Supplementary Information for

Rapid identification of bast fibers in ancient handmade papers by Py-GCxGC/MS based on improved characterization of lignin monomers

Bin Han[1], Yimin Yang[1], Bo Wang[2], Hongen Jiang[1], Michel Sablier[3]

[1]*Department of Archaeology and Anthropology, School of Humanities, University of Chinese Academy of Sciences, 100049 Beijing, China*

[2]*Xinjiang Uygur Autonomous Region Museum, 830000 Urumchi, China*

[3]*Centre de Recherche sur la Conservation (CRC, USR 3224), Muséum national d'Histoire naturelle, Ministère de la Culture et de la Communication, CNRS, 75005 Paris, France.*

Corresponding Authors: michel.sablier@mnhn.fr

Supplemental Methods

Supplemental Figures and Tables

Figure S1. Handmade papers used in ancient funeral rituals and ceremonies. (a) Picture and line graph showing that paper offerings resemble coin currency unearthed from tomb in central China of the Southern Song Dynasty (1127-1279 AD), (b) and (c) are line graphs indicating the making of joss paper (ghost money) for Astana cemetery in this study, (d) is the line graph bundle up of the joss paper (also see Figure 1 in the main text); (e) is the paper coffin unearthed from the Astana cemetery.

Figure S2. Map of the location of the Astana Cemeteries, Xinjiang, northwest China. (A) The location of Xinjiang in China. (B) The location of Turpan (the red rectangular area). (C) The location of the Astana Cemeteries.

Figure S3. Microscopic pictures examined for the studied Astana paper samples (i)-(iv).

Figure S4. Microscopic morphological features of the stained fibers from sample (i): (a)-(b), sample (ii): (c), sample (iii): (d)-(e), and sample (iv): (f)-(g).

Figure S5. Approximate content range of lignin in fibers used for the large-scale production of papers throughout the history of ancient China. The chart was drawn according to the data in the reference (Chen XP, et al., Journal of Silk, 2013,50(12):1-6; Liu F, et al, Journal of Henan Agricultural Sciences, 2011,40(10):120-122; Deng X., et al, Journal of Fungal Research, 2007(02):93-97).

Figure S6. Full chromatograms (a) of the Py-GC/MS analysis of the four Astana paper samples (i)-(iv) with the extended ROI region (b). The compound assignment in the ROI refers to Table S1. Test conditions of Py-GC/MS analysis can be found in reference [21] in the main text.

Figure S7. (a) and (b) show the cellulose structure and lignin monomer markers of p-hydroxyphenyl (H unit), guaiacyl (G unit), and syringyl (S unit) structures, respectively; (c) and (d) demonstrate the distribution of cellulose (blue) and lignin (red) in the fiber layers; (e) represents the distribution of different types of lignin in different fiber layers (red circle represents H type while green circle represents G and S type). Abbreviation in the chart: ML (middle lamella), P1 (primary wall), P2 (secondary wall), S1 (outer layer of P2), S2 (middle layer of P2), S3 (inner layer of P2), W (warty layer). The cellulose-lignin distribution in (c) is in reference to Catherine Bajon, Danièle Reis, Brigitte Vian, Le monde des fibres, Belin, 2006; the woody cell structure in (d) and (e) is in reference to Eero Sjöström, Wood chemistry-Fundamentals and Applications 2nd edition, Academic Press Inc., 1993.

Figure S8. Full Py-GCxGC/MS chromatograms obtained from the pyrolysis of the reference handmade paper sample. For the GCxGC separation of carbohydrate-lignin complexes and the grouping features of lignin monomers, refer to Fig. 4 in the man text and Fig. S9 in this file.

Figure S9. The lignin monomer structures listed in Table 1 in the main text.

Table S1 Assignment of plant markers detected by Py-GC/MS in the ROI of the investigated Astana papers with retention time (RT), base peak, main fragment ions, expected molecular weights, assigned formula and most likely attribution of the products.

Table S2 List of marker compounds detected within the ROI in the 2D chromatogram of the Py-GCxGC/MS analysis of the reference ramie, hemp, flax and jute paper samples and Astana paper samples with their respective retention times ($^{1}t_{R,}{}^{2}t_{R}$), molecular weights (MW), main peaks in the corresponding mass spectrum (base peak in bold), formula, identification and presence for each paper sample composition.

Supplemental Methods

Py-GCxGC/MS Analysis. Py-GCxGC/MS analysis was conducted with a Shimadzu QP 2010 Ultra mass spectrometer (Shimadzu, Champs-sur-Marne, France) equipped with a two-stage thermal modulator ZX 2 (Zoex, Houston, USA). Pyrolysis was performed using a vertical micro-furnace-type pyrolyzer PY-3030D (Frontier Lab, Fukushima, Japan) directly connected to the injection port of the gas chromatograph. The sample cups and the paper samples were weighed similarly to those used during Py-GC/MS experiments. A sample of 70 μg was weighed here. The sample cup was placed on top of the pyrolyzer at near ambient temperature. The sample cup was introduced into the furnace at 500 °C, and the temperature program of the gas chromatograph oven was started. The pyrolyzer interface was held at 320 °C. An Optima-5HT column (30 m × 0.25 mm I.D., 0.25 μm film thickness, Macherey-Nagel, Hoerdt, France) was used as the first dimension column, and a Zebron ZB-50 column (2.8 m × 0.1 mm I.D., 0.1 μm film thickness, Phenomenex, Le Pecq, France) was used as the second dimension column and for the loop modulator system. The separation was carried out at an initial constant pressure of 300 kPa in a constant pressure mode (which resulted in a flow rate varying from 1.0 mL/min to 0.4 mL/min during the acquisition) using Helium Alphagaz 2 as the carrier gas (Air Liquide, Bagneux, France). The ZX 2 two-stage thermal modulator employs a closed cycle refrigerator/heat exchanger to produce a -90 °C cooled air jet regularly modulated with a pulsed hot air jet. The optimized modulation period for two-dimensional chromatogram collection was 9 s with a programmed hot pulse of 0.350 s. A multi-step temperature program was used for the hot jet set at 200 °C for 30 min and subsequently raised to 280 °C until the end of the acquisition. The oven temperature was initially held for 1 min at 100 °C and then ramped at 2 °C min^{-1} to 325 °C, where it was held for 25 min. The total duration of GC analysis was 138.5 min. The injector was held at 280 °C and used in split mode (1:30 of the total flow). The mass spectrometer was operated at 20,000 $u.s^{-1}$, with a scan range from 50 to 500 u, using electron ionization at 70 eV. The interface was kept at 300 °C, and the ion source was at 200 °C. Data processing of the 2D raw data was achieved using GC Image software, version 2.4 (Lincoln, Nebraska). Identification of components was performed by comparing the mass spectra of unknown components with reference compounds from the NIST MS library (2011) and interpretation of the main fragmentations.

