{-1}), and where the smoothing window value has been optimized via the classification success rate, with values typically between 7 and 43 (Woutersen et al. 2018; Jardine et al. 2019; 2021). The phylogeny was generated using the phylo.maker function from the V.PhyloMaker2 package (Jin & Qian 2019; 2022) for R (R Core Team 2013), and phylogenetic distances were measured as the pairwise cophenetic distances calculated from the branches of the phylogeny (Faith 1992).

All data analysis was carried out in R (v4.2.1; R Core Team 2013), using RStudio (v2022.07.0+548; RStudio Team 2022) and the packages ape (v5.6-2; Paradis & Schliep 2019), baseline (v1.3-4; Liland et al. 2010), prospectr (v0.2.4; Stevens & Ramirez-Lopez 2022) and V.PhyloMaker2 (v0.1.0; Jin & Qian 2022). All data and code for recreating these analyses are available to download from figshare (Jardine et al. 2023).

RESULTS

Overview of untreated sporomorph chemistry

Taken together, the FTIR spectra (Fig. 2) for the untreated samples exhibit peaks associated with lipids, proteins, carbohydrates, and sporopollenin-bound phenolic compounds, as expected (Table 3). For example, lipid peaks occur at c. 1740 and 1460 cm^{-1}, proteins at c. 1650 and 1550 cm^{-1}, peaks associated with the carbohydrate intine (the inner pollen wall) between 1200 and 900 cm^{-1}, and sporopollenin peaks at c. 1680, 1605, 1515, 1170, 850 and 830 cm^{-1}. However, there are a number of taxon-specific differences in the heights of the various peaks, reflecting differences in the underlying contributions of respective structural moieties. For

TABLE 3. Band assignments for the main peaks in the ATR-FTIR spectra, based primarily on Lutzke et al. (2020) and Bağcıoğlu et al. (2015).

Band position (cm^{-1})	Assignment	Major component
1740	Ester ν(C=O)	Lipids
1710	Carboxylic acid ν(C=O)	Sporopollenin
1680	Ester ν(C=O)	Sporopollenin
1650	Amide I ν(C=O)	Proteins
1630	Alkenyl ν(C=C)	Sporopollenin
1605	Aromatic ring ν(C=C)	Sporopollenin
1580	Aromatic ring ν(C=C)	Sporopollenin
1550	Amide II δ(NH), ν(C-N)	Proteins
1515	Aromatic ring ν(C=C)	Sporopollenin
1460	$\delta_{scissors}$(CH$_2$)	Lipids
1170	Aromatic δ(CH), ν(CO)	Sporopollenin
1070	Polysaccharides	Carbohydrate intine
1030	Polysaccharides	Carbohydrate intine
1010	Polysaccharides	Carbohydrate intine
850	Aromatic γ(CH)	Sporopollenin
830	Aromatic γ(CH)	Sporopollenin

Note that the band positions are approximate, and vary among taxa and treatments.

example, *Bassia*, *Liquidambar* and *Lycopodium* have a more pronounced lipid peak at 1740 cm^{-1}, while this is barely expressed in *Calluna*, *Betula* and *Secale*, and is intermediate in height in the other taxa. The broad carbohydrate band centred on *c.* 1050 cm^{-1} is clearly present in most taxa but not expressed in *Picea* and *Lycopodium*. Phenolic peaks representing the sporopollenin wall are most clearly expressed in *Picea* but more variably in the other taxa. In general the protein-related amide I and II peaks at 1650 and 1550 cm^{-1}, respectively, are more clearly expressed in the angiosperms relative to the gymnosperms and *Lycopodium*.

Impact of processing on sporomorph chemistry

All the processing approaches have the anticipated effect of removing labile compounds, leaving sporopollenin behind and thus making the sporopollenin-related peaks in FTIR spectra more prominent. There are, however, substantial differences in how completely the non-sporopollenin peaks are removed, and what impact the processing approaches have on the sporopollenin chemistry. For each processing technique, the main spectral changes are summarized below and in Table 4. Figures 3–7 show both the spectra for each taxon following treatment, and the difference between the treated and untreated spectra (one figure per treatment).

Enzymes/solvents. This treatment leads to a removal of the labile compounds, with the protein, lipid and carbohydrate peaks decreasing in size or disappearing entirely (Fig. 3). The carbohydrate band decreased relatively less in some taxa, such as the gymnosperms *Taxus*, *Cryptomeria* and *Juniperus*, suggesting that the enzyme treatment may not have been similarly effective in all cases. The sporopollenin peaks become more prominent, especially the 1680 cm^{-1} carbonyl peak (Lutzke *et al.* 2020), which is less obvious in most of the untreated spectra. Similarly, the C=C peak at *c.* 1580 cm^{-1} becomes more prominent, especially in the angiosperm taxa. A carboxylic acid peak at 1710 cm^{-1}, which has been identified in previous studies of isolated sporopollenin (Fraser *et al.* 2014a; Jardine *et al.* 2017; 2021; Lutzke *et al.* 2020), appears as a shoulder on the main 1680 cm^{-1} carbonyl peak.

TABLE 4. Summary of key effects of the processing treatments used in this study, and their potential strengths and weaknesses for (palaeo)palynological applications.

| Treatment	Key effects	Strengths	Weaknesses
Enzymes and solvents	Removes most labile compounds except for carbohydrates in some samples; little or no change to sporopollenin observed	Mostly effective at isolating sporopollenin without altering the chemical structure	Time consuming (*c.* 2 weeks) and challenging to carry out in a standard palynological laboratory; incomplete removal of some labile compounds
Acetolysis	Removes labile compounds; alters sporopollenin by reducing the phenolic content and adding new peaks at *c.* 1160 and 1030 cm^{-1}	Efficient at isolating sporopollenin; quick to use (treatment of ≤10 min) and can be readily applied to small sample quantities; established palynological technique	Alteration of sporopollenin chemistry
Sulfuric acid (H$_2$SO$_4$)	Removes most labile compounds but less effective with proteins; alters sporopollenin by reducing the phenolic content and adding new peaks between 1200 and 1000 cm^{-1}	Reasonably fast treatment (1 h) and can be readily applied to small sample quantities	Alteration of sporopollenin chemistry; incomplete removal of some labile compounds; density of H$_2$SO$_4$ makes subsequent washing and centrifugation challenging
Anhydrous hydrogen fluoride (HF-Py)	Removes most labile compounds but some lipid and protein peaks remain; slight reduction in the phenolic content of the isolated sporopollenin	Can be readily applied to small sample quantities; only limited alteration of sporopollenin chemistry relative to acetolysis and H$_2$SO$_4$	Incomplete removal of some labile compounds; requires use of expensive and harzardous reagent
Potassium hydroxide (KOH)	Incomplete removal of labile compounds, especially carbohydrates and lipids, and new peak at *c.* 1660 cm^{-1} suggests change in protein structure; no obvious changes to sporopollenin observed	Reasonably fast treatment (1 h) and can be readily applied to small sample quantities; does not alter sporopollenin chemistry	Inefficient at isolating sporopollenin, with more evidence of remaining labile compounds compared to the other treatments
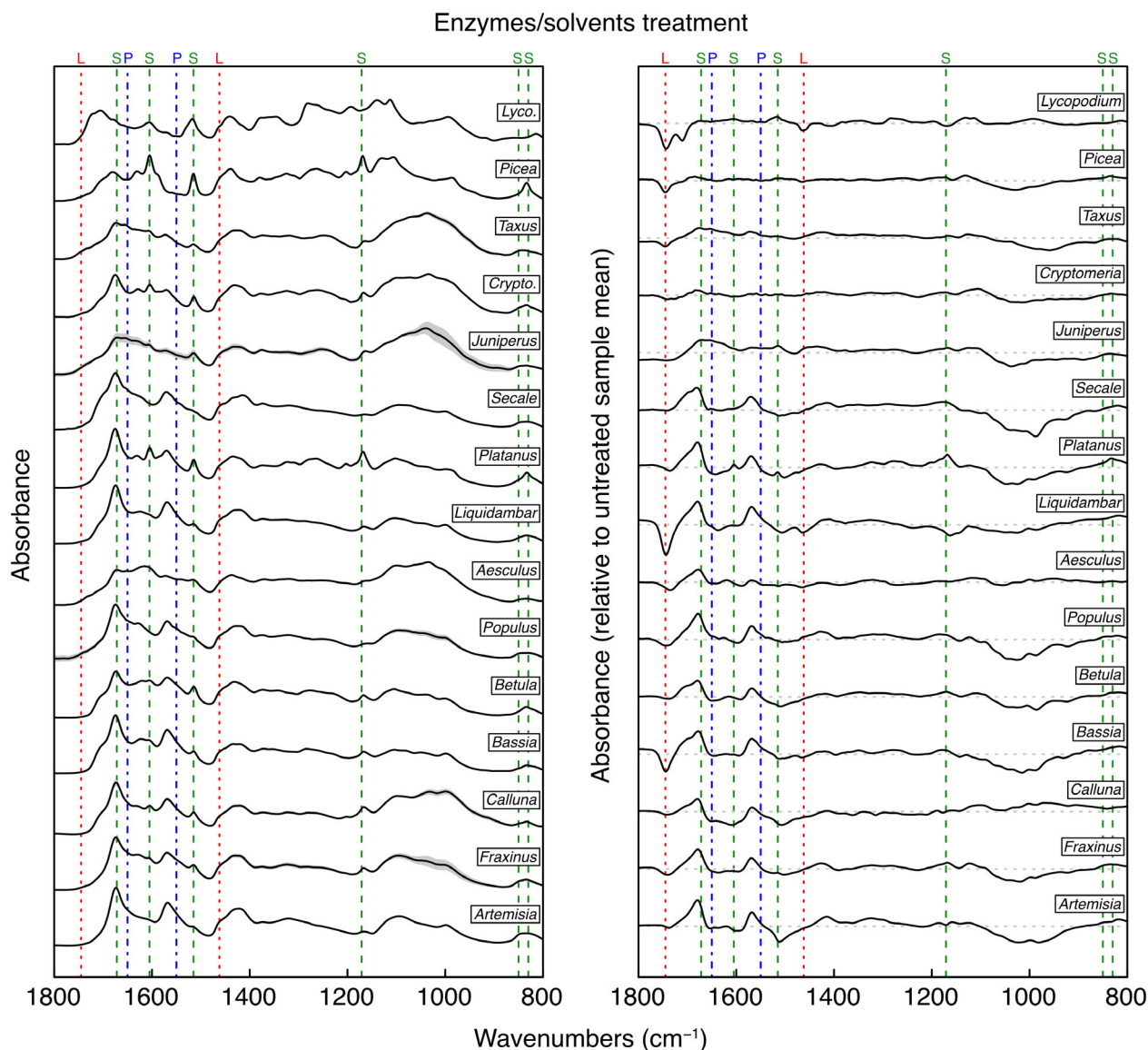
FIG. 3. Stacked FTIR spectra for the enzymes/solvents treatment sample set. Left panel shows the mean spectrum for each taxon, with grey shaded regions showing ±1 standard deviation about the mean. Right panel shows the difference between the mean spectrum and the mean untreated spectrum for each taxon, with the horizontal dashed grey lines showing zero difference in each case (see Fig. 2 for the untreated spectra). Key peaks are marked with dotted/dashed lines. *Abbreviations*: L, lipids; S, sporopollenin; P, proteins (see also Table 3); *Lyco.*, *Lycopodium*; *Crypto.*, *Cryptomeria*.

