Ink in secular and religious documents from the Pelliot collection (BnF, Paris): A study of ingredients and uses of ink in1st millennium Tocharian, Sanskrit and Chinese manuscripts from the Kucha and Dunhuang regions

E. Perruchini, A. Michelin, K. Laclavetine, G.-J. Pinault, M. Sablier

Ink in secular and religious documents from the Pelliot collection (BnF, Paris): A study of ingredients and uses of ink in 1st millennium Tocharian, Sanskrit and Chinese manuscripts from the Kucha and Dunhuang regions

E. Perruchini [a,*], A. Michelin [b], K. Laclavetine [c], G.-J. Pinault [d], M. Sablier [b]

[a] *Laboratoire de Recherche, Département des restaurateurs, Institut National du Patrimoine, 93300 Aubervilliers, France*
[b] *Centre de Recherche sur la Conservation (CRC, UAR 3224), Museum national d'Histoire naturelle, Ministère de la Culture et de la Communication, CNRS, 75005 Paris, France*
[c] *Centre de recherche et de restauration des musées de France (C2RMF), 75001 Paris, France*
[d] *Ecole Pratique des Hautes Etudes, 75014 Paris, France*

* Corresponding author.
E-mail addresses: elsa.perruchini@inp.fr (E. Perruchini), anne.michelin@mnhn.fr (A. Michelin), kilian.laclavetine@culture.gouv.fr (K. Laclavetine), georges.pinault@ephe.psl.eu (G.-J. Pinault), michel.sablier@mnhn.fr (M. Sablier).

A B S T R A C T

This paper presents the study of ink from first millennium manuscripts discovered in the Kucha and Dunhuang regions (West China) from the "Pelliot collection" stored at the *Biblioth` eque nationale de France* (Paris, France). The sixteen analyzed manuscripts written in black ink were selected partly because of their diverse character- istics (script, type, etc.) and partly because of the high possibility for their accurate dating between 4th to the 8th century. This period is defined by an important economic growth as well as cultural and artistic activities, facilitated by the fact that Kucha and Dunhuang played a key role in the passage of the Silk Road network connecting China, India and western Central Asia.

FTIR and Raman spectroscopies were first employed as they are non-destructive and non-invasive methods. They indicated that the ink present on the manuscripts was composed of soot and protein binder, however they did not provide direct insight on the precise origin of the soot (e.g pine soot or lampblack) and the protein binder used in the fabrication of inks. The analysis of the inked manuscripts by pyrolysis-comprehensive two-dimen- sional gas chromatography/mass spectrometry (Py-GCxGC-MS) – for which only a small amount of sample (microgram quantities) is required - and the exploration of the obtained multivariate datasets combining Hier- archical Clustering and PCA provide invaluable information on Asian inks, their manufacture and use in different historical periods. Our research suggests that the majority of the inks were made with pine soot. Furthermore, our results reveal different chemical signatures that may be indicators of difference in the production of pine soot (use of various species of pine wood, combustion parameters, etc.). Moreover, diverse proteinaceous binders and additives, known to have been used in ancient Chinese ink manufacture to ensure ink's consistency and for their antimicrobial and aromatic properties, were identified.

Keywords: Chinese inks, Py-GCxGC/MS, Hierarchical Clustering, PCA, Soot, Animal glue, Kucha and Dunhuang manuscripts

1. Introduction

The *Biblioth` eque nationale de France* (BnF) in Paris (France) holds an invaluable collection of Oriental manuscripts, including the extensive and rich Paul Pelliot (1878–1945) collection containing Buddhist manuscripts dating from the first millennium CE with a large number of those manuscripts not written in Sanskrit nor in Chinese but in To-charian. The Tocharian languages are closely related to the Indo-European linguistic family and were used during the 1st millennium AD in the west of China, in the actual Xinjiang region.

The texts in Pelliot's collection are almost all in Tokharian B and were discovered for the most part in the ruins of Buddhist sites from the Kucha area, which particularly prospered thanks to its location on the Silk Road linking China to the West (Walter, 1998). Some of these manuscripts were discovered in Dunhuang, located in the extreme west of the actual Gansu province, marking the starting point of the Silk Roads to the West. These trade routes spanning from China through India and Central Asia, used by pilgrims, allowed for the transmission

Table 1
List of the 16 inked manuscripts analysed for this study and their specificities (period, city and region of discovery, script, type).

Manuscript ID	Date	Writing support	City	Region	Script	Script 2	Type 1	Type 2
PK AS 7F	7th century	Hemp	Douldour Aqour	Kucha	Koutchean	NTBa	Religious	Literary
PK AS 8B	7th century	Hemp	Subaˇsi	Kucha	Koutchean	NTB semi-cursive	Religious	Magical
PK AS 12M	6th century	Hemp	Douldour Aqour	Kucha	Koutchean	NTB archaic	Religious	Buddhist drama
PK AS 13A	NA	Hemp	NA	Kucha	Koutchean	NTBa	Religious	Buddhist legend
PK NS 6	7th century	Hemp	Subaˇsi	Kucha	Koutchean	NTBa	Religious	Magical
PK NS 95	7th century	Hemp	NA	NA	Koutchean	NTBa	Religious	Vinaya
PK DAM 507 36	8th century	Hemp	Douldour Aqour	Kucha	Koutchean	NTB Cursive	Secular	Judicial
PK DAM 507 40–42	8th century	Hemp	Douldour Aqour	Kucha	Koutchean	NTB Cursive	Religious	Vinaya
PK LC XXVI	8th century	Hemp	Subaˇsi	Kucha	Koutchean	NTB Cursive	Secular	Admin/Economic
PK LC A Brāhmī	8th century	Hemp	Douldour Aqour	Kucha	Koutchean	NTB Cursive	Secular	Admin/Economic
PK LC A Chinese	8th century	Hemp	Douldour Aqour	Kucha	Chinese	NTB Cursive	Secular	Admin/Economic
PC DA 1	721 CE	Hemp	Douldour Aqour	Kucha	Chinese	NA	Secular	Military
PK AS 3B	7th-8th century	Hemp	Dunhuang	Dunhuang	Koutchean	NTBab	Secular	Medicinal recipes
PS VERT 35	6th century	Hemp	NA	NA	Sanskrit	T r3	Religious	Literary
PS VERT 16	4th-5th century	Hemp	NA	South of Turkestan?	Sanskrit	T v1	Religious	Literary
PS SUTRA DS3	7th century	Hemp	NA	NA	Sanskrit	NTBa	Religious	Literary

and exchange of religion, culture, as well as techniques and artistic forms (Neelis, 2011).

Therefore, these religious and secular texts are of particular interest to linguists, philologists and historians who want to reconstruct the history of Buddhist written culture between 400 and 1000 CE. Since October 2018, the project "History of the Tocharian Texts of the Pelliot Collection – HisTochText" aims at a multidisciplinary study of these manuscripts using all available methods. Its originality lies in combining textual analysis proper, carried out with the resources of philology, with the material analysis of the support of the manuscripts- which are mostly on paper and a minority on wood- and coloured materials such as ink. When it comes to Asian inks, it is surprising to realize how little we know about the fabrication processes and ingredients involved. Moreover, the majority of the information are related to Chinese ink. Chinese ink recipes dating back to the 5th century CE reveal that three main ingredients in different proportions were mixed together to form inksticks: 1) a pigment (soot of various origins), 2) an organic binder (animal glue) with the dual function of holding these pigments in the ink sticks and binding these carbon pigments to the surface of the writing surface and finally 3) additives (mainly organic) whose role were to colour the ink (e.g. plant extracts), to ensure its consistency by working as consolidating agents to keep the inksticks from cracking during the drying stage of the moulding process (e.g. lacquer tree sap, vegetable oils) or added for their antimicrobial and aromatic properties (e.g. camphor, sandalwood, cloves, musk, pomegranate peels) (Chi-chen, 1930; Needham and Tsuen-Hsuin, 1985; Franke, 1962). Although the ingredients used in ink production are generally few in number, the exact composition, preparation and quantity of each ingredient as well as the way in which they were mixed are parameters that can vary considerably. This information is rarely available to us, even though these parameters will play a role in the quality, colour, gloss, etc. of the ink obtained. Moreover, seldom is known about the purpose(s) and context(s) in which these inks were used (Ren et al., 2022).

In an attempt to fill this gap, in this extremely fragmented history of the development and use of Asian inks and although the techniques used for the manufacture of the inks present on the Tokharian, Sanskrit and Chinese manuscripts from the Kucha region and Dunhuang may have been influenced by other historical origins (such as the Indian world, other regions of Central Asia), their chemical analysis will provide us with invaluable information on Asian inks, their manufacture, and their use.