Data pretreatment and PCA/cluster analysis. In the pyrograms, only the compounds issued from lignin pyrolysis were considered for multivariate data analysis. During Py-GCxGC/MS experiments, compound identification and area integration in the total ion chromatogram (TIC) were performed manually within GC Image software, version 2.4 (Lincoln, Nebraska). The TIC peak areas (either raw or normalized values) were used as responses for multivariate data analysis to explore the most characteristic features of each paper sample. In the data matrix for PCA modeling, when one compound is absent, its corresponding descriptor value was set at 0 in the matrix.

After the data pretreatment procedure, the data were subjected to principal component analysis (PCA). Each analysis was considered as an individual sample in PCA, and the peak areas of the compounds distributed within the lignin monomers acted as variables. The PCA analyses were performed using Unscrambler X (Version 10.4) software from CAMO Software AS (Oslo, Norway). In the process of PCA modeling, data were centered, and different variable weighting parameters (weight as the same value 1 and weight as 1/STD) were performed. Here, STD is defined as the standard deviation calculated among each variable, which indicates how much the data are scattered for this response. Therefore, the weight of each variable is no longer related to its value but to the way it differs among samples. The data were mean centered in the course of PCA. Regarding clustering analysis, the default Euclidean distance measure was used. The results of hierarchical cluster analysis are illustrated graphically in a tree diagram called a dendrogram. In the dendrograms, a cluster is usually defined as two or more leaves that are joined by a common node and a branch. The height at the bottom of the dendrogram represents the distance between the samples or clusters.