Acetolysis. While this treatment appears to be effective in removing labile compounds, it also decreases the heights of the phenolic peaks at *c.* 1680, 1605, 1580, 1515, 1170, 850 and 830 cm^{-1}, simplifying the spectra and thus indicating an impact on the general structure of the sporopollenin (Fig. 4). Conversely, the 1710 cm^{-1} carboxyl peak is more prominent in the acetolysed spectra, relative to the enzymes/solvents spectra (Figs 3, 4). Consistent with previous studies (Domínguez *et al.* 1998; Jardine *et al.* 2015; 2017; 2020; 2021; Woutersen *et al.* 2018;

Lutzke *et al.* 2020), prominent peaks appear at *c.* 1160 and 1030 cm^{-1} (Fig. 4).

H$_2$SO$_4$. As with acetolysis, this treatment appears to efficiently remove labile compounds while also reducing the intensity of the aromatic peaks (Fig. 5). Again additional peaks also appear in the 1200 to 1000 cm^{-1} range. The form of these peaks differs to treatment with acetolysis, however, with a pronounced peak at 1200 cm^{-1} and a relatively lower 1030 cm^{-1} peak. The maintenance of the
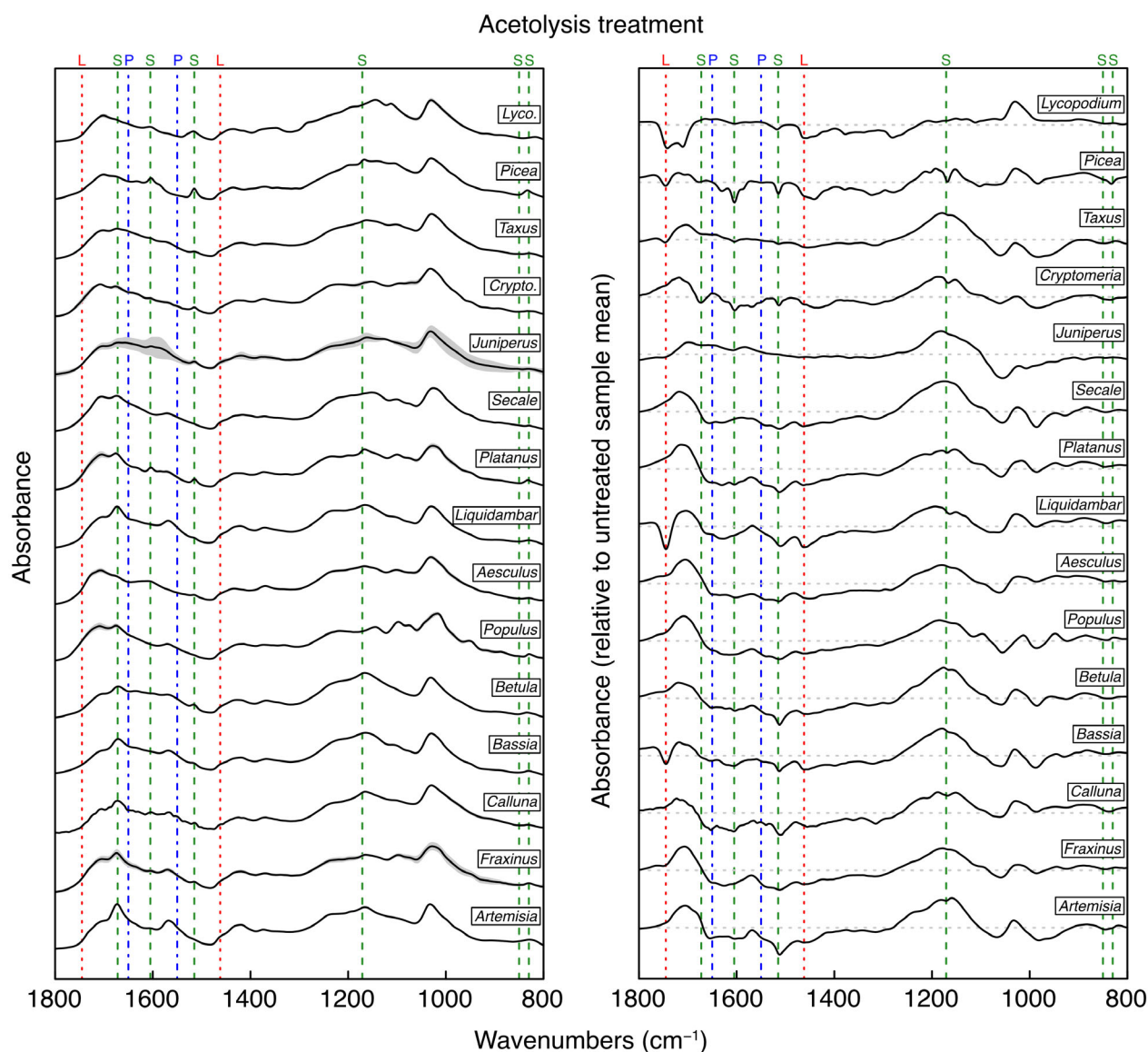
FIG. 4. Stacked FTIR spectra for the acetolysis treatment sample set. Left panel shows the mean spectrum for each taxon, with grey shaded regions showing ±1 standard deviation about the mean. Right panel shows the difference between the mean spectrum and the mean untreated spectrum for each taxon, with the horizontal dashed grey lines showing zero difference in each case (see Fig. 2 for the untreated spectra). Key peaks are marked with dotted/dashed lines. *Abbreviations*: L, lipids; S, sporopollenin; P, proteins (see also Table 3); *Lyco.*, *Lycopodium*; *Crypto.*, *Cryptomeria*.

Amide I peak at 1650 cm^{-1} in some taxa (e.g. *Fraxinus*, *Populus*) suggests that the proteins were also not fully removed by H$_2$SO$_4$ in these samples (Fig. 5).

HF-Py. This treatment removes most labile compounds with apparently less of an obvious impact on the sporopollenin chemistry compared to acetolysis and H$_2$SO$_4$ (Fig. 6). However, the aromatic peaks are slightly decreased in height relative to the intensities after the enzymes/solvents treatment, and in some cases the labile compounds are incompletely removed, such as with the

1745 cm^{-1} lipid peak in *Liquidambar* and *Bassia*, and a peak or shoulder at 1650 cm^{-1} in several taxa suggesting remaining proteins (Fig. 6).