Determination of chemical composition of writing materials generates important data for addressing cultural and historical questions that cannot be solved by historical and philological methods alone (Han et al., 2019; Abdul Razak et al., 2017; Ren et al., 2018; Wei et al., 2012; Luo et al., 2019). In its individual materiality, each manuscript contains evidence of the production processes involved in ancient Asian ink making. The traces that can be measured and ascribed to certain ingredients provide insights into the history of a manuscript. There are at least two prerequisites for the successful implementation of such measurements. The first and the most important one for the investigation of historical objects is the use of techniques that are non-destructive or require only minimal sampling. In case of sampling, the latter should preferably stay unchanged by analysis and available for further studies. In this respect, Raman spectroscopy and Fourier transform infrared spectroscopy (FTIR) were first used as they offer a non-invasive/non-destructive approach to the study of inks, in particular by identifying the nature of the soot and the presence of proteinaceous material (Ren et al., 2018). However, while they provide a very useful and simple methodology to detect the presence of inorganic pigment and organic binders, they rarely detect the precise nature of material used in the manufacture of ancient and/or degraded complex mixtures (Ren et al., 2022). On the other hand, pyrolysis-comprehensive two-dimensional gas chromatography/mass spectrometry (Py-GCxGC/MS) micro-destructive analysis represents a valuable method for obtaining qualitative and semi-quantitative information on a great diversity of materials found in inks (Perruchini et al., 2022). For instance, this technique allows us to detect Polycyclic Aromatic Hydrocarbons (PAHs) which are compounds characteristic of soot. Previous works have looked at the formation and occurrence of PAHs, in order to determine the source of the soot (softwood vs hardwood) (Guillon et al., 2013; Achten et al., 2015; Orasche et al., 2013; Avagyan et al., 2016; Rogge et al., 1998; Gonçalves et al., 2011; Simoneit et al., 2000) as well as the impact of various combustion conditions (temperature, presence of bark during the combustion, the moisture content of the wood, amount of wood load in a combustion chamber, the type of stoves and ovens, etc.) (Avagyan et al., 2016; Rogge et al., 1998; Gonçalves et al., 2011; Simoneit, 2002; Weimer et al., 2008; Pettersson et al., 2011; Boman et al., 2011; Shen et al., 1994; Kim Oanh et al., 1999). This study presents new data using Py-GCxGC/MS analysis to identify the nature and manufacturing processes of soot as well as the other ingredients used in the composition of ancient inks and highlight similarities/differences with what is known about Chinese ink manufacturing techniques.

2. Material and method

2.1. Description of the manuscripts

As a part of this extensive study, 16 manuscripts from the Pelliot Collection stored in the BnF were selected for their representativeness, historical value, and a high possibility for their accurate dating (Table 1) (Luo et al., 2019). They all presented texts written in black ink (Figs. D50–D64 in Supplementary material) and the writing support has been identified as hemp in almost every manuscript often mixed with

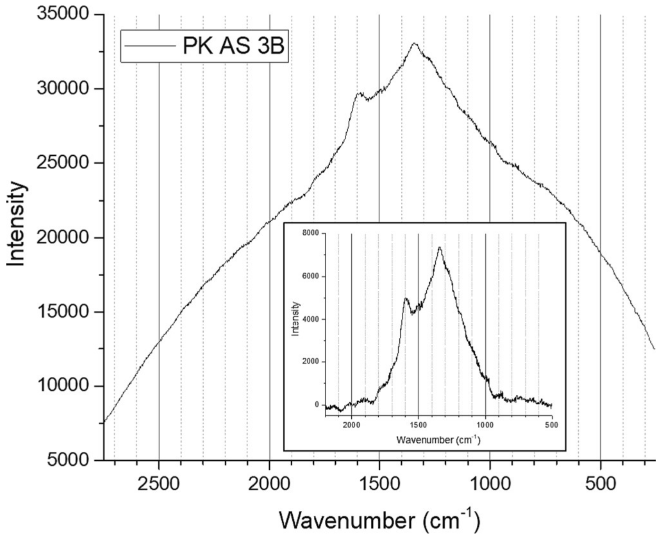

Fig. 1. Raman measurements realized on one inked spots of the manuscript PK AS 3B showing two peaks at around 1600 and 1345 cm^{-1}, which are characteristic of black carbon (zoom with baseline in the window).

other fibers thanks to macroscopic and microscope observations (Table 1).

The majority of the manuscripts are dated to the second half of the 1st millennium CE and were discovered in ruins of Buddhist sites, in the Xinjiang region (West of China) and belong to a relatively circumscribed area around Kucha (cities of Douldour Aqour and Subaˇsi), in the Tarim basin, on the northern fringe of the Taklamakan desert. Ten manuscripts are written in various styles of the Tocharian Brāhmī script and two in Chinese. One of the manuscripts, PK AS 3B, is a medical document discovered in Dunhuang, in the actual Gansu province (West of China), East of the Taklamakan desert. Finally, three manuscripts in Sanskrit were also analyzed for this study.

2.2. *Reference material*

In an attempt to identify characteristic compounds of soot, in addition to referring ourselves to ancient texts sources and literature, this work incorporates the use of several reference samples (Perruchini et al., 2022): 1) a traditional black Chinese inkstick from the old Hu Kai Wen ink workshop made from industrially produced oil soot and labelled as blackHKW, 2) a traditional Chinese black inkstick from the "Chinese Brush Painting Set" of Jane Dwight made from oil soot and labelled as blackJD, 3) a sample of pine soot ink, made from the soot of pine wood which is produced by the anoxic burning of pine wood and labelled as blackIP.

2.3. *FTIR spectroscopy*

A preliminary investigation was conducted using FTIR spectroscopy on a selected number of manuscripts with the objective of gathering information about the presence of organic materials.

The infrared spectra were recorded using a Bruker Alpha portable FTIR spectrometer with a reflectance sampling module for contactless analysis (at a working distance of 15 mm) coupled to a video camera in order to locate the analysed area. The analytical spots are about 5 mm in diameter. Spectra were acquired over the range of 7000–375 cm^{-1} with a spectral resolution of 4 cm^{-1} and 128 scans per spectrum. A gold