Supplemental Figures and Tables

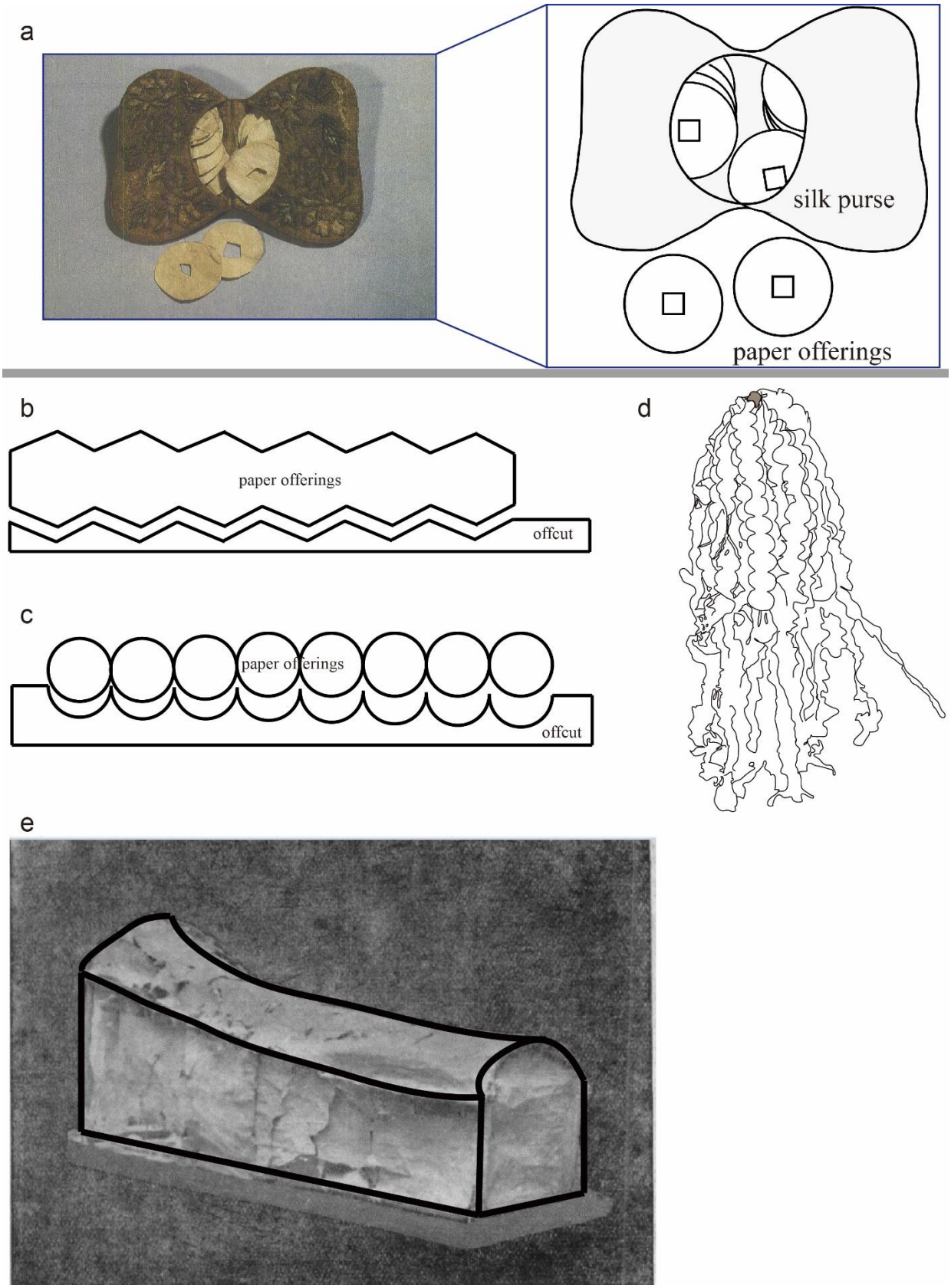

Figure S1 Handmade papers used in ancient funeral rituals and ceremonies. (a) Picture and line graph showing that paper offerings resemble coin currency unearthed from tomb in central China of the Southern Song Dynasty (1127-1279 AD), (b) and (c) are line graphs indicating the making of joss paper (ghost money) for Astana cemetery in this study, (d) is the line graph bundle up of the joss paper (also see Figure 1 in the main text); (e) is the paper coffin unearthed from the Astana cemetery.

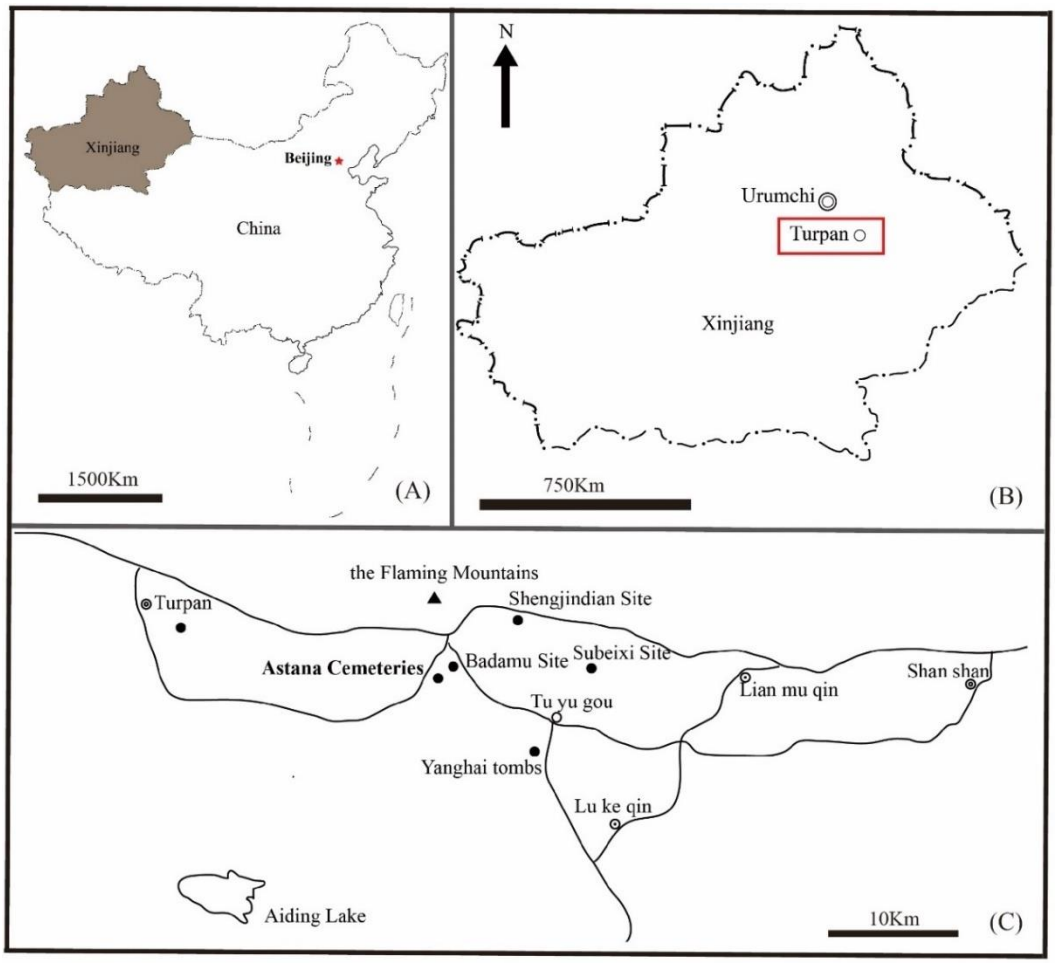

Figure S2 Map of the location of the Astana Cemeteries, Xinjiang, northwest China. (A) The location of Xinjiang in China. (B) The location of Turpan (the red rectangular area). (C) The location of the Astana Cemeteries.