KOH. Following this treatment, a number of labile compound peaks remain across the taxa, specifically peaks in the 1200 to 900 cm^{-1} range suggesting incomplete removal of the carbohydrate intine, and peaks or shoulders at 1745 and 1460 cm^{-1} suggesting incomplete removal of lipids (e.g. *Liquidambar*) (Fig. 7). A new peak at *c.* 1660 cm^{-1} also appears in most taxa following
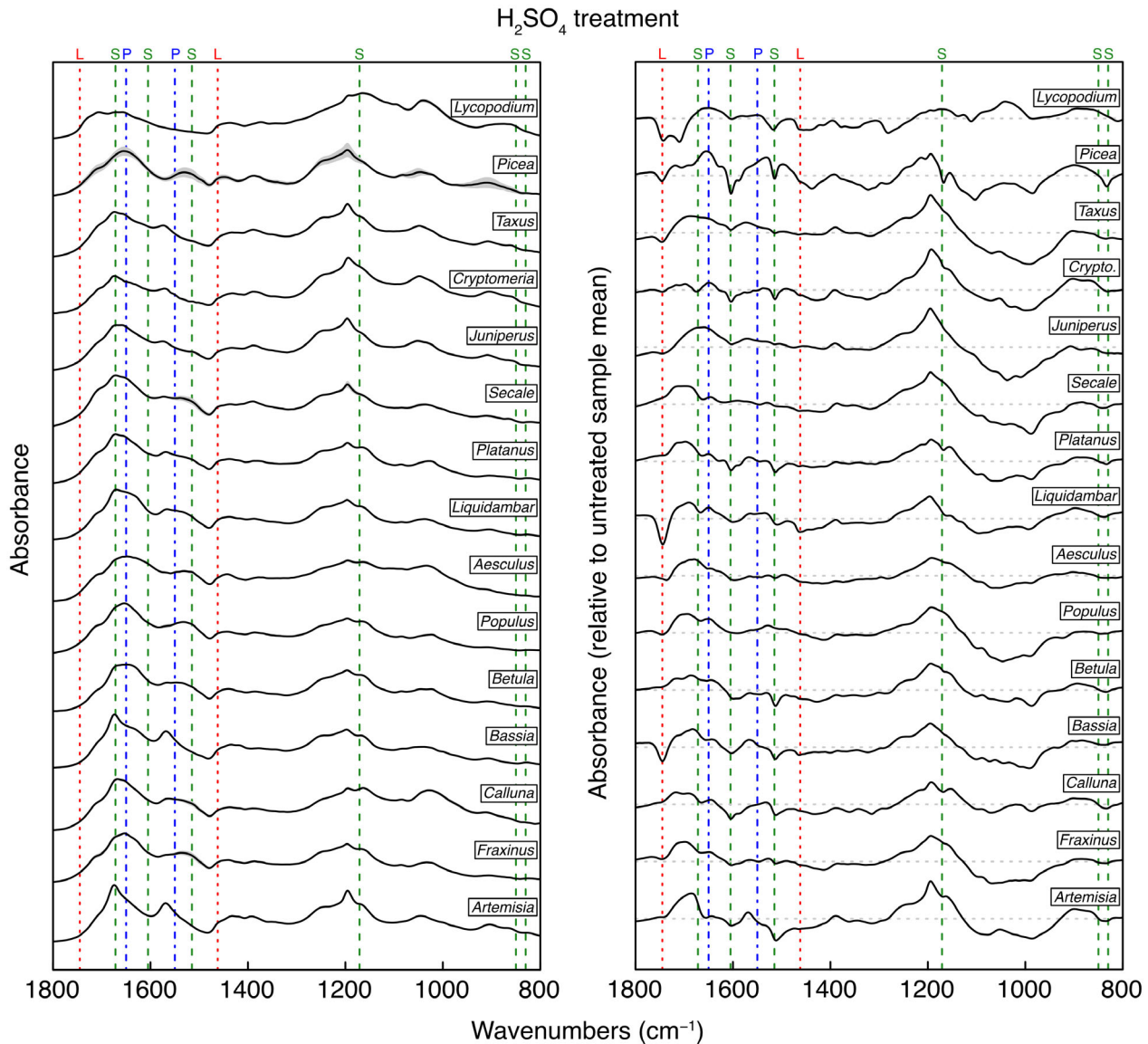
FIG. 5. Stacked FTIR spectra for the H$_2$SO$_4$ treatment sample set. Left panel shows the mean spectrum for each taxon, with grey shaded regions showing ±1 standard deviation about the mean. Right panel shows the difference between the mean spectrum and the mean untreated spectrum for each taxon, with the horizontal dashed grey lines showing zero difference in each case (see Fig. 2 for the untreated spectra). Key peaks are marked with dotted/dashed lines. *Abbreviations*: L, lipids; S, sporopollenin; P, proteins (see also Table 3); *Crypto.*, *Cryptomeria*.

Multivariate analysis

For the untreated samples and non-derivatized spectra, the inter-taxon Euclidean distances are positively correlated with phylogenetic distances, demonstrating that closely-related taxa are chemically more similar compared to more distantly related taxa (Fig. 8). A PCA of the taxon mean spectra shows that the angiosperms group together in the upper left quadrant of the plot (lower end of PC1 and upper end of PC2), separating out from *Lycopodium*, which occurs at the upper end of PC1, and the gymnosperms, which occur between these two extremes (Fig. 9). *Picea* occurs separately to the other three gymnosperm taxa, consistent with their placement on the phylogeny (Fig. 1). Similarly, the superasterid taxa all cluster together, although the superrosids are more spread out in ordination space (Fig. 9). The PC loadings show that the heights of the 1650
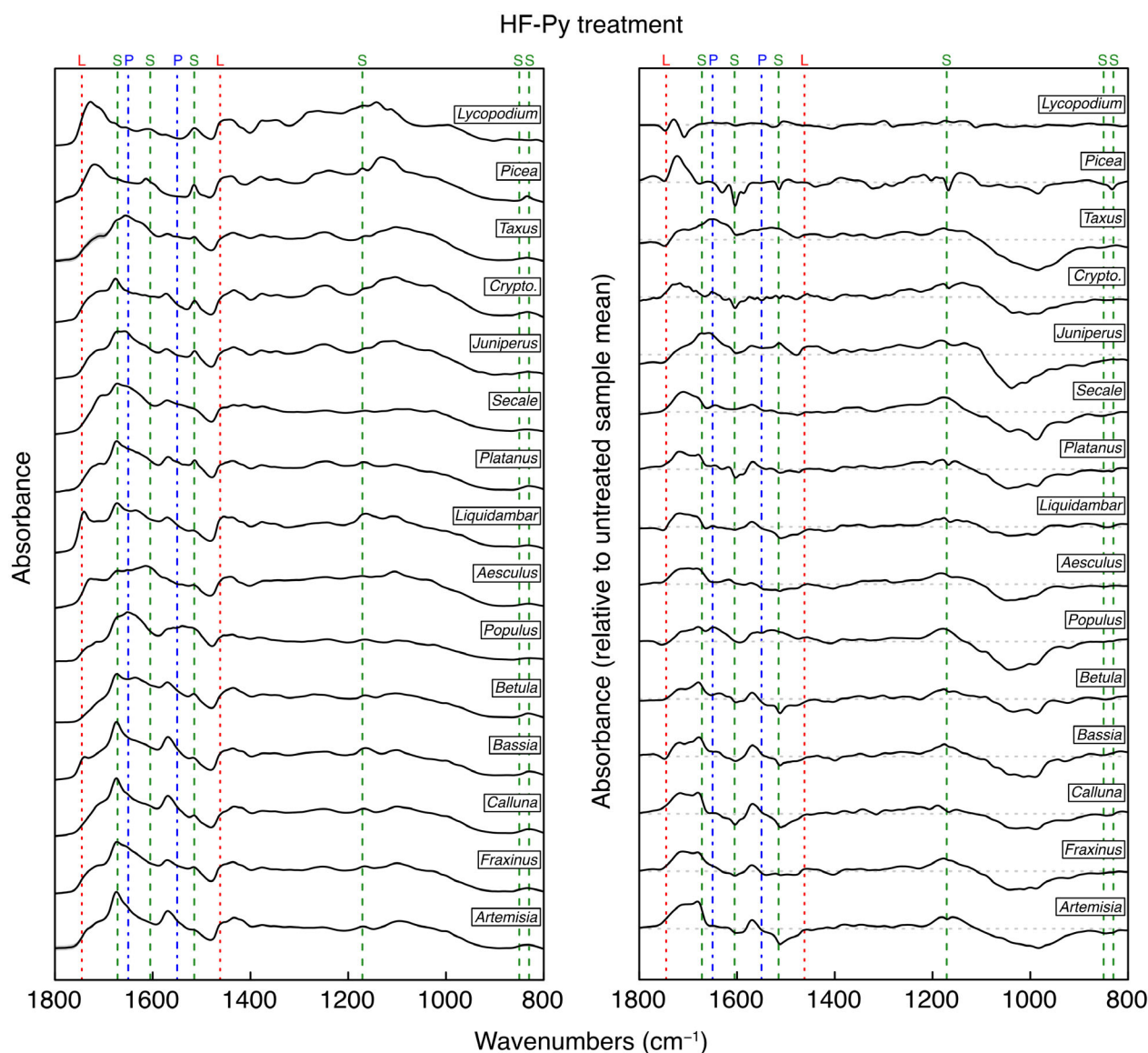
treatment with KOH, which possibly represents a change in protein structure leading to a shift of the Amide I band to higher wavenumbers.

FIG. 6. Stacked FTIR spectra for the HF-Py treatment sample set. Left panel shows the mean spectrum for each taxon, with grey shaded regions showing ±1 standard deviation about the mean. Right panel shows the difference between the mean spectrum and the mean untreated spectrum for each taxon, with the horizontal dashed grey lines showing zero difference in each case (see Fig. 2 for the untreated spectra). Key peaks are marked with dotted/dashed lines. *Abbreviations*: L, lipids; S, sporopollenin; P, proteins (see also Table 3); *Crypto.*, *Cryptomeria*.

and 1550 cm^{-1} protein peaks, and the 1020 cm^{-1} carbohydrate peak, are negatively correlated with PC1 (that is, the peak heights are relatively lower in taxa at the upper end of PC1, and relatively higher in taxa at the lower end of PC1), while the 1710 and 1170 cm^{-1} sporopollenin peaks are positively correlated with PC1 (Fig. 10; Table 3). For PC2, peaks at 1740, 1515 and 1170 cm^{-1} are positively correlated, while the 1020 cm^{-1} peak is negatively correlated.

The inter-taxon distances for the enzymes/solvents, H$_2$SO$_4$ and HF-Py treatments are all positively correlated