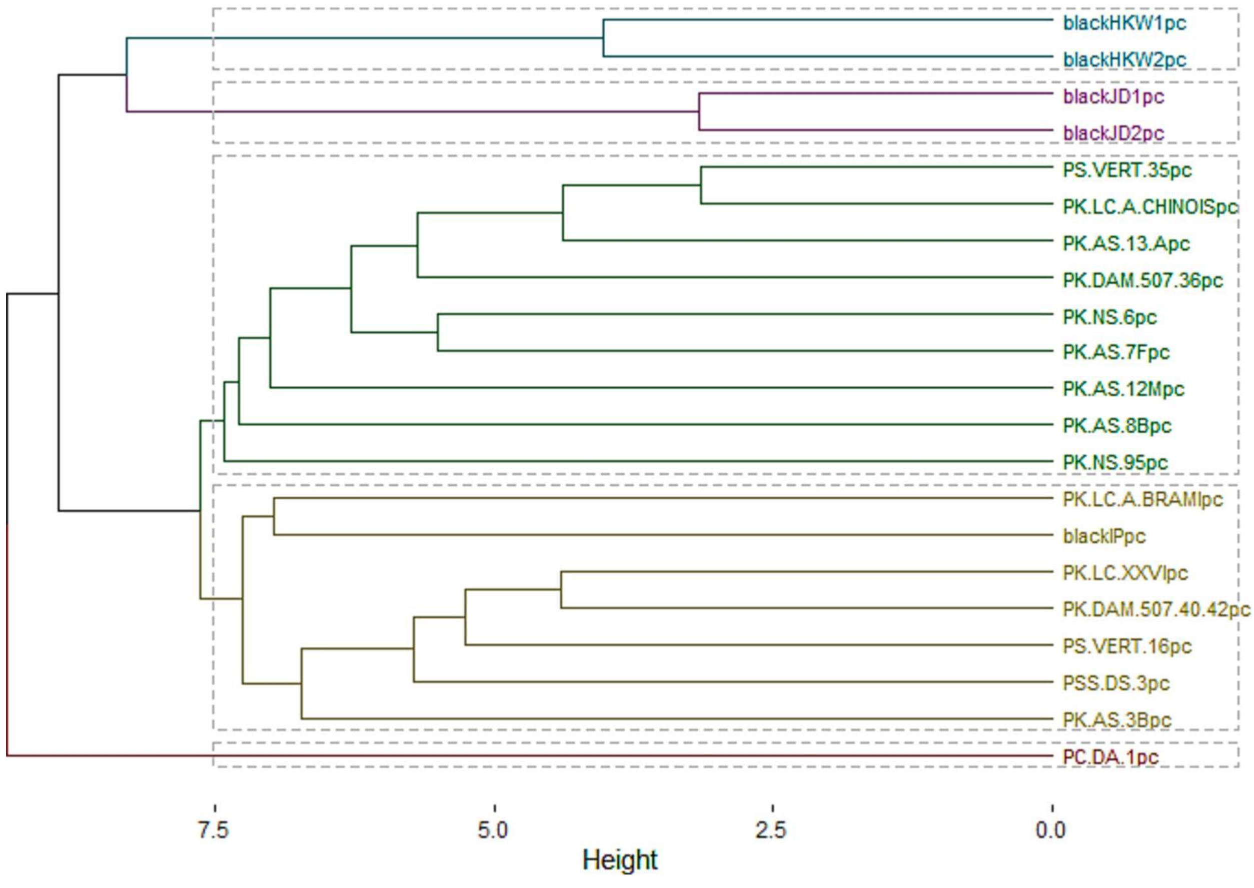

Fig. 2. Dendrogram of the hierarchical clustering of the analysed inked manuscript and the modern reference samples of traditional inks based on the distributions of PAHs, characteristic compounds of soot. Average linkage method, c = 0,84, k = 5.

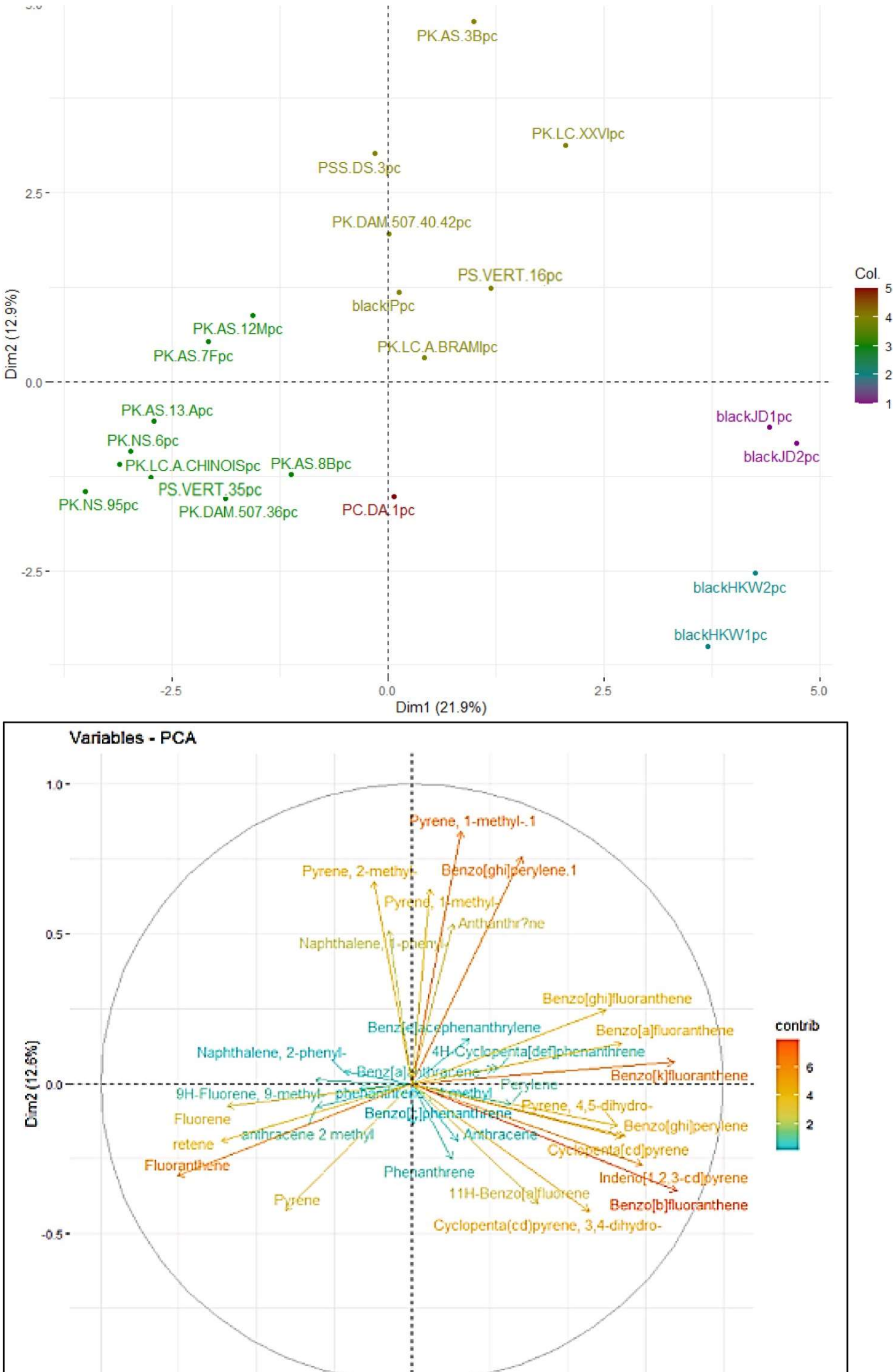

Fig. 3. Above - the first principal component (describing 35% of total variance) from pca of the analysed inked manuscript and the modern reference samples of traditional inks based on the distributions of PAHs, characteristic compounds of soot. the colour has been attributed according to the five clusters of the dendrogram. below - Graphical output of the variables' contributions for the first principal component with components in red and orange contributing the most to the definition of the first principal component. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mirror was used to acquire a background spectrum. The Bruker's OPUS software was used to process the data.

2.4. Raman spectroscopy

Raman spectroscopy was used on the same selection of manuscripts with the objective of gathering information about the presence of

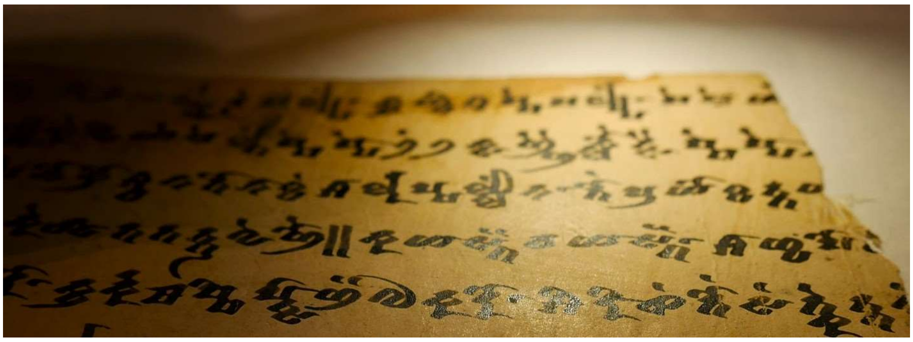

Fig. 4. PK AS 3B ©Emilie Arnaud Nguyeˆnˇ.

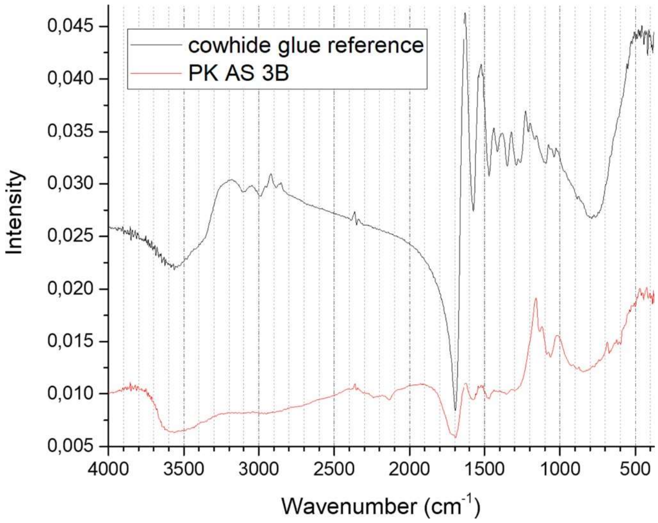

Fig. 5. Reflectance FTIR spectra of an inked spot of the manuscript PK AS 3B (in red) showing peaks at around 1630 and 1530 which are characteristic of proteinaceous binders (example of a cowhide glue reference in grey). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

inorganic materials, such as soot.