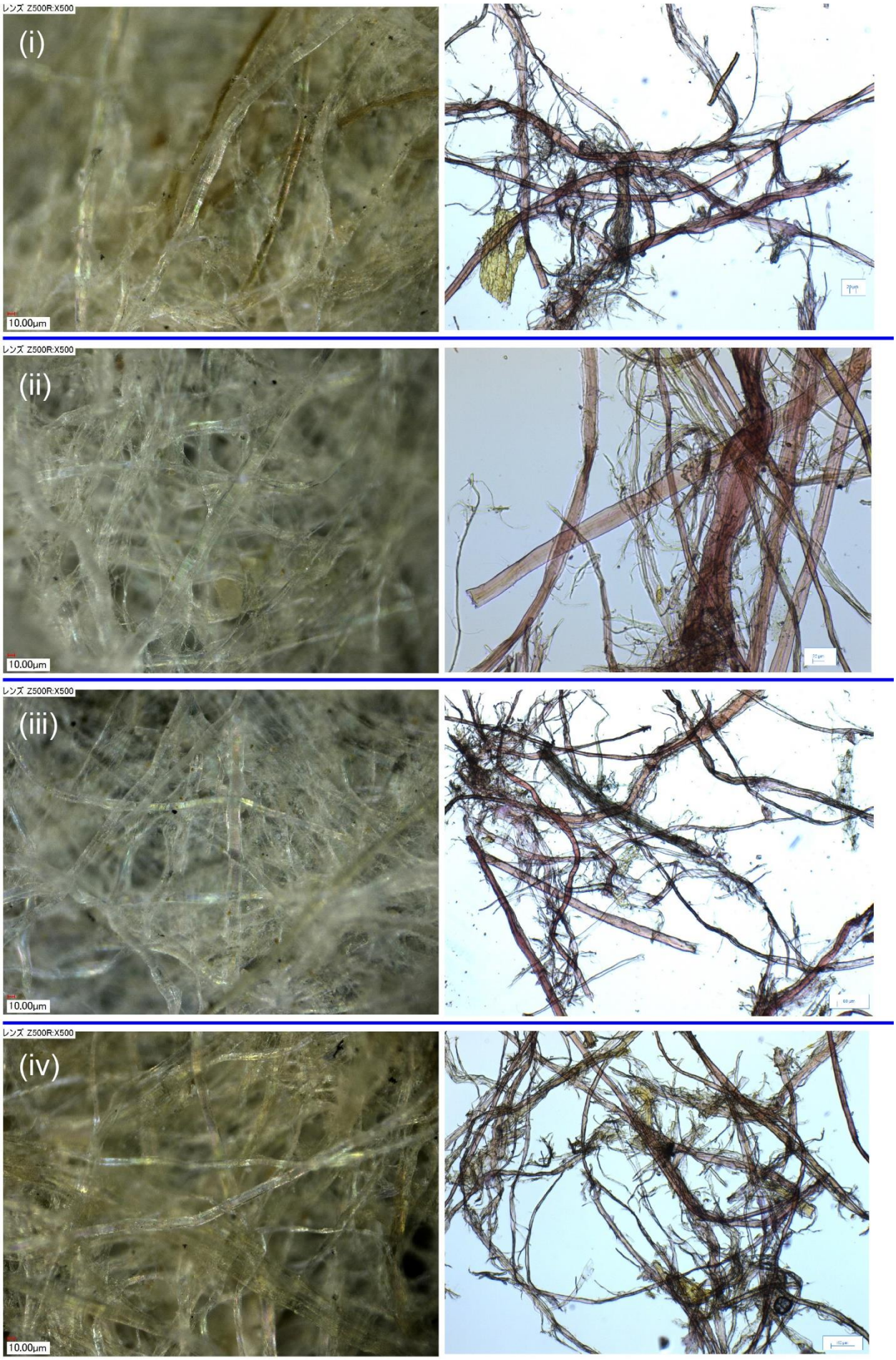

Figure S3 Microscopic pictures examined for the studied Astana paper samples (i)-(iv).

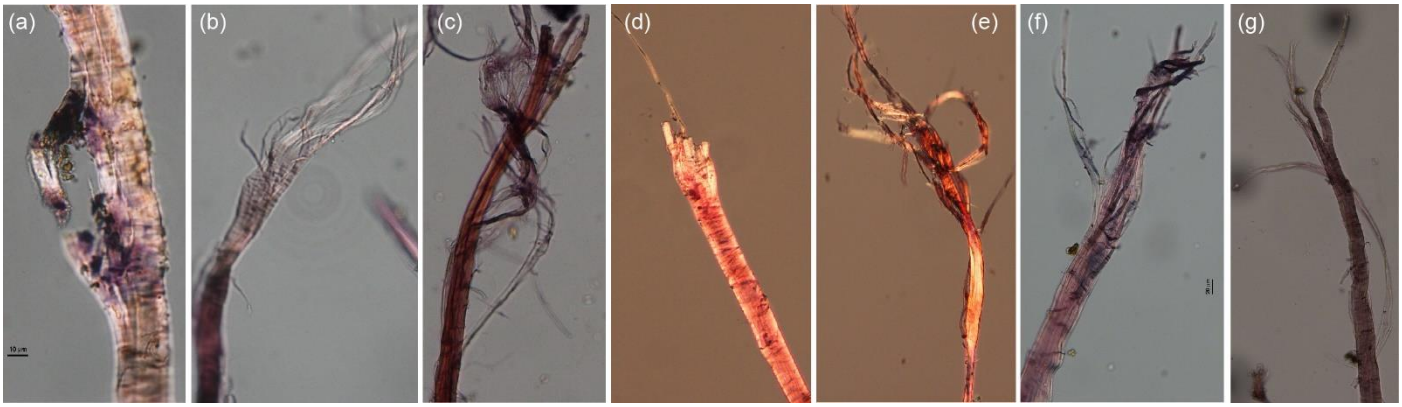

Figure S4 Microscopic morphological features of the stained fibers from sample (i): (a)-(b), sample (ii): (c), sample (iii): (d)-(e), and sample (iv): (f)-(g).