with each other, the phylogenetic distances, and to the inter-taxon distances for the untreated samples (Spearman's *rho* = 0.4–0.8) (Fig. 8). Ordinations of these treatments reveals a broadly similar arrangement of taxa as the untreated samples, with the angiosperms separating out from the gymnosperms and *Lycopodium* (Fig. 9). While the loadings vary between the treatments, for PC1 peaks between 1200 and 1000 cm^{-1} are positively correlated while peaks between 1700 and 1500 cm^{-1} are negatively correlated (Fig. 10).
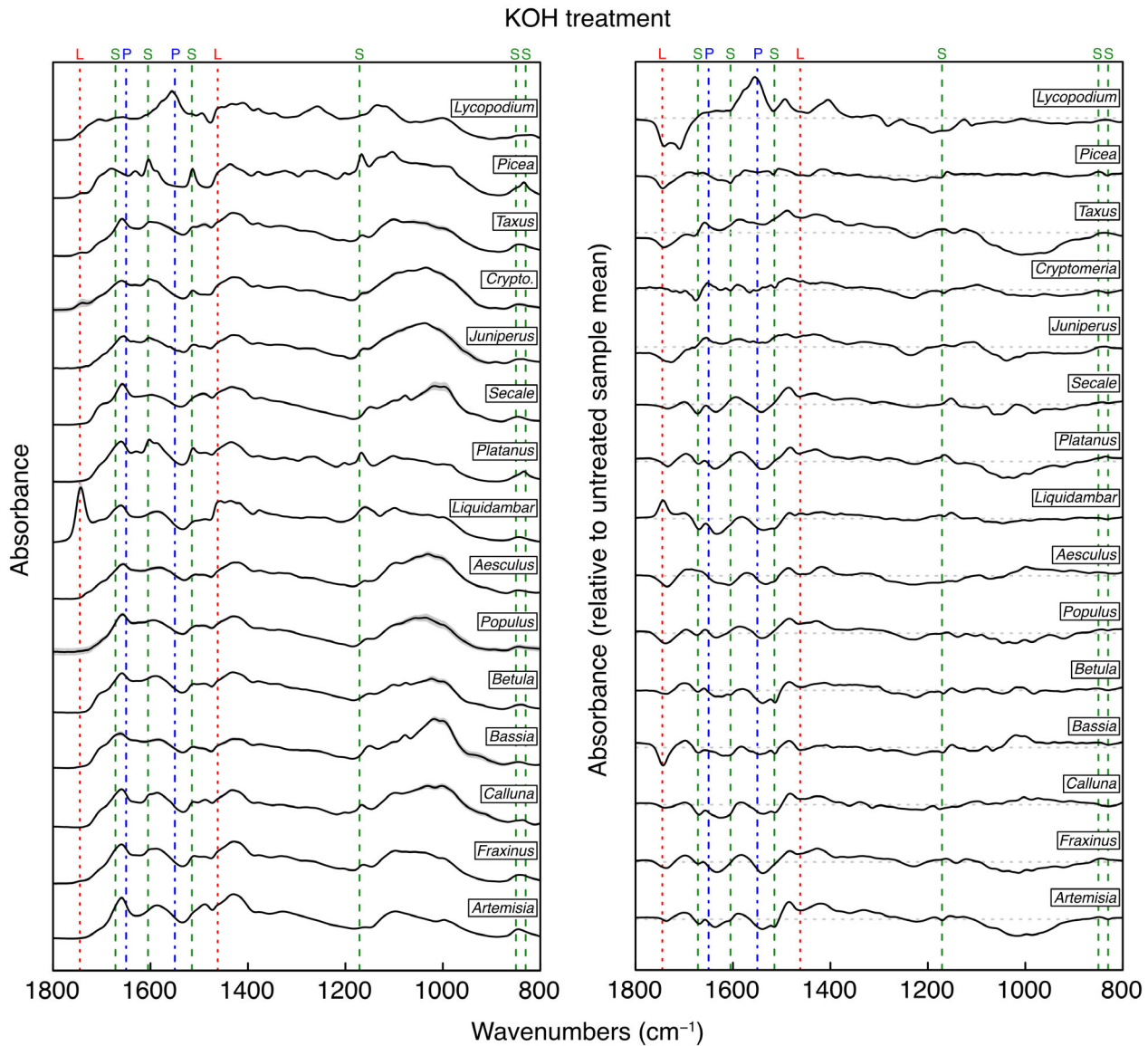
FIG. 7. Stacked FTIR spectra for the KOH treatment sample set. Left panel shows the mean spectrum for each taxon, with grey shaded regions showing ±1 standard deviation about the mean. Right panel shows the difference between the mean spectrum and the mean untreated spectrum for each taxon, with the horizontal dashed grey lines showing zero difference in each case (see Fig. 2 for the untreated spectra). Key peaks are marked with dotted/dashed lines. *Abbreviations*: L, lipids; S, sporopollenin; P, proteins (see also Table 3); *Crypto.*, *Cryptomeria*.

While the inter-taxon distances for the KOH and acetolysis treatments are positively correlated with each other and the other treatments and the phylogenetic distances, the correlation coefficients are lower (Fig. 8). This is especially pronounced for acetolysis which is at most only weakly correlated with the other treatments, and uncorrelated with phylogeny. A PCA of the KOH treated samples reveals a less clear phylogenetic signal in the arrangement of the taxa in ordination space compared to most of the other treatments, although the angiosperms are mostly limited to the lower end of PC2 and the gymnosperms are mostly limited to the upper end (Fig. 9). An exception to this is *Liquidambar*, which occurs at the upper end of PC2; the PC loadings show that this is because of the higher 1745 cm^{-1} lipid peak that remains in these samples, and the relatively pronounced 1170 cm^{-1} peak that it shares with *Picea* (Figs 7, 10). The arrangement of taxa on PC1 is largely controlled by the relative height of the carbohydrate peak, suggesting that incompleteness of the removal of the intine across the taxa overprints much of the underlying phylogenetic signal in these spectra (Fig. 10).
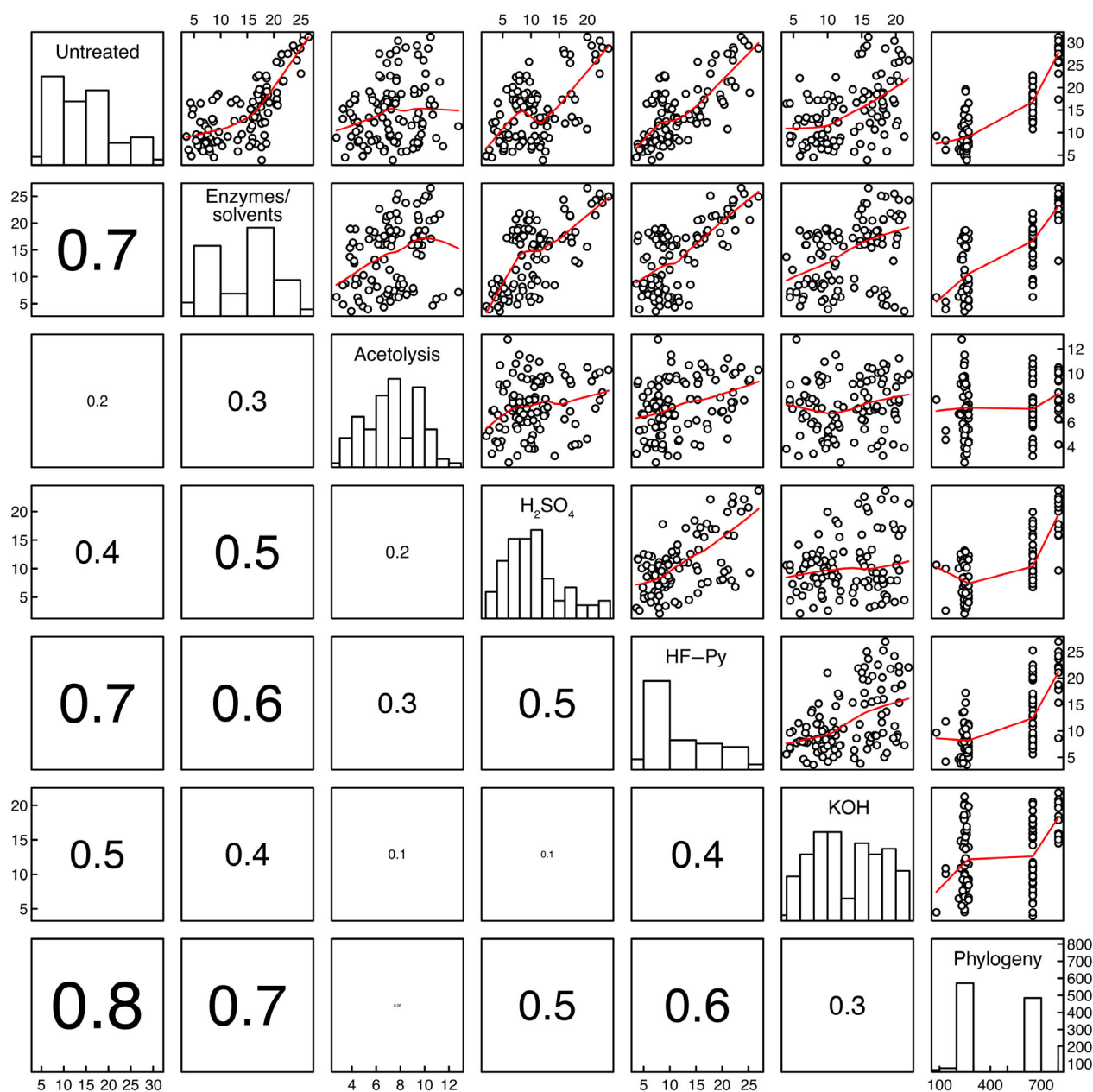
FIG. 8. Pairs plot for the non-derivatized taxon mean spectra and phylogenetic pairwise distances, showing the relationships among treatments and phylogeny via scatterplots (upper triangle) and Spearman's *rho* correlation coefficients (lower triangle), and the distribution of pairwise distances as histograms on the diagonal.

A PCA of the acetolysed samples shows limited evidence of a phylogenetic signal in the arrangement of the taxa, with considerable overlap between the gymnosperms and angiosperms, and no clear clade-based groupings within the angiosperms (Fig. 9). Inspection of the PC loadings shows that peaks at *c.* 1720 and <1100 cm^{-1} are positively correlated with the PC 1 scores, while sporopollenin peaks at *c.* 1670, 1570 and 1170 cm^{-1} are negatively correlated (Fig. 10).

The inter-taxon Euclidean distances are highest for the untreated and enzymes/solvents samples and lowest for the acetolysed samples, with the H$_2$SO$_4$, HF-Py and KOH samples intermediate between these (Fig. 11). The lower distances among the acetolysed samples are consistent with the observed simplification of the IR spectra (Fig. 4).