The Raman spectra were acquired using a portable Jobin-Yvon Horiba Raman spectrometer: an HE785 coupled with a GaAlAs diode laser (785 nm) and based on a fixed grating of 685lines/mm. This spectrometer uses fiber optics and a SuperHead interface for the measurement, edge filters to eliminate the Rayleigh line, and a Peltier cooled CCD detector for signal recording. In order to avoid any degradation, the laser power at the measurement point is checked and adjusted below 700 μW for the 785 nm lasers. The spectral resolution was 5 cm^{-1} and the recording time was between 12 s and 4 min depending on the response of the area analysed. The spot analysed was focused with a long working distance x40 objective lens. The software LabSpec 5 was used to process the data.

2.5. *Py-GCxGC/MS*

A sample of each manuscript (ca. 150 μg) was then analysed by Py-GCxGC/MS following the method presented in Perruchini *et al*, 2022 (Perruchini et al., 2022). Pyrolysis was performed using a Frontier Lab pyrolyzer PY-3030iD directly connected to a Shimadzu QP2010-Ultra mass spectrometer equipped with a two-stage thermal modulator ZX 2. The GC injector, was operated in split mode at a split ratio of 30:1 and set at 280 ◦C. The GCxGC setting utilizes two columns: an OPTIMA-5HT column (30 m × 0.25 mm I.D., 0.25 μm film thickness, Macherey-Nagel, Hoerdt, France) was used as first dimension column and a Zebron ZB-50 (2.8 m × 0.1 mm I.D., 0.1 μm film thickness, Phenomenex, Le Pecq, France) was used as a second-dimension column. The temperature of the interface was held at 320 ◦C. The sample cup was introduced into the furnace at 500 ◦C, the chosen temperature for the analysis of the paper samples. Data processing of the Py-GCxGC/MS raw data was achieved using GC Image software, version 2.4. In regard to PyGCxGC/MS data processing, relative proportions of the compounds detected were used instead of using peak volumes representatives of compounds relative concentrations, rendering the data suitable for standard statistical tools and allowing us to compare the composition of samples of diverse natures and samples of inked manuscripts for which the quantity of ink analyzed cannot be known. Percentages of compounds detected in each analysis were calculated considering the ratios of each compound of interest to the sum of all the considered compounds.

All statistical analyses and figure preparations were performed using RStudio, version 1.4.1106. For this work, we used two complementary methods: Principal component analysis (PCA) and hierarchical cluster analysis. The same (Euclidean) distance between individuals was used for each method. Hierarchical classification and PCA have a similar perspective, the exploratory analysis of a rectangular table but the mode of representation differs, and we achieve different information. Hence the idea of combining the approaches to obtain a rich methodology, and several points of view which can only reinforce the solidity of our conclusions.

PCA provides principal components which are orthogonal synthetic variables (dimensions, presenting the important variables which have a big weight in explaining variations in a dataset). Since the PCA does not provide any information on the position of the points in the other dimensions, two points close in one set of dimensions can be in the same class in the hierarchical tree (and therefore not too far from each other along the other dimensions) or in two different classes (because they are far from each other along the other dimensions). PCA can thus be used as a pre-processing for hierarchical cluster analysis which will, on the contrary, take into consideration all the variables to class the samples in a dendrogram.

Several parameters have to be considered when working with dendrogram visualization. The way the pairs are combined depends on a different calculation between each combined pair and the other samples. We call “linkage” the measure of dissimilarity between two groups of observations or clusters. Several algorithms can be used in hierarchical cluster analysis to determine how the “linkages” are created, however the most common are: 1) the complete linkage method where the dissimilarity of the groups is measured by the minimum dissimilarity between the variables in the groups, 2) the single linkage method which involves measuring the similarity to the closest pair and finally 3) the group average method, which is the most widely used in hierarchical clustering algorithms, using the average dissimilarity between pairs of points in two groups (Kassambara, 2017). Whatever the method, this pairing process continues until all the elements are merged into a single cluster. There is no absolute rule saying one linkage is better than another (the “linkages” not presented in the texts can be found in

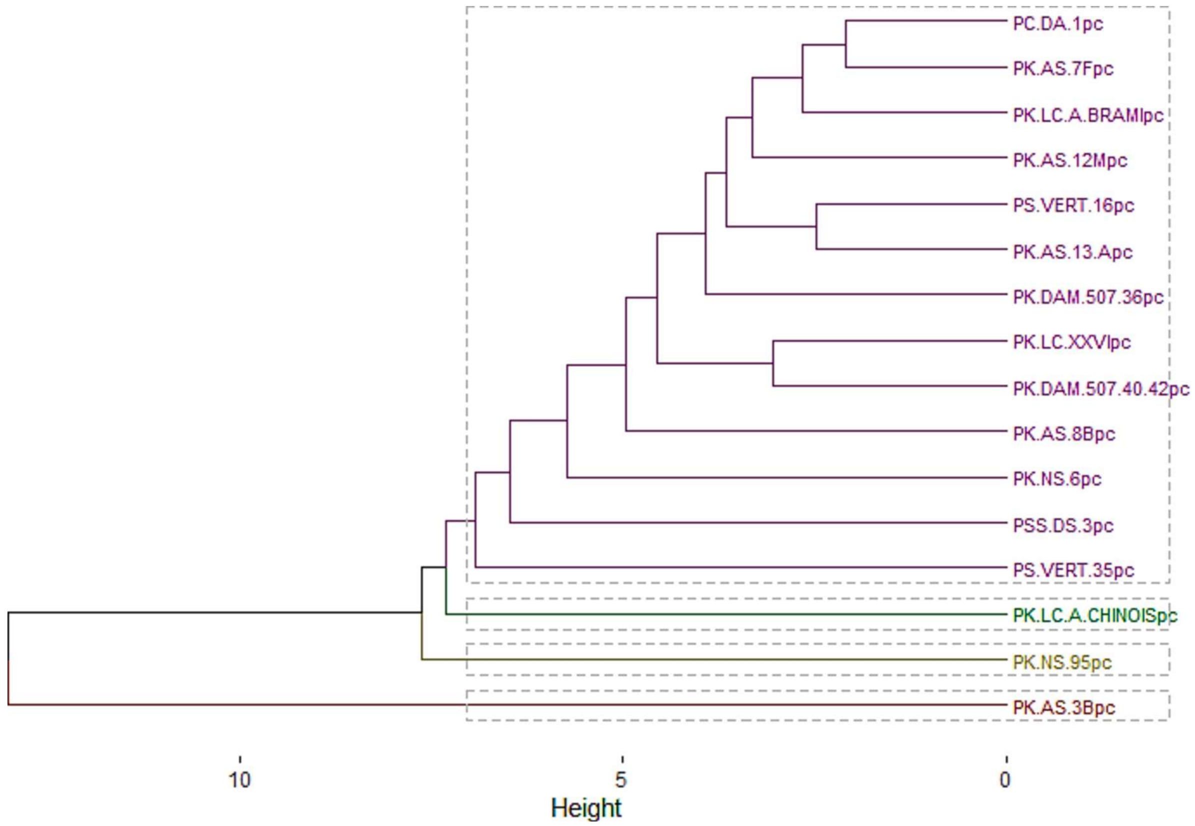

Fig. 6. Dendrogram of the hierarchical clustering of the analysed inked manuscript samples based on the distributions of compounds characteristic of proteinaceous materials. Average linkage method, c = 0,97, k = 4.

Supplementary material), however the cophenetic correlation is an indicator of the quality of a hierarchical cluster or dendrogram, helping us choose the clustering method which best preserves the pairwise distances between the original dataset. A value close to 1 indicates that the dendrogram reflects the original distances. Another important parameter to consider is the optimum number of clusters of the dendrogram (named *k* in this study). There are multiple methods used in order to determine *k* for each dataset, such as the elbow and silhouette methods (Kassambara, 2017). For this work we used the silhouette method. The silhouette analysis can be used to study the separation distance between the resulting clusters. The silhouette plot displays a measure of how close each point in one cluster is to points in the neighboring clusters and thus provides a way to assess parameters like number of clusters visually. This measure has a range of [-1, 1]. Where 1 means that points are very close to their own cluster and far from other clusters and —1 indicates that points are close to the neighboring clusters.