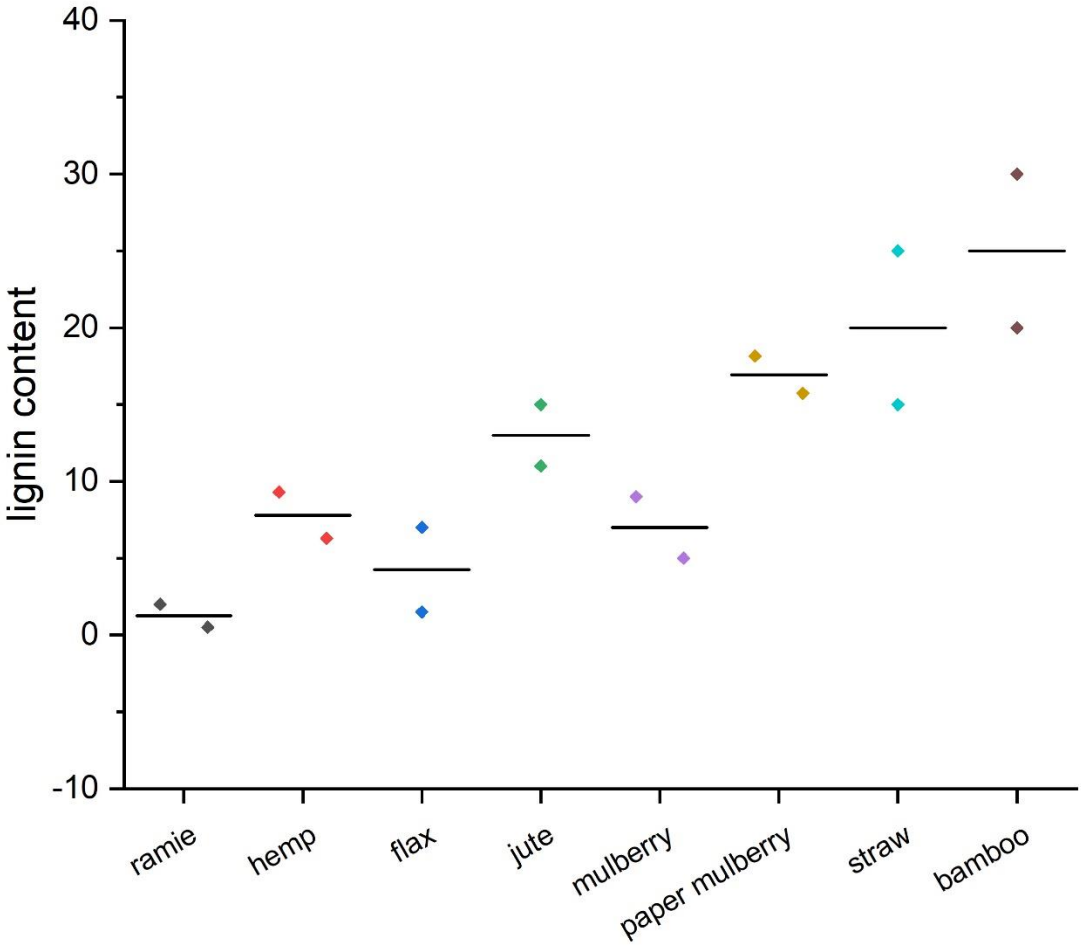

Figure S5 Approximate content range of lignin in fibers used for the large-scale production of papers throughout the history of ancient China. The chart was drawn according to the data in the reference (Chen XP, et al., Journal of Silk, 2013,50(12):1-6; Liu F, et al, Journal of Henan Agricultural Sciences, 2011,40(10):120-122; Deng X., et al, Journal of Fungal Research, 2007(02):93-97).