Repeating these analyses with second derivative spectra reveals broadly similar relationships among the inter-taxon distances to the non-derivatized spectra. The
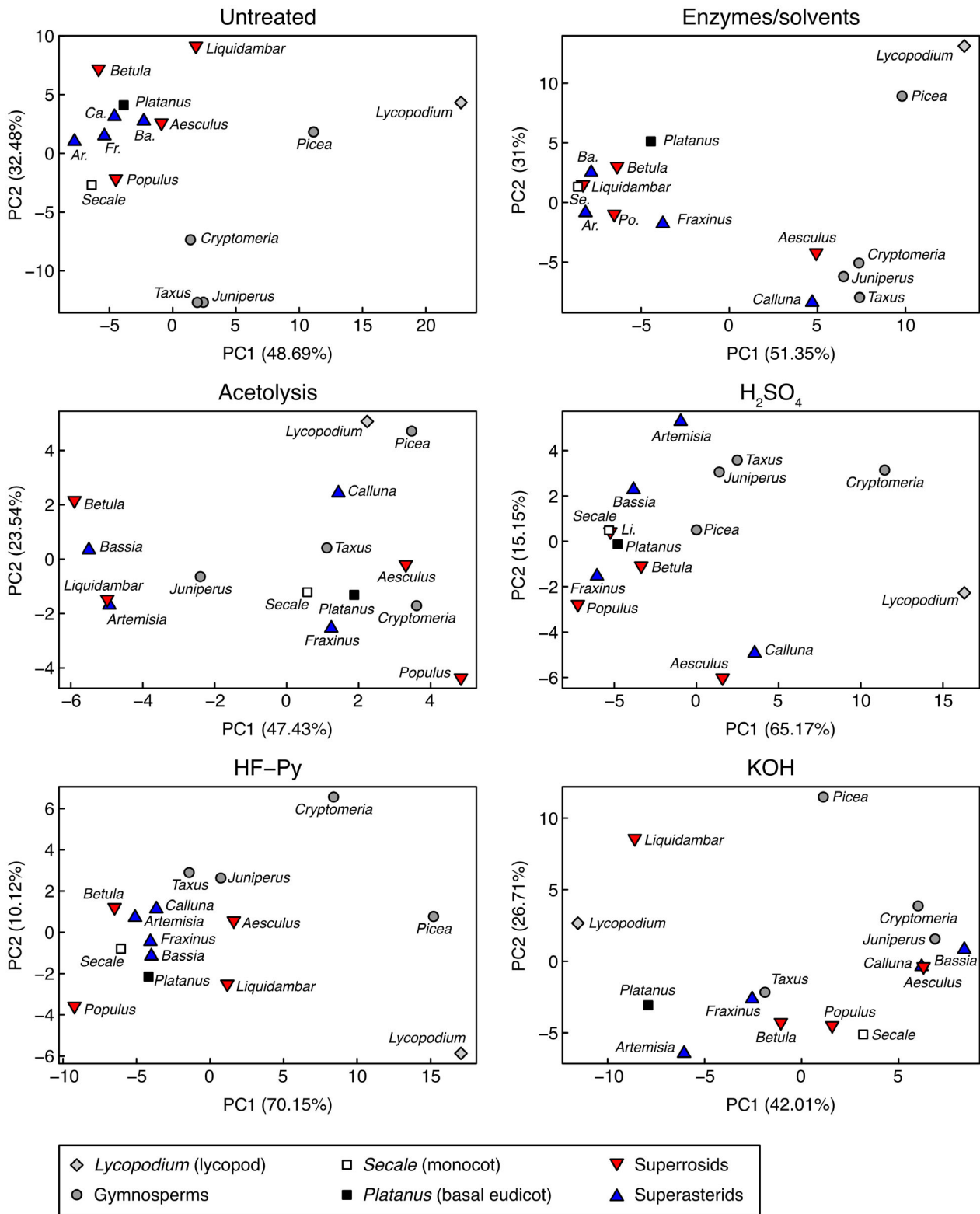
FIG. 9. Principal components analysis scatterplots for the non-derivatized taxon mean spectra, showing PC1 and PC2 for each treatment. Percentage variance explained by each principal component shown in parentheses.
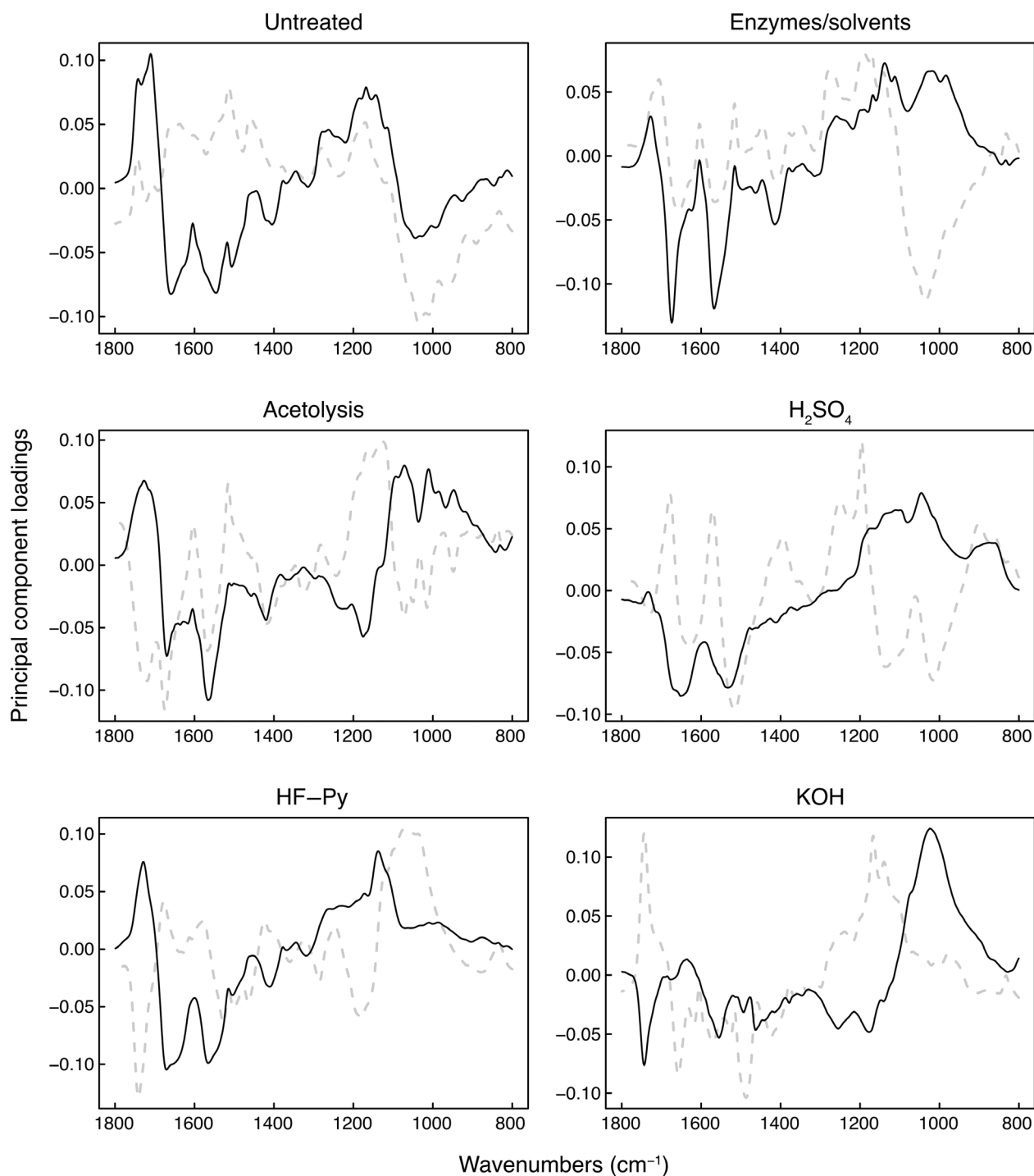
FIG. 10. Principal components analysis loadings plots for the non-derivatized spectra, showing the PC1 (black solid lines) and PC2 (grey dashed lines) loadings for each treatment.

correlation coefficients between the treatments are in general higher with the derivatized spectra, especially for the acetolysed and KOH treated samples (Fig. 12). They are lower for the H_2SO_4 treatment, however, and the correlations between the chemical and phylogenetic distances are also generally weaker with the derivatized spectra (Fig. 12). Comparing the inter-taxon Euclidean distances for each treatment shows a similar pattern to the non-derivatized spectra, with the exception that the H_2SO_4 spectra have similarly low distances to the acetolysed spectra (Fig. 13).
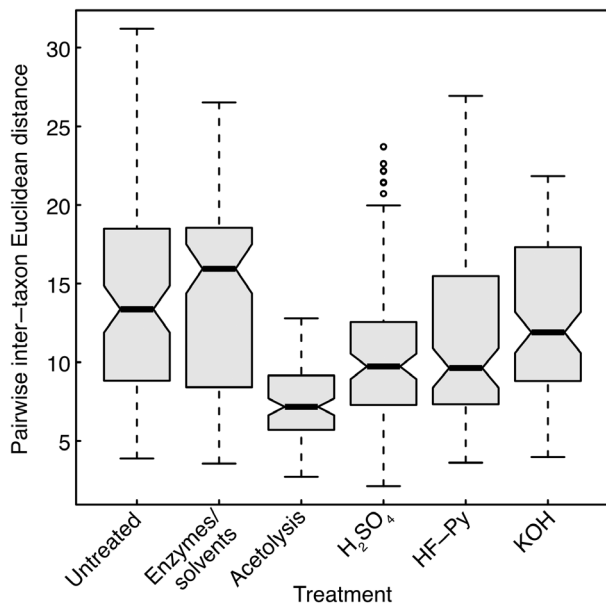
FIG. 11. Box plot of the pairwise inter-taxon Euclidean distances for each treatment, for the non-derivatized spectra. For each box the thick horizontal line shows the median, the edges of the box show the lower and upper quartile, and the whiskers show the extent of the data up to 1.5 times the interquartile range; values beyond this are shown as individual points. Notches around the median lines roughly correspond to a 95% confidence interval for the medians, and were calculated according to McGill et al. (1978).

Applying PCA to the derivatized spectra shows a separation of the angiosperm and non-angiosperm taxa in ordination space for most treatments, and to some extent a separation between the superrosid and superasterid angiosperms (Fig. 14). This pattern is less clear in the untreated spectra, where *Lycopodium* and *Picea* separate out from the rest of the taxa on PC1, and for the KOH treated spectra where *Liquidambar* separates out on PC1 and the other taxa cluster together (Fig. 14). Compared to the non-derivatized spectra the acetolysed samples show a clearer within-clade clustering. The percentage of variance explained on the first two PCs is slightly lower with the derivatized spectra compared to the non-derivatized spectra, but in most cases the first two axes account for *c.* 70% of the variance in the dataset (Fig. 14).