Finally, another important parameter to consider when creating a dendrogram and a PCA is the normalization or non-normalization of the data. Normalizing the data gives equal weight to all variables in the calculation of distances. While this equal treatment of variables avoids any scale effect, it also has the consequence of not distinguishing variables carrying important information. As we shall see later, PCAs can be useful in viewing both standardized and non-standardized data before analysis in order to extract the most information.

Combining multivariate analyses, such as PCA and hierarchical cluster analysis can, thus, help us establish relevant, but sometime hidden, relationships between variables and how classes are created, and better understand on the bases of which variables, individuals are grouped and/or excluded in classes.

3. Results and discussion

The, previously described, multi-analytical study was applied to identify and characterize the different components of the inks used in the 16 manuscripts analyzed. While Raman and FTIR spectroscopy allow us to identify the nature and presence of the pigment and binder used in the fabrication of inks, Py-GCxGC/MS gives access to a complete collection of data characterizing the different ingredients. This characterization is further validated by the use of reference samples.

3.1. Soot characterization

All the Raman spectra obtained from the analysis of the few selected manuscripts (Table 1 in Supplementary material) showed two peaks at around 1600 and 1350 cm^{-1} (Fig. 1), which are characteristic of black carbon such as pine soot and lampblack (Ren et al., 2018).

A large amount of PAHs, were identified using Py-GCxGC/MS.

The clustering analysis of our reference samples of traditional inks and manuscript samples to their PAHs distribution was performed in the hope to characterize the soot used in their fabrication. After normalization of the data, in order to test the robustness of the clustering, several clustering methods (complete, average and single) have been employed and their correlation coefficient calculated. The average linkage presented the highest correlation coefficient (c = 0,84) is thus presented here grouping the samples into clusters according to the similarities with their fingerprints (PAHs distribution). Moreover, to successfully perform clustering, it is necessary to predict the optimum number of clusters *k* the data should be grouped into according to the closeness measure selected. The results of the silhouette criterion provided the highest number of clusters (i.e. 5) and was thus used in the dendrogram visualization (Fig. 2).

In Fig. 2, the traditional inks made with oil soot (cluster 1 and 2) are

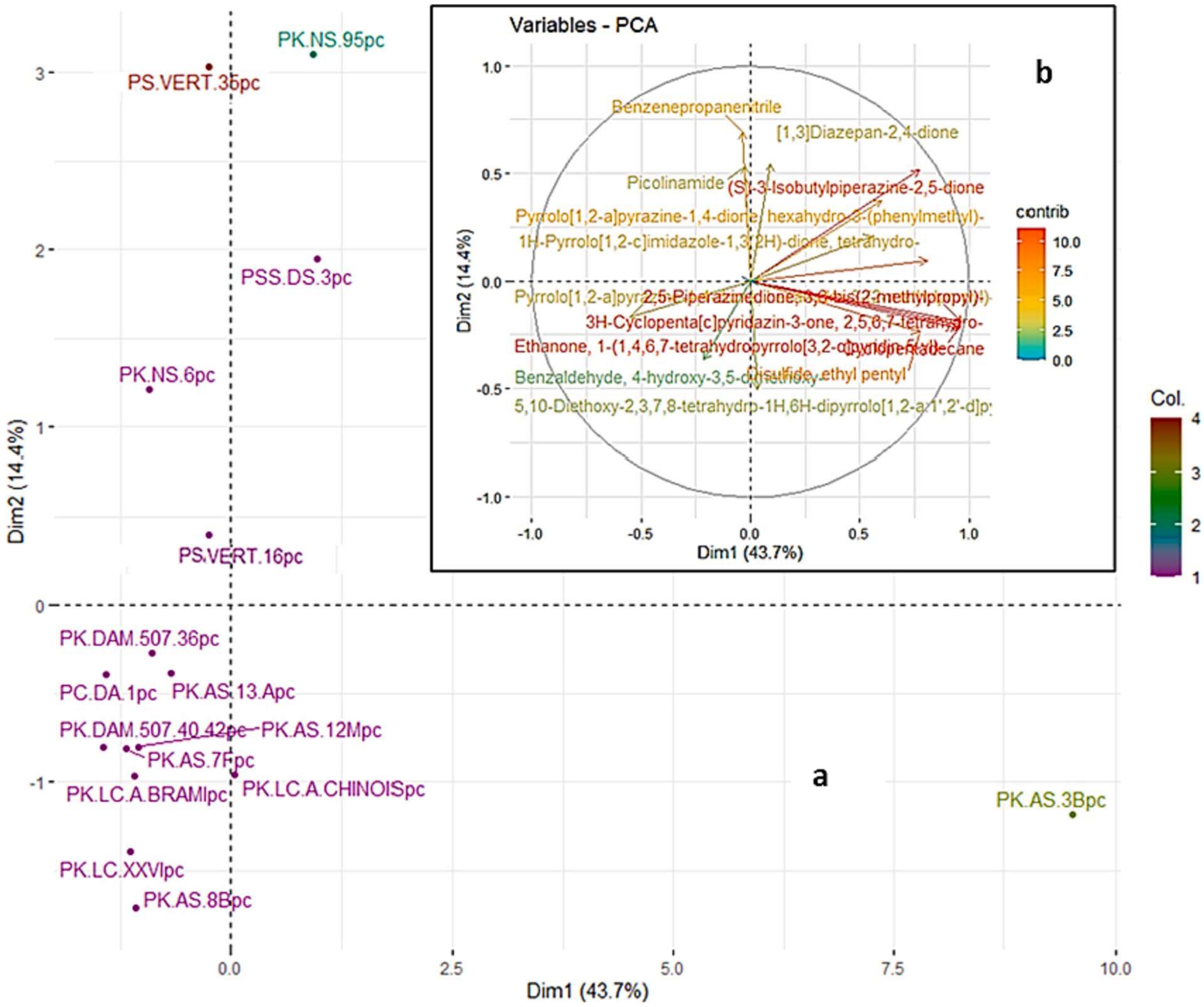

Fig. 7. A) first principal component (describing 58% of total variance) from pca of the analysed inked manuscript samples based on the distributions of compounds characteristic of proteinaceous materials. the colour has been attributed according to the four clusters of the dendrogram. b) Graphical output of the variables' contributions for the first principal component with components in red and orange contributing the most to the definition of the first principal component. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

presented as dissimilar from the majority of our manuscripts grouped with our traditional ink made with pine soot into clusters 3 and 4 coming from the same branch. PC DA 1 is isolated from the rest of our samples in cluster 5.

To better interpret our data, beside the partitioning to clusters according to the closeness of their PAHs distribution, PCAs were also used to reveal the influence of the individual components (or groups of components) on the samples' similarity and variation.

In the first principal component (dimensions 1 and 2), explaining 35 % of the data variance (Fig. 3), modern traditional inks made from oil soot regroup and are defined by a group of variables, mainly Benzo[b] fluoranthene, Indeno[1,2,3-cd]pyrene and Benzo[ghi]perylene which are heavy PAHs 5 and 6 cycles, while the modern traditional ink made with pine soot are defined by lighter PAHs and methylated PAHs (Fig. 3). This is in accordance with previous work on soot characterization which have identified methylated and cyclopenta-fused PAHs as markers of wood soot, and light PAHs as markers of softwood (such as pine) versus heavier PAHs as markers of hardwood (from angiosperms or flowering plants) (Guillon et al., 2013; Achten et al., 2015; Orasche et al., 2013; Avagyan et al., 2016; Rogge et al., 1998; Gonçalves et al., 2011; Simoneit et al., 2000).

The PAHs distribution of the 16 samples of manuscripts is dissimilar from our samples of modern traditional inks made from oil soot but presents high similarities with our sample of modern traditional ink made from pine soot. This seems to indicate that the majority of the manuscripts were made with pine soot, further implied by the fact that retene was detected in the majority of the manuscripts. Indeed, softwood species are prolific resin producers, resulting in emissions of retene if burnt (Guillon et al., 2013; Achten et al., 2015; Orasche et al., 2013; Avagyan et al., 2016; Rogge et al., 1998; Gonçalves et al., 2011; Simoneit et al., 2000). Retene was not detected solely in manuscript PC DA 1. The ink of PC DA 1 is defined mainly by a high percentage of phenanthrene and anthracene and few PAHs detected. All of this could indicate that the ink of PC DA 1 was made with a different type of soot. Indeed, the ink for the manuscript PC DA 1 appeared far away from the rest of the other samples in the first principal component (Fig. 3). However, it is important to note that PC DA 1 is the verso of a two sided document (Trombert, 2000) and thus could have had an impact on the PAHs distribution as it could be the mix of two inks from the verso but also the recto of the document.