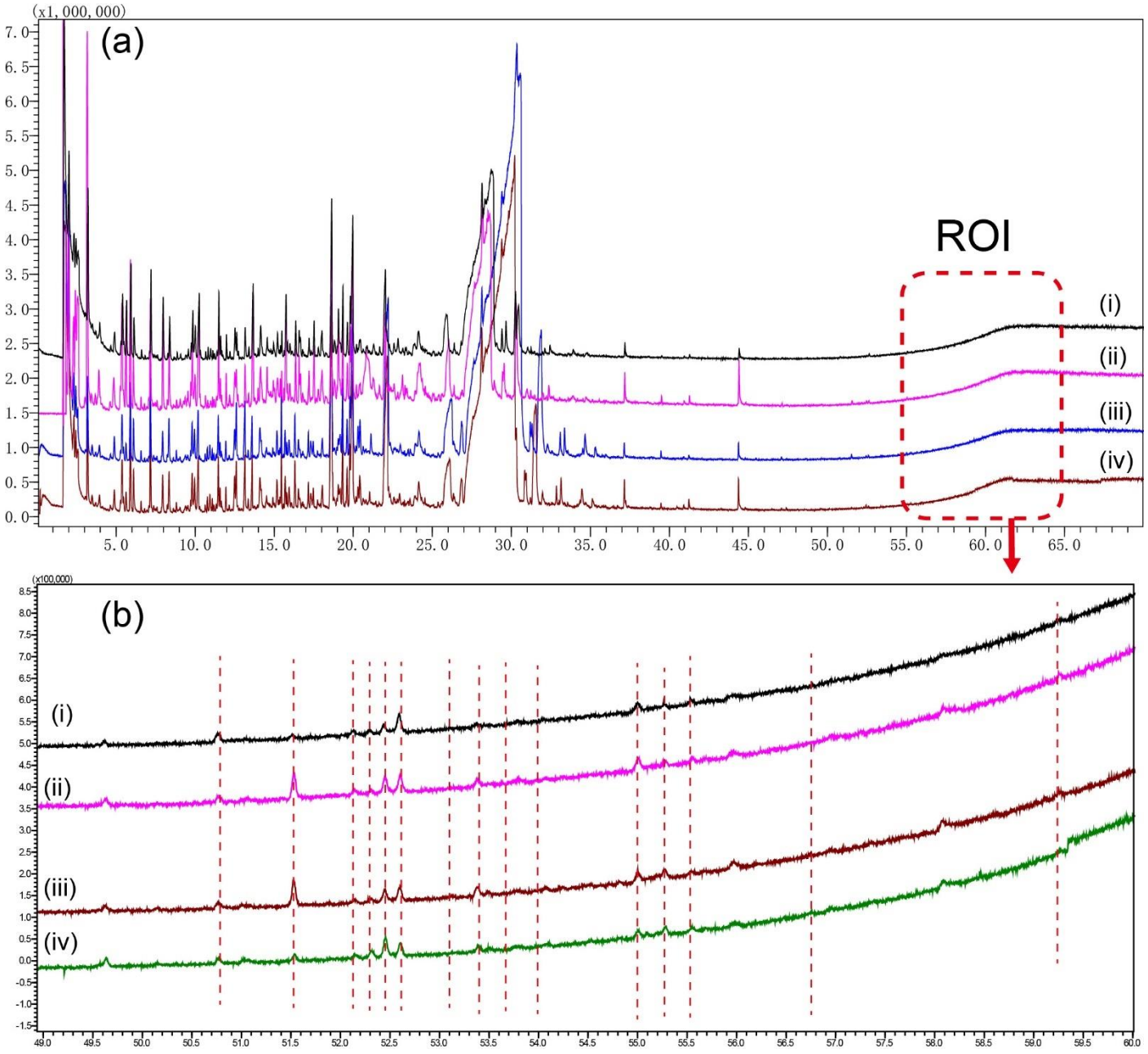

Figure S6 Full chromatograms (a) of the Py-GC/MS analysis of the four Astana paper samples (i)-(iv) with the extended ROI region (b). The compound assignment in the ROI refers to Table S1. Test conditions of Py-GC/MS analysis can be found in reference [21] in the main text.

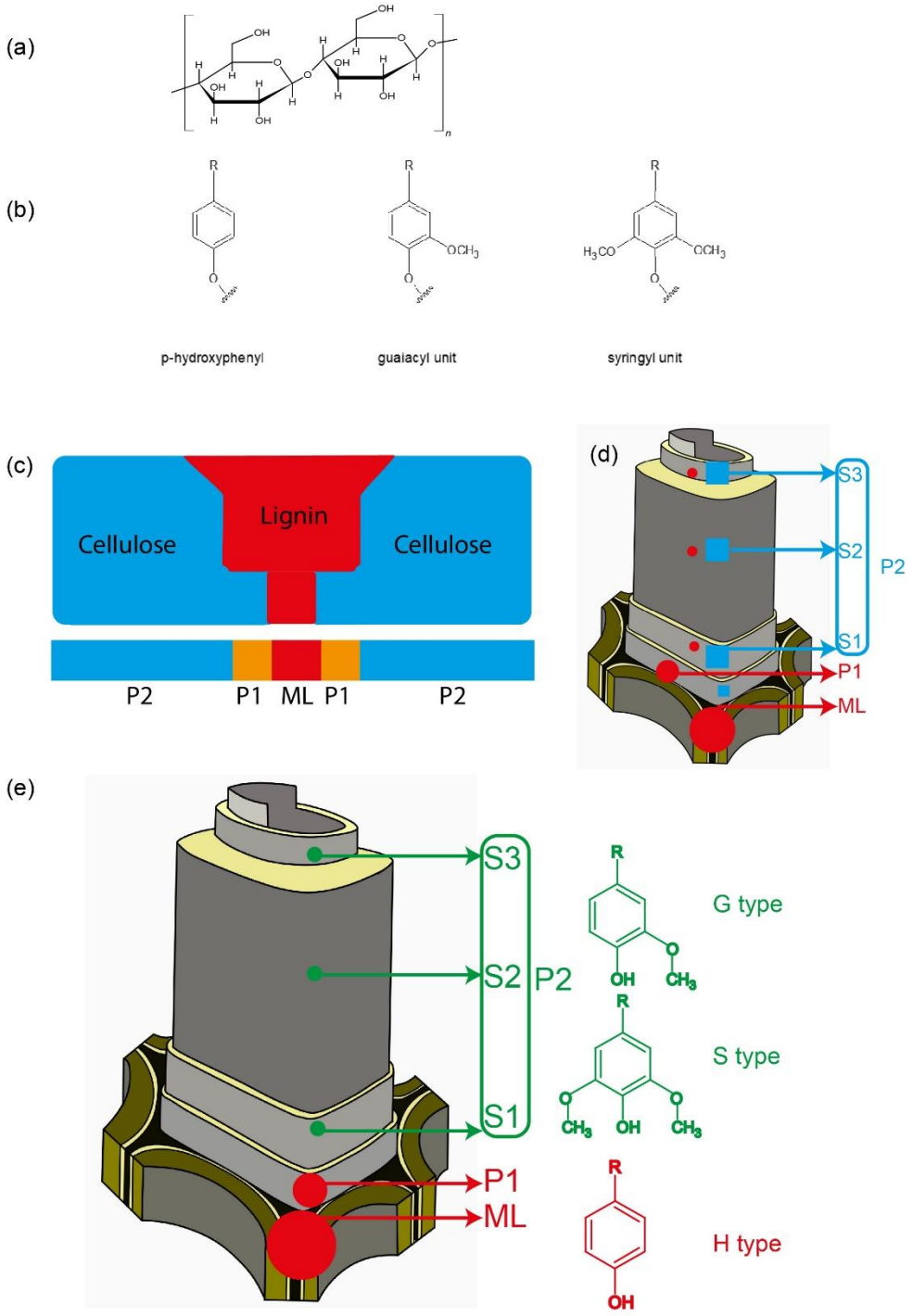

Figure S7 (a) and (b) show the cellulose structure and lignin monomer markers of p-hydroxyphenyl (H unit), guaiacyl (G unit), and syringyl (S unit) structures, respectively; (c) and (d) demonstrate the distribution of cellulose (blue) and lignin (red) in the fiber layers; (e) represents the distribution of different types of lignin in different fiber layers (red circle represents H type while green circle represents G and S type). Abbreviation in the chart: ML (middle lamella), P1 (primary wall), P2 (secondary wall), S1 (outer layer of P2), S2 (middle layer of P2), S3 (inner layer of P2), W (warty layer). The cellulose-lignin distribution in (c) is in reference to Catherine Bajon, Danièle Reis, Brigitte Vian, Le monde des fibres, Belin, 2006; the woody cell structure in (d) and (e) is in reference to Eero Sjöström, Wood chemistry-Fundamentals and Applications 2nd edition, Academic Press Inc., 1993.