DISCUSSION

Overall, these results suggest that using different treatment methods can lead to substantially different residual sporomorph chemistries, either through not completely removing labile compounds or by altering the chemistry of the sporopollenin. This shows that for machine learning-based classifications using sporomorph chemistry it will be essential to work with training sets that have been processed in exactly the same way as the material being classified to ensure that like is being compared with like. However, the overall ranking of inter-taxon distances is similar across most of the treatments, with high positive correlations among distances and a similar arrangement of taxa in ordination space (Figs 8, 9). This suggests that while the different treatments lead to a different sporomorph chemistry, the same underlying taxon-to-taxon relationships can be uncovered. There are exceptions to this, such as acetolysis and to some extent the treatment using KOH, although processing the spectra with derivatives and Savitzky–Golay smoothing increases the correlation with the enzymes/solvents treatment, suggesting that a common underlying chemical pattern can, to some extent, be recovered with these methods as well (Figs 12, 14).

These results support a phylogenetic signal in sporomorph chemistry, with a separation between the angiosperms and non-angiosperms in ordination space, and some evidence for clade-based groupings within these higher taxa. While the correlation between the inter-taxon chemical and phylogenetic distances in the processed samples shows that sporopollenin chemistry contains a phylogenetic signal (Woutersen *et al.* 2018; Nierop *et al.* 2019; Jardine *et al.* 2021), the decrease of this correlation with processing suggests that the labile compounds carry a phylogenetic signal as well, and it is the combination of these different components that drives the relationship with phylogeny in fresh sporomorph chemistry (Zimmermann & Kohler 2014; Bağcıoğlu *et al.* 2015). Interestingly, the correlations between the inter-taxon distances and phylogeny are generally lower with the derivatized spectra compared to the non-derivatized spectra (Figs 8, 12), and future analyses of sporopollenin evolutionary history should carry out data analysis both with and without spectral processing to take the impact of this step into account.

The treatments used here each have pros and cons for use in palynological analyses, linked both to their ability to isolate sporopollenin efficiently while not altering its chemical structure, but also to their general applicability for (palaeo)palynological research (Table 4). In particular, collecting samples for classification training sets will mean going beyond sporomorph types that can be purchased in bulk or collected from abundant sporomorph producers (such as *Lycopodium*, conifers or birches), and instead working with small samples from plants that produce limited quantities of sporomorphs, sampled in natural settings, botanical gardens or herbaria. Ideally this will also involve sampling a number of individual plants per species to incorporate intra-specific variation and chemical responses to environmental gradients (Zimmermann
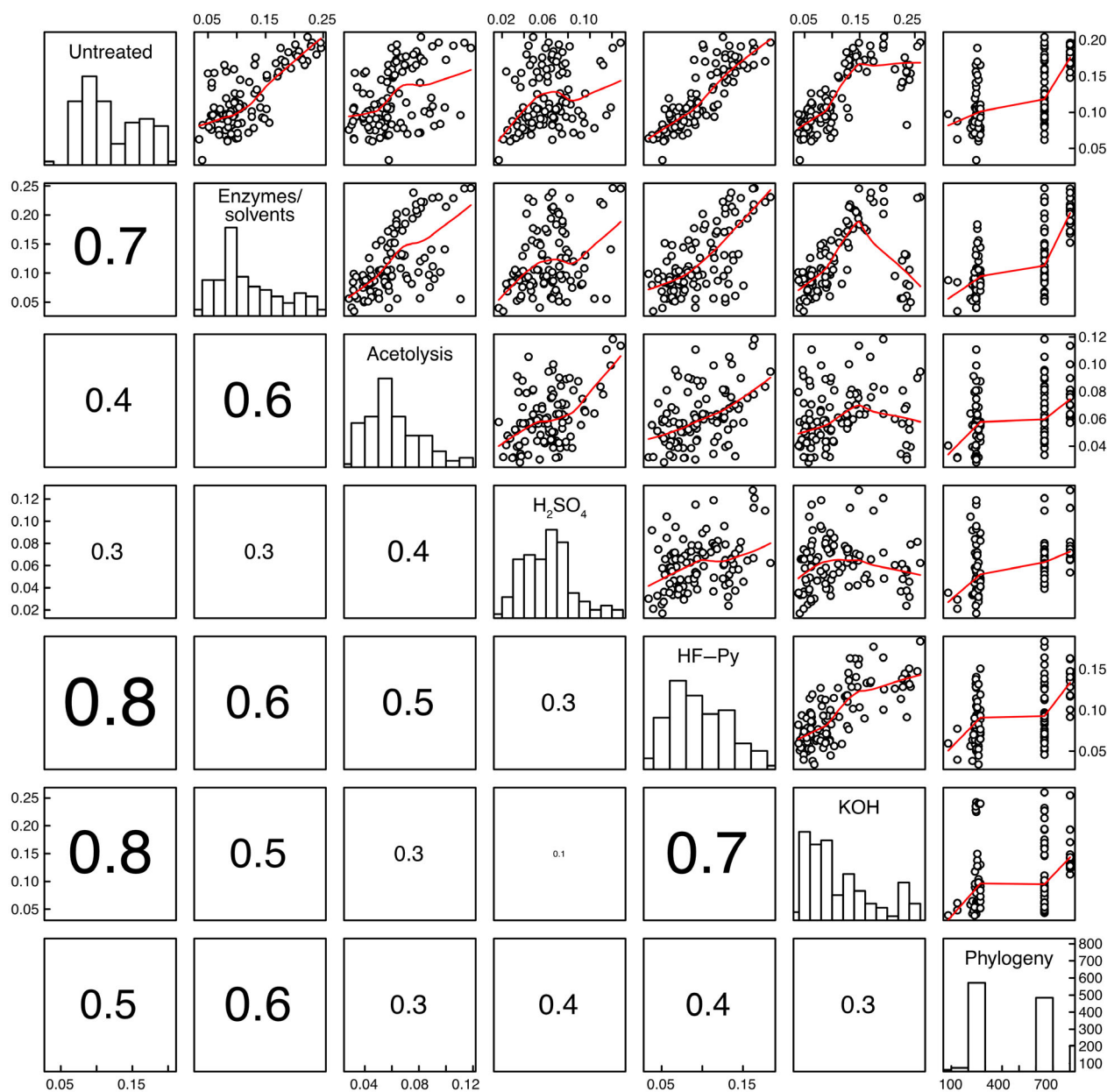
FIG. 12. Pairs plot for the 2nd derivative taxon mean spectra and phylogenetic pairwise distances, showing the relationships among treatments and phylogeny via scatterplots (upper triangle) and Spearman's *rho* correlation coefficients (lower triangle), and the distribution of pairwise distances as histograms on the diagonal.

& Kohler 2014; Muthreich *et al.* 2020), again limiting potential sample sizes because multiple individuals cannot be pooled together. Such research is also likely to involve efficiently processing large numbers of samples in standard palynology laboratory set-ups, rather than in dedicated chemistry laboratories.

The enzymes/solvents treatment appears to remove most labile compounds while leaving the sporopollenin biopolymer largely unaffected, although the carbohydrate intine was incompletely removed in some taxa (Fig. 3). One key difference in our study compared to previous

applications of this approach (e.g. Lutzke *et al.* 2020) is that, due to equipment constraints tied to processing 15 samples at once in a palynological laboratory, the samples were not continuously stirred during the enzyme digestion step but were instead manually stirred at intervals across the 72-hour treatment. Since this included leaving the samples unattended overnight, it is possible that this led to the intine not being fully digested, and this is something that would need careful consideration if this approach were to be used again. However, the wider applicability of this technique also needs to be taken into
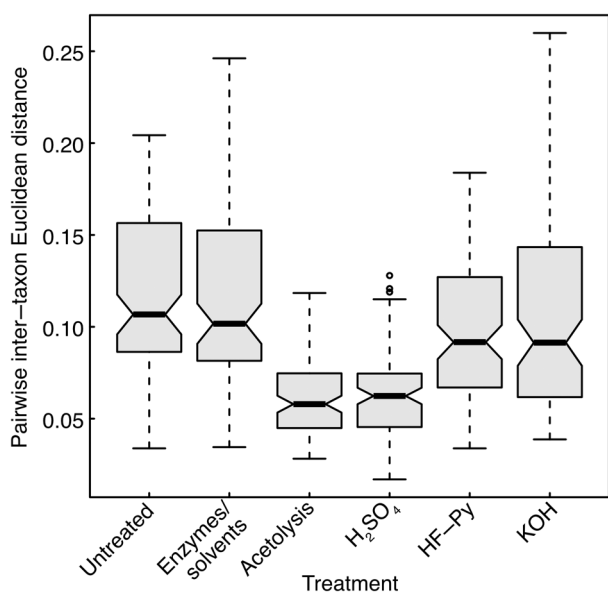
FIG. 13. Box plot of the pairwise inter-taxon Euclidean distances for each treatment, for the 2nd derivative spectra. For each box the thick horizontal line shows the median, the edges of the box show the lower and upper quartile, and the whiskers show the extent of the data up to 1.5 times the interquartile range; values beyond this are shown as individual points. Notches around the median lines roughly correspond to a 95% confidence interval for the medians, and were calculated according to McGill et al. (1978).

account, since the processing protocol takes c. 2 weeks to work through, including treatment with multiple reagents and many cycles of centrifugation and decanting. This approach is therefore unlikely to be useful for high-throughput processing and working with very small sample sizes, limiting its use for developing chemotaxonomic libraries (Julier et al. 2016; Zimmermann et al. 2016; Jardine et al. 2019), analysing sporomorph chemistry evolution (Fraser et al. 2012; Nierop et al. 2019) or calibrating palaeoclimate proxies (Blokker et al. 2005; Lomax et al. 2008; Jardine et al. 2016; 2017).