Furthermore, two subgroups in green and marron in Fig. 3 were identified in the principal component analysis and in accordance with our classification (Fig. 2). The subgroup in marron (consisting of blackIP, PSS DS3, PS VERT 16, PK LC XXVI, PK DAM 507 40–42, PK LC A Brāmī and PK AS 3B) is characterized by a group of methylated and phenylated PAHs. The other in green (consisting of PS VERT 35, PK DAM

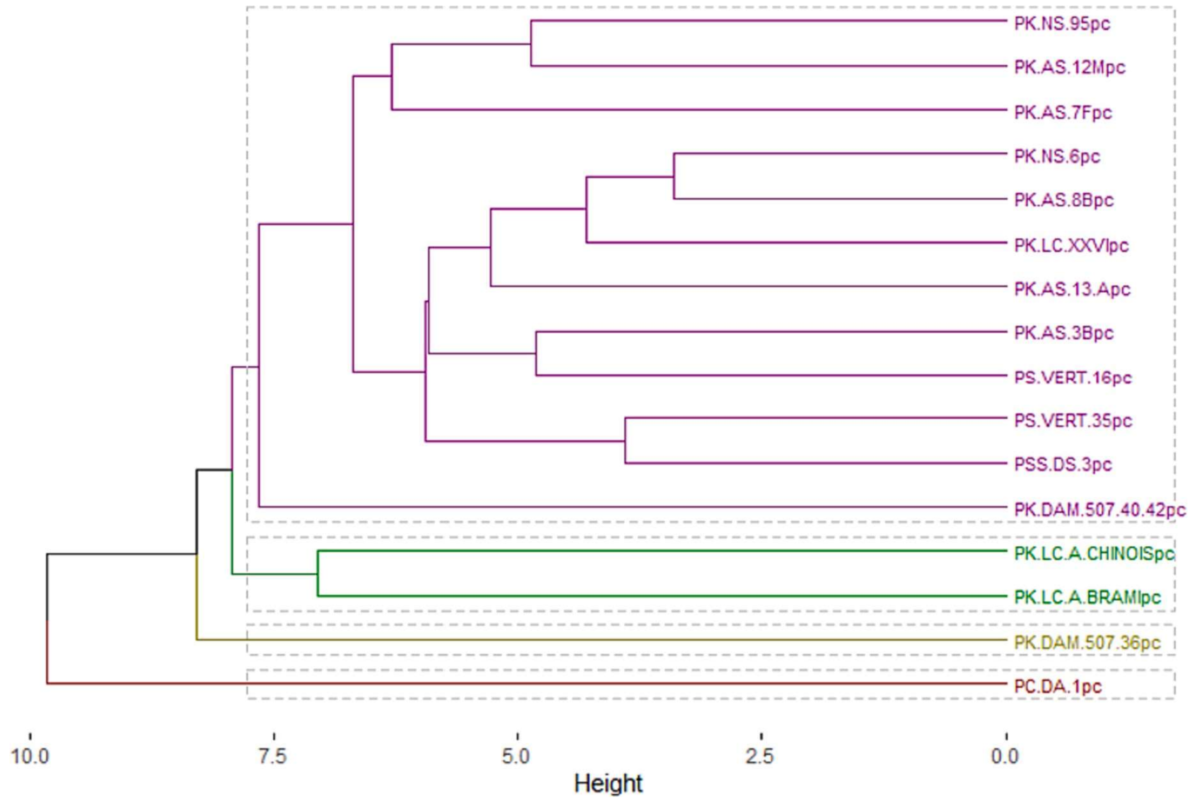

Fig. 8. Dendrogram of the hierarchical clustering of the analysed inked manuscript samples based on the distributions of compounds attributed to additives presence. Average linkage method, c = 0,89, k = 4.

507 36, PK LC A Chinese, PK AS 13A, PK NS 95, PK AS 7F, PK AS 12 M, PK AS 8B and PK NS 6) is characterized by a group of variables, light PAHs such as fluorene, fluoranthene and pyrene (Fig. 3). These differences in PAHs distribution could indicate a difference in soot combustion conditions used in ink production, as they have been shown to also impact PAHs distribution (Avagyan et al., 2016; Rogge et al., 1998; Gonçalves et al., 2011; Simoneit, 2002; Weimer et al., 2008; Pettersson et al., 2011; Boman et al., 2011; Shen et al., 1994; Kim Oanh et al., 1999).

3.2. Binder

Only a few of the spectra obtained by FTIR analysis of the selection of the inked manuscripts were interpretable (Table 1 and Fig. B24 in Supplementary material). Indeed, in reflectance, a bright material is generally needed to get a clear signal. And in fact, the inks that appeared shinier under light (Fig. 4) were the ones for which we had a clear FTIR signal (Fig. 5). Furthermore, degradation can also affect the FTIR signal. It is thus impossible to conclude on the nature of the binder used on the material not reflective enough or too degraded.

However, several of the FTIR results (Fig. 5) show the presence of protein with the identification of the characteristic amide bands, at 1630 cm^{-1}, assigned to C = O stretching vibration and, at 1530 cm^{-1}, associated to C–N stretching and N–H bending vibrations (Pellegrini et al., 2016; Belbachir et al., 2009). As mentioned, animal glue has been commonly used as the binding media in Chinese ink (Chi-chen, 1930; Needham and Tsuen-Hsuin, 1985), however, FTIR analysis did not provide direct insight on the precise animal origin of the protein binder used in the fabrication of inks.

Cluster analysis and principal component analysis were conducted on 26 characteristic compounds of protein substances detected by Py-GCxGC/MS based on the analysis of reference materials and similar studies on animal glues. The main markers of proteins are diketopiperazines (DKPs) formed upon pyrolysis of amino acid chains, as well as other compounds such as cyclopentadecane and picolinamide.

Py-GC–MS has been proven to be adequate to identify egg binder based on the presence of characteristic compounds but in the case of other proteinaceous substances their differentiation is not so obvious (Perruchini et al., 2022; Chiavari et al., 1998). Beside the technique's limits, other elements could explain the difficulty to identify precisely the origins of proteinaceous materials used in ink making. First, the unknown method of binder preparation in ancient recipes (which part of the animal and conditions of preparation were used) will have an impact on the final product and its composition. Moreover, we know from the literature that a mix of proteinaceous materials of different origins could have been used.

What the clustering analysis puts into light is that the proteinaceous material used in PK AS 3B appears to be different from the rest of the manuscripts (Fig. 6). The distinction of the proteinaceous material from the rest of the manuscripts is more challenging, but PK NS 95 and PS VERT 35 while presenting similarities are also presenting differences that put them apart from the rest of the main cluster constituted of the rest of the manuscripts (Fig. 6).

The principal component analysis gives us important information with the first principal component explaining 58 % of the data variance (Fig. 7). PK AS 3B is characterized by the high contribution of cyclopentadecane and a group of three DKPs as visualised in the first principal component. PK NS 95 can be distinguished by a high contribution of the DKP [1,3]diazepan-2,4-dione in the first principal component, while PS VERT 35 is characterised by a high contribution of picolinamide. In the second principal component, it is interesting to note that PK NS 95, PS VERT 35, PSS DS 3, PK NS 6 and PS VERT 16 all present a higher proportion of benzepropanenitrile (Fig. 7 and Fig. B24 in Supplementary material).