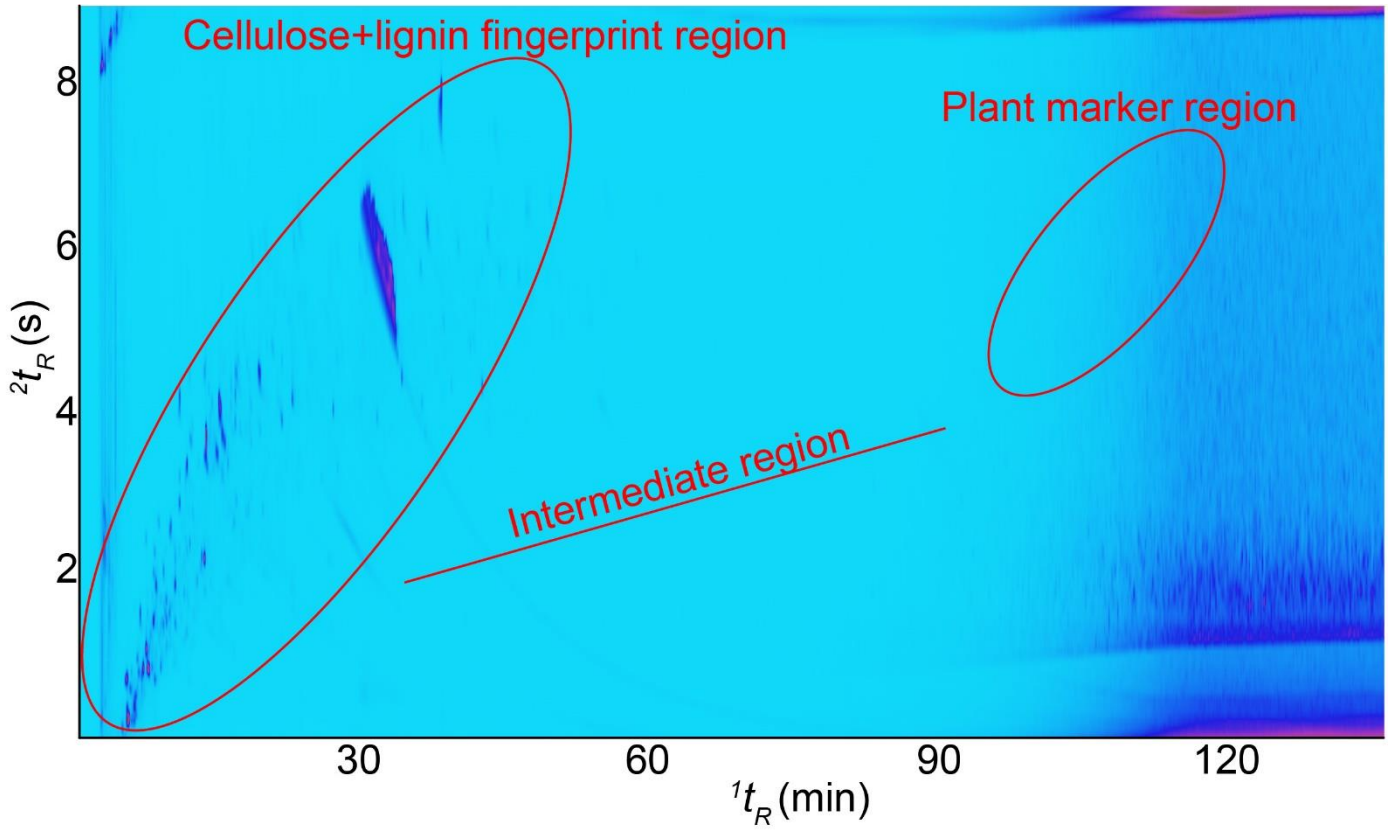

Figure S8 Full Py-GCxGC/MS chromatograms obtained from the pyrolysis of the reference handmade paper sample. For the GCxGC separation of carbohydrate-lignin complexes and the grouping features of lignin monomers, refer to Fig. 4 in the man text and Fig. S9 in this file.