Processing sporomorphs with KOH, while a standard approach for morphological studies (Traverse 2007), appears to be less suitable for chemical analyses, due to the incomplete removal of both carbohydrates and lipids (Fig. 7). However, it is perhaps unsurprising that not all labile compounds are removed by a 1-hour treatment with KOH, given that the standard approach used in sporopollenin isolation for microencapsulation applications involves refluxing with acetone for c. 12 hours to remove lipids, KOH for c. 12 hours to remove proteins, and phosphoric acid for 7 days to remove residual proteins and the carbohydrate intine, as well as multiple washes with solvents (Gonzalez-Cruz et al. 2018). While this approach and variations of it result in empty and clean sporopollenin shells suitable for microencapsulation (Mundargi et al. 2016;

Gonzalez-Cruz et al. 2018; Uddin et al. 2019), as with the enzymes/solvents approach it is unlikely to be useful for routine palynological applications involving large quantities of, what are often, small samples.

Both the H_2SO_4 and HF-Py treatments were positively correlated with the enzymes/solvents and untreated samples, and would be suitable for working with smaller sample sizes. However, both approaches failed to remove all the labile compounds, suggesting longer treatment times would be needed, and the H_2SO_4 treatment appears to alter sporopollenin chemistry, with a reduction in the height of the phenolic peaks (Fig. 5). HF-Py is also expensive to purchase and hazardous to use, and since palynologists are actively looking for alternatives to HF for routine sample processing (Riding & Kyffin-Hughes 2011; Santos & Ledru 2022) this treatment seems less likely to attain widespread use.

Acetolysis leaves a clear imprint on the processed material, with a reduction in the height of phenolic peaks and the introduction of new absorbance bands at c. 1160 and 1030 cm^{-1} (Fig. 4). This means that the resultant biopolymer is not unaltered sporopollenin (Lutzke et al. 2020) and inter-taxon chemical differences are reduced compared to most of the other treatments (Figs 11, 13). Processing the spectra with derivatives increases the correlation with all the other treatments, at least in terms of relative inter-taxon Euclidean distances and the arrangement of taxa in ordination space, although the correlation coefficients are still relatively low in most comparisons (Fig. 12), and determining whether this approach can be applied to future studies of the evolutionary history of sporopollenin chemistry requires further research. Acetolysis is, however, convenient and quick to use and gives consistent results, with all labile compounds removed, and is applicable to small samples sizes. Chemical classification with acetolysed material does appear to be feasible as well (Jardine et al. 2021), even if unaltered sporopollenin chemistry is not being used as the basis for classification, although again this requires further validation to ensure that the results are consistent and reliable across both extant and fossil specimens, and are not systematically biased by the processing procedure used. Similarly, UV-B irradiance reconstructions have been achieved using acetolysed sporomorphs (Jardine et al. 2016; 2020), and a direct comparison between UV-B absorbing phenolic compound levels in acetolysed and unacetolysed subfossil grass pollen grains revealed a strong positive correlation (Jardine et al. 2016), despite the reduction in phenolic content that acetolysis causes. Pragmatically, acetolysis may still be a suitable approach for future chemopalynological investigations, once the various factors (convenience of use and applicability to small sample sizes vs impacts on sporopollenin chemistry) have been considered, but only if it can be shown that
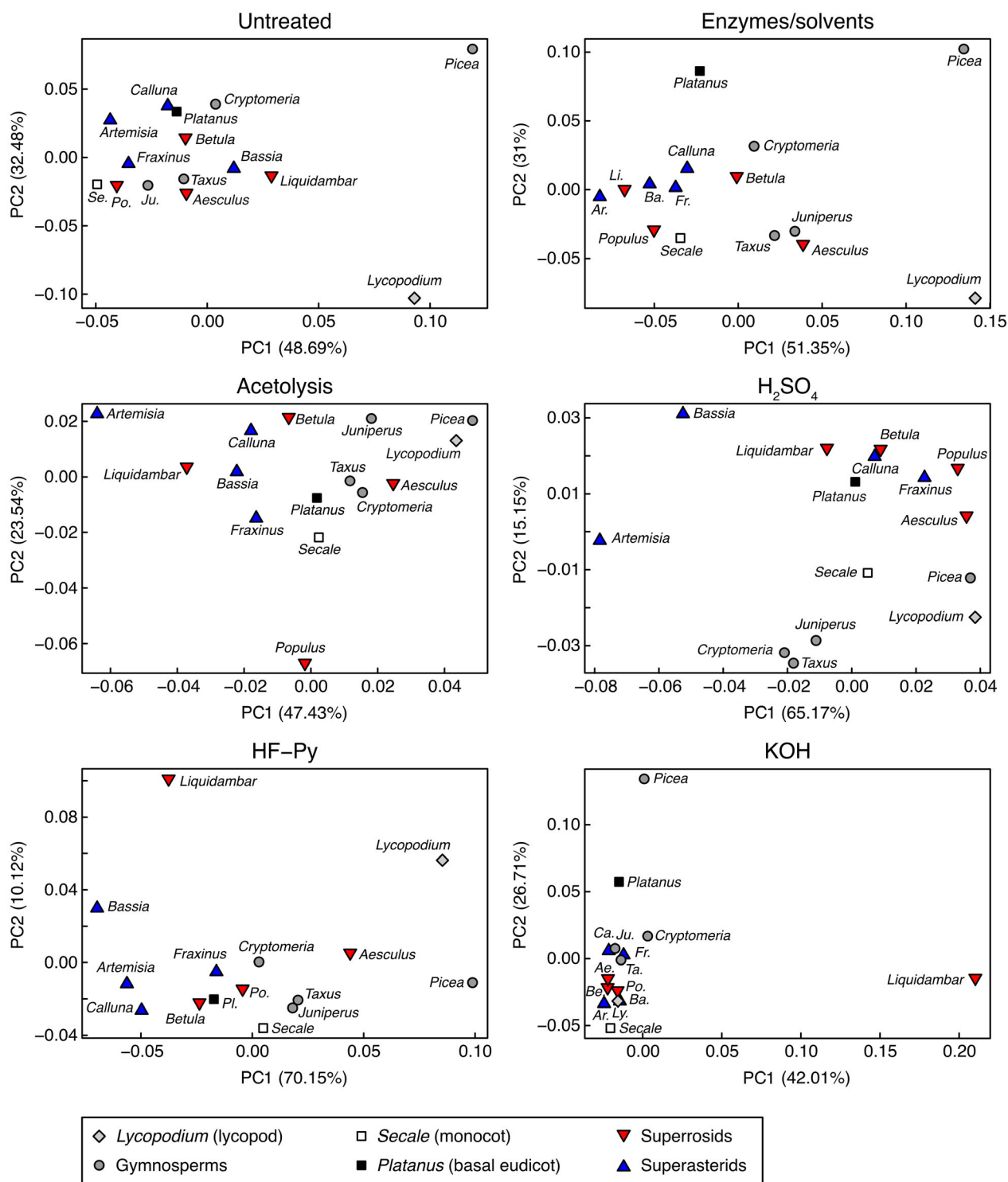
FIG. 14. Principal components analysis scatterplots for the 2nd derivative taxon mean spectra, showing PC1 and PC2 in each case. Percentage variance explained by each principal component shown in parentheses.

the chemical alterations inherent in this approach do not invalidate the assumptions of the research.

The broader findings from this study (that no one approach both efficiently isolated the sporopollenin exine and left it unaltered) suggests that palynologists will need to continue investigating new methods, or adapt those used here, to try to find a better solution to the problem of processing extant sporomorphs for chemical analysis. We have emphasized keeping methods broadly applicable, including limiting techniques as far as possible to those that can operate in standard palynological laboratories, and ideally without the use of hazardous and expensive reagents such as HF, but being open to collaborating with chemists and incorporating novel techniques and equipment is likely to help solve some of the issues raised in this study.

While here we have used FTIR spectroscopy as a measurement tool, a range of vibrational and mass spectroscopic methods are available that can provide complementary information on different aspects of sporomorph chemistry, and future analyses of the kind carried out in this study would benefit from incorporating these. Like FTIR, Raman spectroscopy is a vibrational technique that provides information on the molecular structure of analysed samples, but with differing sensitivity to different functional groups, so that bands that are weakly detected in one approach are often strong in the other (Mayo 2003; Olcott Marshall & Marshall 2014). Although Raman spectroscopy has previously been used to analyse sporomorphs (e.g. Schulte *et al.* 2008; Bağcıoğlu *et al.* 2015; Diehn *et al.* 2020), strong fluorescence can swamp the Raman signal and therefore complicate measurements, most significantly with the isolated sporopollenin and associated bio/geopolymers analysed in palaeopalynological research (Olcott Marshall & Marshall 2014; Bernard *et al.* 2015). However, Fourier transform (FT) Raman appears to overcome this issue (Zimmermann 2010; Kenđel & Zimmermann 2020), and so may be a viable option for comparing sporopollenin isolation approaches and analysing fossil sporomorphs. While mass spectroscopic analysis of sporomorphs has traditionally been problematic because of the chemical inertness of the sporopollenin biopolymer, thermally assisted hydrolysis and methylation–pyrolysis–gas chromatography–mass spectrometry (THM–py–GC–MS) has successfully been used to provide detailed information on the identity and abundance of the specific compounds comprising sporopollenin (Blokker *et al.* 2005; Watson *et al.* 2007; 2012; Willis *et al.* 2011; Bell *et al.* 2018). Incorporating THM–py–GC–MS measurements would therefore be a natural extension of the FTIR-based analysis presented here, especially in terms of understanding the potential of the different isolation techniques to drive changes in sporopollenin chemistry. Future work in this area will also benefit from incorporating (sub)fossil specimens that have been treated with a range of chemical treatments, to better understand how extant and fossil datasets can be directly integrated in the same analysis.

CONCLUSION

In this study, we have compared sporopollenin isolation approaches using a range of taxa from across the tracheophyte phylogeny. The results of this research show that there is a phylogenetic signal both in fresh sporomorphs and in isolated sporopollenin. While different treatments lead to different sporomorph chemistries, either because some labile compounds remain following processing or because the sporopollenin itself is altered, the various treatments broadly converge on a common signal in terms of relative chemical similarities and differences among taxa. Although there are exceptions to this finding, such as processing with acetolysis and to some extent KOH, analysing derivatized spectra increases the overall similarity with other treatments. Future studies of chemical palynology will need to consider the trade-offs between the different processing options available, including the efficacy of the treatment, its impact on sporopollenin chemistry, and its applicability to small sample sizes and high-throughput processing of potentially large batches of samples.