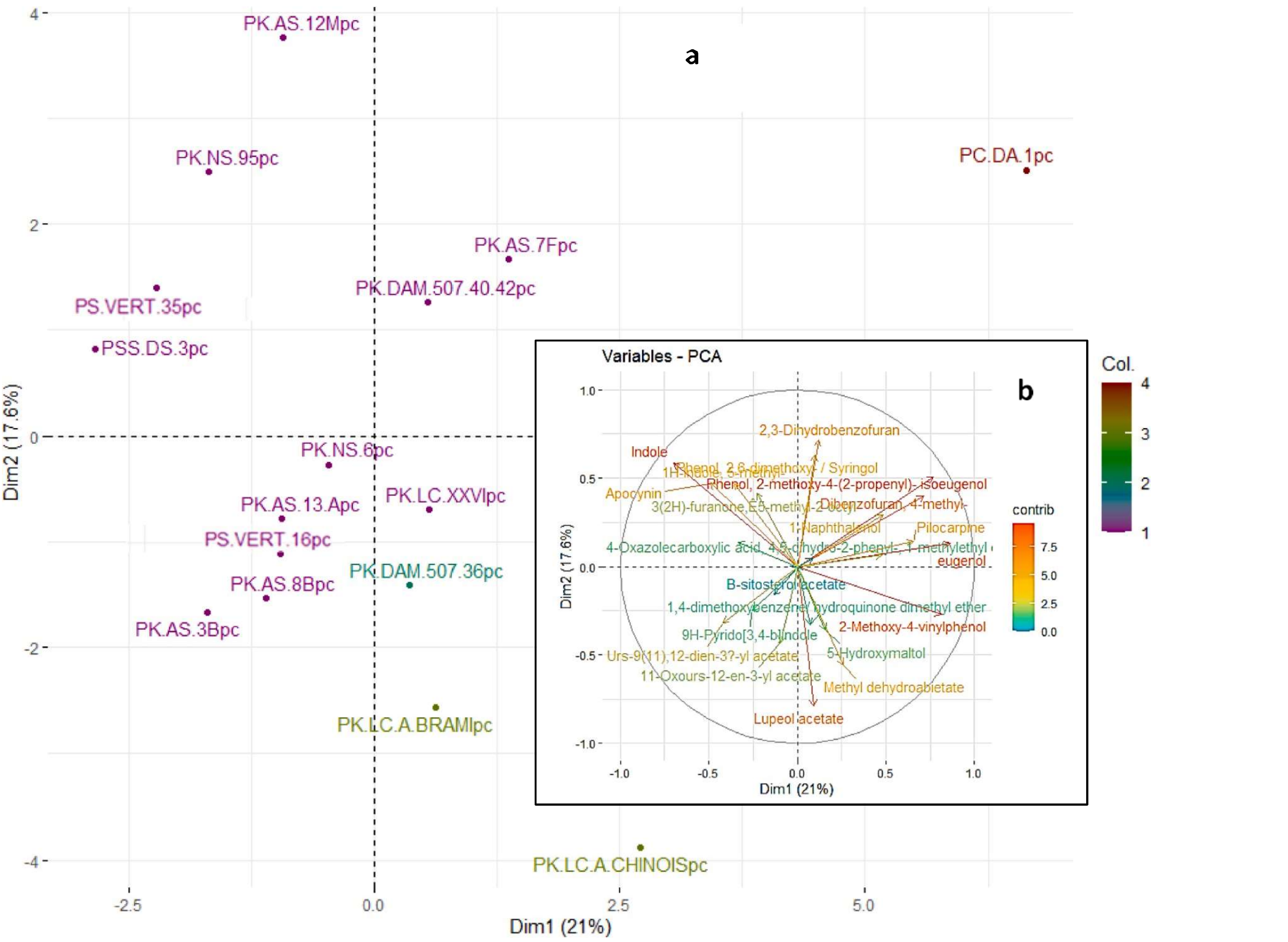

Fig. 9. A) the first principal component (describing 38% of total variance) from pca of the analysed inked manuscript samples based on the distributions of compounds attributed to additives presence. the colour has been attributed according to the four clusters of the dendrogram. b) Graphical output of the variables' contributions for the first principal component with compounds in red and orange contributing the most to the definition of the first principal component. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Other additives

FTIR and Raman analysis were not able to give us information on possible additives used in the fabrication of the inks applied to manuscripts.

Regarding the additives, our classification reveals that diverse additives would have been used in the fabrication of inks. In the study, four clusters were identified. One consists in PC DA 1 distinct from the rest of the manuscripts. The variation for the rest of the manuscripts is more ambiguous but PK DAM 507 36 as well as PK LC A CHINOIS and PK LC A BRĀMĪ seem to distinguish themselves from the rest (Figs. 8 and 9).

On the basis of the compounds identified and the available historical data, we were able to identify several additives. First, indole and 1H-indole,5-methyl, two compounds attributed in previous works to the presence of eggs white and/or egg yolk (Wei et al., 2012; Perruchini et al., 2022; Yao et al., 2021) were detected in the majority of the samples apart from PCDA 1 and PK LC A Chinois. PC DA1 is characterized by a high presence of phenylpropenes: isoeugenol and eugenol which are produced by plants, are found in abundance in cloves mentioned in Chinese ink recipes as additives serving as preservatives for the inkstick at least from the 6th century CE (Needham and Tsuen-Hsuin, 1985). However, we cannot discard the fact that isoeugenol and eugenol may also be found in other plants. The identification of phenolic compounds, dibenzofuran, 4-methyl- and pyroglutamic acid derivatives may prove the presence of pomegranate peels in the majority of our manuscripts from the 7th and 8th century. Pomegranate peels were used in the manufacture of Chinese ink during the Tang Dynasty (618–907) (Needham and Tsuen-Hsuin, 1985). These compounds were not identified in PK AS 12 M dated to the 6th century, as well as in the three manuscripts in Sanskrit and PK AS 3B from Dunhuang. Traces of pine resin (presence of retene, methyl dehydroabietate, and 7-Oxodehydroabietic acid methyl ester) (Wei et al., 2012; Brettell et al., 2015; Chiavari et al., 1995) were identified solely in six manuscripts, all from Douldour Aqour in the Kucha Region (PK AS 7F, PK AS 12 M, PC DA 1, PK LC A and PK AS 13A). Finally, other compounds have been identified but are more ambiguous as they can be found in multiple natural products. For instance, 5-Hydroxymaltol identified solely in PK DAM 507 36, is a derivative of maltol, is a substance that can be found in *Penicillium echinulatum* (Anderson et al., 1988) as well as diverse plants and fruits (Demirci et al., 2018; Kim et al., 2021). Contrary to the analysis of soot markers and proteinaceous binders, the analysis of other additives was revealed to be more delicate for the interpretation of data since a complete set of reference samples is missing and requires extensive studies of historical texts for exhaustive studies if alike available. However, thanks to mass spectrometric identifications, the distinction between groups of manuscripts was made possible on the basis of similarities in the addition of those specific additives.

4. Conclusions

The inks of 16 manuscripts from first millennium manuscripts discovered in the Kucha and Dunhuang regions (West China) from the “Pelliot collection” stored at the BnF (Paris, France) were characterised by FTIR, Raman and Py-GCxGC/MS analyses.

This study allows for the first time the identification of characteristic signatures of ingredients and production techniques of ancient inks from manuscripts of the Pelliot collection, even on small quantities of inked manuscript samples.

While Raman and FTIR have the advantage of being non-invasive and non-destructive techniques that can in some cases confirm the presence of certain products of interest, they do not provide sufficiently precise information to study the precise composition of the inks applied to manuscripts. They are nevertheless essential to justify a sample and to allow a micro-invasive and micro-destructive analysis like Py-GCxGC/MS which can provide much more information on the identification of the compounds present.

Based on the relative abundances of the main PAHs and through comparison with modern inksticks, the ink of the sixteen manuscripts was identified as pine-soot ink. Furthermore, and while we cannot pinpoint precisely which factor(s) have impacted the molecular profile identified by Py-GCxGC/MS (species of pine wood, geographical origin of the wood, combustion parameters, etc.), our results reveal that different pine soot may have been produced and used in the fabrication of the ink. It would be interesting to find out if it is also possible to study how soot production, in particular each combustion parameters, can impact PAH signatures of inks. To do so, in future studies, it would be worth analyzing reference inks made with different soot production conditions.

While the identification of the proteinaceous binder (or binders) used in ink production from the manuscripts appears to be beyond the reach of this study, this method could be applied to explore differences and similarities in proteinaceous binder(s) uses in ancient inks fabrication. Our clustering analysis puts into light, that apart from PK AS 3B, the proteinaceous binder(s) used in the inks of the 15 other manuscripts shared strong similarities.

Regarding the additives, our results suggest that a diverse range of additives would have been used in the fabrication of inks. While their sure identification cannot be achieved without the analysis of a large set of reference samples, by comparing molecular profiles, our study, at least to some degree, explores similarities and differences in additives used in the fabrication of inks.

Although the present study cannot, without reference samples, precisely characterize all the ingredients which were used in the fabrication of inks from first millennium manuscripts, we demonstrated that this multi-analytical method is effective to study ancient inks, present even at trace level, and bring to light similarities and differences in their fabrication. The method is therefore well suited for the analysis of aged manuscripts and questioned documents, as well as other rare and precious samples. With the information collected with Py-GCxGC/MS, it would be possible to explore if the observed differences are in direct correlation with an evolution of manufacture of inks over time (temporal), or link to regional traditions (spatial) or dependent of other specificities of the manuscripts such as the type of documents (religious or secular), the script, or even the writing surface (based on a previous microscopic study of paper fibers of the manuscripts of the Pelliot collection). These analytical results will consequently be of great interest for further philological investigations.