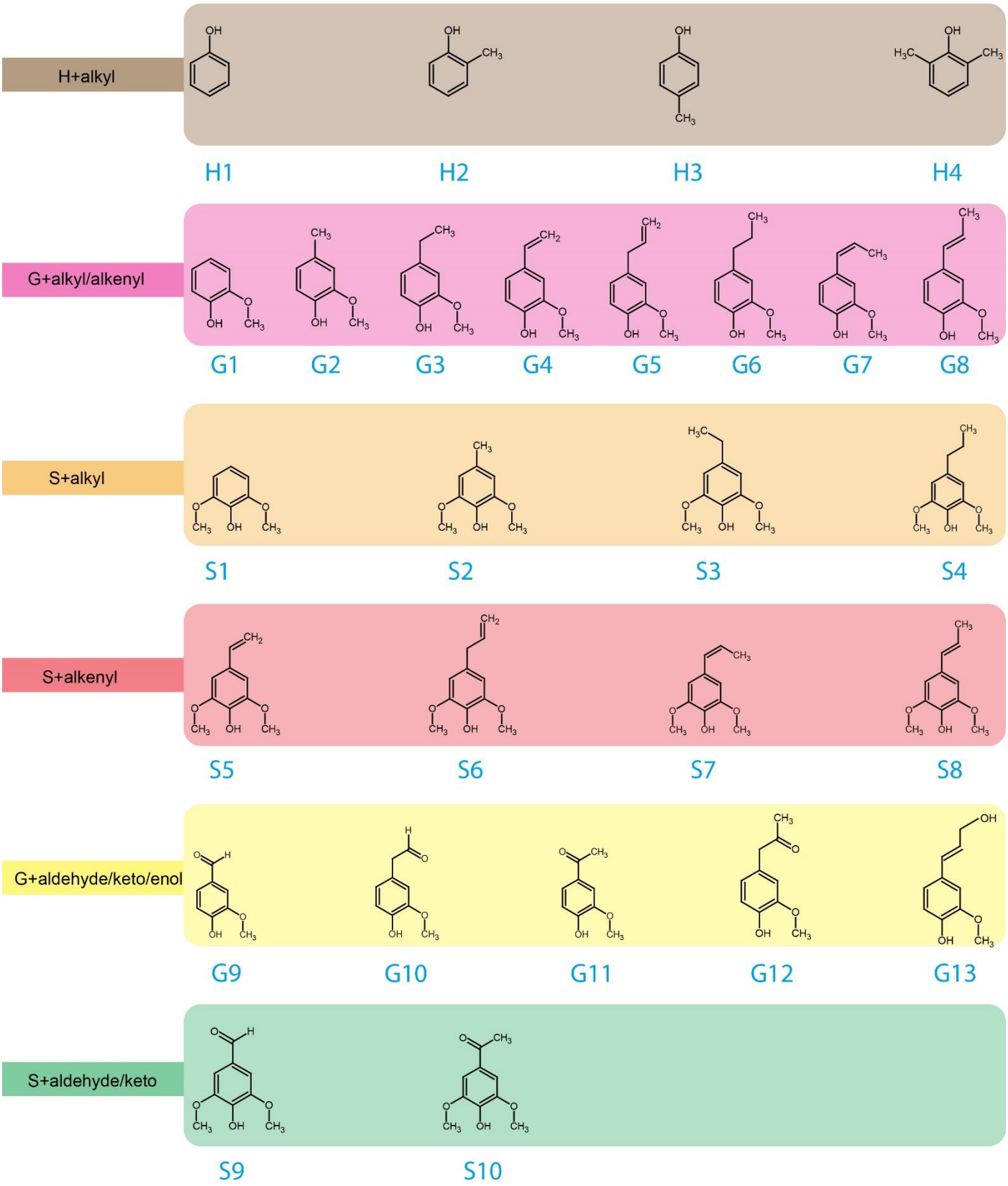

Figure S9 The lignin monomer structures listed in Table 1 in the main text.

Table S1 Assignment of plant markers detected by Py-GC/MS in the ROI of the investigated Astana papers with retention time (RT), base peak, main fragment ions, expected molecular weights, assigned formula and most likely attribution of the products.

Ret. Time (min)	Base peak	Main fragment ions	MW	formula	attribution
50.788	59	59,72,55,43,69	281	$C_{18}H35NO$	9-octadecenamide (oleyl amide)
51.123	59	59,72,43,69,55	NA	NA	long chain amide
51.530	69	69,81,95,121,93	410	$C_{30}H_{50}$	squalene
51.699	57	57,43,97,55,85	NA	NA	long chain alkene
52.138	105	105,95,91,43,81	368	$C_{27}H_{44}$	cholesta-2,4-diene or isomer
52.314	135	135,143,81,95,366	366	$C_{27}H_{44}O$	cholesta-4,6-dien-3-ol,
52.599	147	147,81,105,91,145,368	368	$C_{27}H_{44}$	cholesta-3,5-diene or isomer
53.111	57	57,43,71,97,85	NA	NA	long chain alkane
53.379	57	57,43,83,9,55	424	$C_{28}H_{56}O_2$	dodecyl hexadecanoate supposedly
53.666	59	59,72,55,43,69	NA	NA	long chain alkane amide
53.963	59	59,72,43,57,60	NA	NA	long chain alkane amide
55.019	135	135,81,55,95,43,119	NA	NA	cholestane structure
55.544	105	105,43,107,81,95	386	$C_{27}H_{46}O$	cholest-5-en-3-ol (cholesterol)
56.749	174	174,159,161,382,187	382	$C_{27}H_{42}O$	cholesta-3,5-dien-7-one
59.202	137	137,43,55,109,81	NA	NA	cholestane derivative

Table S2 List of marker compounds detected within the ROI in the 2D chromatogram of the Py-GCxGC/MS analysis of the reference ramie, hemp, flax and jute paper samples and Astana paper samples with their respective retention times ($^1t_{R,}{}^2t_R$), molecular weights (MW), main peaks in the corresponding mass spectrum (base peak in bold), formula, identification and presence for each paper sample composition.

NO.	^1t_R	^2t_R	MW	m/z	formula	identification	hemp	ramie	flax	jute	72TAM184:5	72TAM184:2	72TAM518:9	72TAM518:9
1	92.45	2.85	368	**145**,81,95,105,121	$C_{27}H_{44}$	NA Cholestane	x	x		x	x	x		x
2	93.5	2.25	410	**69**,81,68,95,67	$C_{30}H_{50}$	squalene	x	x	x	x		x	x	x
3	95.3	3.0	370	**91**,95,215,316,73	$C_{27}H_{46}$	cholest-2-ene				x				
4	95.75	3.25	368	**105**,91,55,95,81	$C_{27}H_{44}$	NA cholestane	x	x		x	x	x		x
5	96.20	3.40	384	**135**,143,81,119,95	$C_{27}H_{44}O$	cholesta-4,6-dien-3-ol,(3β)-	x	x		x	x	x		x
6	96.95	3.35	368	**147**,368, 145,81,91	$C_{27}H_{44}$	NA cholestane (cholesta-3,5-diene supposedly)	x	x	x	x	x	x	x	x
7	102.65	3.65	394	**135**,143,81,149,57		NA	t	x	x		x	x	x	x
8	103.4	3.6	396	**147**,91,145,105,107	$C_{29}H_{48}$	NA	t	t	x			x	x	x
9	107.3	5.05	382	**174**,187,161,159,91	$C_{27}H_{42}O$	Cholesta-3,5-dien-7-one	x	x	x	x	x	x	x	x
10	110.6	5.15		**174**,207,161,195	$C_{29}H_{48}$	NA			x				x	x
11	113.45	5.30	410	**174**,161,187,175,410	$C_{29}H_{46}O$	stigmasta-3,5-dien-7-one		x		t	x	x	x	x

t trace