Author contributions. **Conceptualization** PE Jardine (PEJ), MS Kent (MSK), WT Fraser (WTF), K-H Knorr (KHK), BH Lomax (BHL); **Data Curation** PEJ; **Formal Analysis** PEJ; **Funding Acquisition** PEJ, WTF, BHL; **Investigation** PEJ; **Methodology** PEJ, KHK; **Resources** PEJ, KHK; **Visualization** PEJ; **Writing – Original Draft Preparation** PEJ; **Writing – Review & Editing** MSK, WTF, KHK, BHL.

DATA ARCHIVING STATEMENT

Research Data and R code associated with this article is available in FigShare: https://doi.org/10.6084/m9.figshare.22182790

Editor. Benjamin Bomfleur

REFERENCES

AHLERS, F., THOM, A., LAMBERT, J., KUCKUK, R. and WIERMANN, R. 1999. ^{1}H NMR analysis of sporopollenin from *Typha angustifolia*. *Phytochemistry*, **50**, 1095–1098.

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.

BELL, B. A., FLETCHER, W. J., RYAN, P., GRANT, H. and ILMEN, R. 2017. Stable carbon isotope analysis of *Cedrus atlantica* pollen as an indicator of moisture availability. *Review of Palaeobotany & Palynology*, **244**, 128–139.

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.

BELL, B. A., FLETCHER, W. J., CORNELISSEN, H. L., CAMPBELL, J. F. E., RYAN, P., GRANT, H. and ZIEL-HOFER, C. 2019. Stable carbon isotope analysis on fossil *Cedrus* pollen shows summer aridification in Morocco during the last 5000 years. *Journal of Quaternary Science*, **34**, 323–332.

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.

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

COJOCARU, R., MANNIX, O., CAPRON, M., MILLER, C. G., JOUNEAU, P.-H., GALLET, B., FALCONET, D., PACUREANU, A. and STUKINS, S. 2022. A biological nanofoam: the wall of coniferous bisaccate pollen. *Science Advances*, **8**, eabd0892.

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

DIEGO-TABOADA, A., MAILLET, L., BANOUB, J. H., LORCH, M., RIGBY, A. S., BOA, A. N., ATKIN, S. L. and MACKENZIE, G. 2013. Protein free microcapsules obtained from plant spores as a model for drug delivery: ibuprofen encapsulation, release and taste masking. *Journal of Materials Chemistry B*, **1**, 707–713.

DIEHN, S., ZIMMERMANN, B., TAFINTSEVA, V., SEIFERT, S., BAĞCIOĞLU, M., OHLSON, M., WEIDNER, S., FJELLHEIM, S., KOHLER, A. and KNEIPP, J. 2020. Combining chemical information from grass pollen in multimodal characterization. *Frontiers in Plant Science*, **10**, 1788.

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.

DOMÍNGUEZ, E., MERCADO, J. A., QUESADA, M. A. and HEREDIA, A. 1999. Pollen sporopollenin: degradation and structural elucidation. *Sexual Plant Reproduction*, **12**, 171–178.

FAITH, D. P. 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation*, **61**, 1–10.

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.

FRASER, W. T., 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.

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

FRASER, W. T., LOMAX, B. H., JARDINE, P. E., GOSLING, W. D. and SEPHTON, M. A. 2014b. Pollen and spores as a passive monitor of ultraviolet radiation. *Frontiers in Ecology & Evolution*, **2**, 12.

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.

GRIENER, K. W., NELSON, D. M. and WARNY, S. 2013. Declining moisture availability on the Antarctic Peninsula during the Late Eocene. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **383–384**, 72–78.

JANSONIUS, J. and MacGREGOR, D. C. (eds). 1996. *Palynology: Principles and applications*. American Association of Stratigraphic Palynologists Foundation, 1330 pp.

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.

JARDINE, P. E., FRASER, W. T., LOMAX, B. H., 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.

JARDINE, P. E., 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.

JARDINE, P. E., 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.

JARDINE, P. E., FRASER, W. T., GOSLING, W. D., ROBERTS, C. N., EASTWOOD, W. J. and LOMAX, B. H. 2020. 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.

JARDINE, P. E., 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. 2021. Sporopollenin chemistry and its durability in the geological record: an integration of extant and fossil chemical data across the seed plants. *Palaeontology*, **64**, 285–305.

JARDINE, P. E., KENT, M. S., FRASER, W. T., KNORR, K.-H. and LOMAX, B. H. 2023. Data and code for "Uncovering a phylogenetic signal in plant biopolymer chemistry: a comparison of sporopollenin isolation approaches for use in palynological research". Figshare. https://doi.org/10.6084/m9.figshare.22182790

JIN, Y. and QIAN, H. 2019. V.PhyloMaker: an R package that can generate very large phylogenies for vascular plants. *Ecography*, **42**, 1353–1359.

JIN, Y. and QIAN, H. 2022. V.PhyloMaker2: an updated and enlarged R package that can generate very large phylogenies for vascular plants. *Plant Diversity*, **44**, 335–339.

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.

KENĐEL, A. and ZIMMERMANN, B. 2020. Chemical analysis of pollen by FT-Raman and FTIR spectroscopies. *Frontiers in Plant Science*, **11**, 352.

KORASIDIS, V. A., WING, S. L., NELSON, D. M. and BACZYNSKI, A. A. 2022. Reworked pollen reduces apparent floral change during the Paleocene-Eocene Thermal Maximum. *Geology*, **50**, 1398–1402.

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.

LILAND, K. H., ALMØY, T. and MEVIK, B.-H. 2010. Optimal choice of baseline correction for multivariate calibration of spectra. *Applied Spectroscopy*, **64**, 1007–1016.

LIU, F., PENG, H., MARSHALL, J. E. A., LOMAX, B. H., BOMFLEUR, B., KENT, M. S., FRASER, W. T. and JARDINE, P. E. 2023. Dying in the sun: direct evidence for elevated UV-B radiation at the end-Permian mass extinction. *Science Advances*, **9**, eabo6102.

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.

LOMAX, B. H., FRASER, W. T., 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.

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.

MACKENZIE, G., BOA, A. N., DIEGO-TABOADA, A., ATKIN, S. L. and SATHYAPALAN, T. 2015. Sporopollenin, the least known yet toughest natural biopolymer. *Frontiers in Materials*, **2**, 1–5.

MAYO, D. W. 2003. Survey of infrared and Raman group frequencies. 355–398. *In* MAYO, D. W., MILLER, F. A. and HANNAH, R. W. (eds) *Course notes on the interpretation of infrared and Raman spectra*. Wiley.

McGILL, R., TUKEY, J. W. and LARSEN, W. A. 1978. Variations of box plots. *The American Statistician*, **32**, 12–16.

MOORE, P. D., WEBB, J. A. and COLLINSON, M. E. 1991. *Pollen analysis*. Blackwell Scientific, 216 pp.

MUNDARGI, R. C., POTROZ, M. G., PARK, J. H., SEO, J., TAN, E. L., LEE, J. H. and CHO, N. J. 2016. Eco-friendly streamlined process for sporopollenin exine capsule extraction. *Scientific Reports*, **6**, 19960.

MUTHREICH, F., ZIMMERMANN, B., BIRKS, H. J. B., VILA-VIÇOSA, C. M. and SEDDON, A. W. R. 2020. Chemical variations in *Quercus* pollen as a tool for taxonomic identification: implications for long-term ecological and biogeographical research. *Journal of Biogeography*, **47**, 1298–1309.

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.

NELSON, D. M., URBAN, M. A. and HU, F. S. 2014. Spatiotemporal variation in the origin of C_4 grasses: $\delta^{13}C$ analysis of grass pollen from the southeastern United States. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **396**, 227–231.

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.

OLCOTT MARSHALL, A. and MARSHALL, C. P. 2014. Vibrational spectroscopy of fossils. *Palaeontology*, **58**, 201–211.

PARADIS, E. and SCHLIEP, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, **35**, 526–528.

R CORE TEAM. 2013. R: a language and environment for statistical computing. Version 4.2.1. http://www.R-project.org

RIDING, J. B. and KYFFIN-HUGHES, J. E. 2011. A direct comparison of three palynological preparation techniques. *Review of Palaeobotany & Palynology*, **167**, 212–221.

ROZEMA, J., BROEKMAN, R. A., BLOKKER, P., MEIJKAMP, B., de BAKKER, N., van de STAAIJ, J., van BEEM, A., ARIESE, F. and KARS, S. M. 2001a. 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: Biology*, **62**, 108–117.

ROZEMA, J., 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. 2001b. (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.

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

ROZEMA, J., 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.

RSTUDIO TEAM. 2022. RStudio: integrated development environment for R. Version 2022.07.0+548. https://www.rstudio.com/

SANTOS, R. D. A. and LEDRU, M.-P. 2022. Acid-free protocol for extracting pollen from Quaternary sediments. *Palynology*, **46**, 1–8.

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.

SEDDON, A. W. R., 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.

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.

STEVENS, A. and RAMIREZ-LOPEZ, L. 2022. Prospectr: miscellaneous functions for processing and sample selection of spectroscopic data. R package version 0.2.4. https://cran.r-project.org/web/packages/prospectr/index.html

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

UDDIN, M. J., ABIDI, N., WARZYWODA, J. and GILL, H. S. 2019. Investigation of the fate of proteins and hydrophilicity/hydrophobicity of lycopodium clavatum spores after organic solvent-base-acid treatment. *ACS Applied Materials & Interfaces*, **11**, 20628–20641.

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

VERDÚ, M., GÓMEZ-APARICIO, L. and VALIENTE-BANUET, A. 2012. Phylogenetic relatedness as a tool in restoration ecology: a meta-analysis. *Proceedings of the Royal Society B*, **279**, 1761–1767.

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.

WATSON, J. S., 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.

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.

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

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

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

ZUUR, A. F., IENO, E. N., WALKER, N. J., SAVELIEV, A. A. and SMITH, G. M. 2009. *Mixed effects models and extensions in ecology with R*. Springer.