Funding

This publication belongs to the research conducted under the HisTochText project. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 788205).

Acknowledgements

For giving us access to the Pelliot Collection and the authorization to sample manuscripts, the authorization and support of Dr. Isabelle Le Masne de Chermont, General Curator, Director of the Manuscripts Department of the Bibliothèque nationale de France (BnF) is gratefully acknowledged.

We have also benefited from the efficient support of Dr. Laurent Héricher, General Curator, Head of the Oriental division of the Manuscripts Department of the BnF. We would like to express our gratitude to Dr. Jérôme Petit, Curator of the Indian Manuscripts at the BnF for sharing invaluable information on the typology and history of the manuscripts of the Pelliot Collection. In addition, we are indebted to Athanaric Huard and Emilie Arnaud-Nguyen, PhD students at the EPHE, PSL, and team members working on the HisTochText project, for their precious help at various stages of our investigations on the original manuscripts.

We thank Dr. Ludovic Bellot-Gurlet and Dr. Céline Paris for their assistance in the data acquisition with their Raman spectrometers from MONARIS laboratory (UMR 8233, Sorbonne Université).

We would like to express our gratitude to Dr. Narimane Dorey for their help on how to improve the clarity and style of the different graphs presented in this paper.

References

Abdul Razak, R., Barkeshli, M., Othman, R., 2017. Traditional malay black ink: an analysis of its recipes and characteristics. Adv. Sci. Lett. 23, 6189–6193. https://doi.org/10.1166/asl.2017.9233.

Achten, C., Beer, F.T., Stader, C., Brinkhaus, S.G., 2015. Wood-specific polycyclic aromatic hydrocarbon (PAH) patterns in soot using gas chromatography-atmospheric pressure laser ionization-mass spectrometry (GC-APLI-MS). Environ. Forensic 16, 42–50. https://doi.org/10.1080/15275922.2014.991004.

Anderson, H.A., Bracewell, J.M., Fraser, A.R., Jones, D., Robertson, G.W., Russell, J.D., 1988. 5-Hydroxymaltol and mycophenolic acid, secondary metabolites from Penicillium echinulatum. Trans. Br. Mycol. Soc. 91, 649–651. https://doi.org/ 10.1016/S0007-1536(88)80040-8.

Avagyan, R., Nystro¨m, R., Lindgren, R., Boman, C., Westerholm, R., 2016. Particulate hydroxy-PAH emissions from a residential wood log stove using different fuels and burning conditions. Atmos. Environ. 140, 1–9. https://doi.org/10.1016/j.atmosenv.2016.05.041.

Belbachir, K., Noreen, R., Gouspillou, G., Petibois, C., 2009. Collagen types analysis and differentiation by FTIR spectroscopy. Anal. Bioanal. Chem. 395, 829–837. https:// doi.org/10.1007/s00216-009-3019-y.

Boman, C., Pettersson, E., Westerholm, R., Bostro¨m, D., Nordin, A., 2011. Stove performance and emission characteristics in residential wood log and pellet combustion, part 1: pellet stoves. Energy Fuels. 25, 307–314. https://doi.org/10.1021/ef100774x.

Brettell, R.C., Schotsmans, E.M.J., Walton Rogers, P., Reifarth, N., Redfern, R.C., Stern, B., Heron, C.P., 2015. 'Choicest unguents': molecular evidence for the use of resinous plant exudates in late Roman mortuary rites in Britain. J. Archaeol. Sci. 53, 639–648. https://doi.org/10.1016/J.JAS.2014.11.006.

Chiavari, G., Fabbri, D., Rocco, M., Bocchini, P., Galletti, G., 1995. Pyrolysis gas chromatography-mass spectrometry of natural resin used for artistic objects. Chromatographia 41, 273–281. https://doi.org/10.1007/BF02688040.

Chiavari, G., Gandini, N., Russo, P., Fabbri, D., 1998. Characterisation of standard tempera painting layers containing proteinaceous binders by pyrolysis (/methylation)-gas chromatography-mass spectrometry. Chromatographia 47, 420–426. https://doi.org/10.1007/BF02466473.

Chi-chen, W., 1930. Notes on Chinese Ink. Metrop. Mus. Stud. 3, 114–133. https://doi.org/10.2307/1522772.

Demirci, M.A., Ipek, Y., Gul, F., Ozen, T., Demirtas, I., 2018. Extraction, isolation of heat-resistance phenolic compounds, antioxidant properties, characterization and purification of 5-hydroxymaltol from Turkish apple pulps. Food Chem. 269, 111–117. https://doi.org/10.1016/j.foodchem.2018.06.147.

Franke, H., 1962. Kulturgeschichtliches über die chinesische Tusche. Bayerische Akademie der Wissenschaften, München.

Gonçalves, C., Alves, C., Fernandes, A.P., Monteiro, C., Tarelho, L., Evtyugina, M., Pio, C., 2011. Organic compounds in PM2.5 emitted from fireplace and woodstove combustion of typical Portuguese wood species. Atmos. Environ. 45, 4533–4545. https://doi.org/10.1016/j.atmosenv.2011.05.071.

Guillon, A., Le Menach, K., Flaud, P.-M., Marchand, N., Budzinski, H., Villenave, E., 2013. Chemical characterization and stable carbon isotopic composition of particulate polycyclic aromatic hydrocarbons issued from combustion of 10 mediterranean woods. Atmos. Chem. Phys. 13, 2703–2719. https://doi.org/10.5194/acp-13-2703-2013.

Han, B., Vial, J., Sakamoto, S., Sablier, M., 2019. Identification of traditional East Asian handmade papers through the multivariate data analysis of pyrolysis-GC/MS data. Analyst 144, 1230–1244. https://doi.org/10.1039/C8AN01898G.

Kassambara, A., 2017. Practical guide to cluster analysis in r: unsupervised machine learning, 1st ed.,. STHDA.

Kim, S.-L., Choi, H.S., Ko, Y.-C., Yun, B.-S., Lee, D.-S., 2021. 5-hydroxymaltol derived from beetroot juice through lactobacillus fermentation suppresses inflammatory effect and oxidant stress via regulating NF-kB, MAPKs Pathway and NRF2/HO-1 expression. Antioxidants. 10, 1324. https://doi.org/10.3390/antiox10081324.

Kim Oanh, N.T., Bætz Reutergårdh, L., Dung, N.T., 1999. Emission of polycyclic aromatic hydrocarbons and particulate matter from domestic combustion of selected fuels. Environ. Sci. Technol. 33, 2703–2709. https://doi.org/10.1021/es980853f.

Luo, Y., Chen, J., Yang, C., Huang, Y., 2019. Analyzing ancient Chinese handmade Lajian paper exhibiting an orange-red color. Heritage Science. 7, 61. https://doi.org/10.1186/s40494-019-0306-6.

Needham, J., Tsuen-Hsuin, T., 1985. Science and Civilisation in China: Volume 5, Chemistry and Chemical Technology, Part 1. Paper and Printing. Cambridge University Press.

J. Neelis, Chapter Four. Old Roads In The Northwestern Borderlands, in: Brill, Leiden, The Netherlands, 2011: pp. 229–256. https://doi.org/10.1163/ej.9789004181595.i-372.27.

Orasche, J., Schnelle-Kreis, J., Scho¨n, C., Hartmann, H., Ruppert, H., Arteaga-Salas, J., Zimmermann, R., 2013. Comparison of emissions from wood combustion. part 2: impact of combustion conditions on emission factors and characteristics of particle-bound organic species and polycyclic aromatic hydrocarbon (PAH)-related toxicological potential. Energy Fuels 27, 1482–1491. https://doi.org/10.1021/ef301506h.

Pellegrini, D., Duce, C., Bonaduce, I., Biagi, S., Ghezzi, L., Colombini, M.P., Tin`e, M.R., Bramanti, E., 2016. Fourier transform infrared spectroscopic study of rabbit glue/ inorganic pigments mixtures in fresh and aged reference paint reconstructions. Microchem. J. 124, 31–35. https://doi.org/10.1016/j.microc.2015.07.018.

Perruchini, E., Pinault, G.-J., Sablier, M., 2022. Exploring the potential of pyrolysis-comprehensive two-dimensional gas chromatography/mass spectrometry in the characterization of Chinese inks of ancient manuscripts. J. Anal. Appl. Pyrol. 164, 105503 https://doi.org/10.1016/j.jaap.2022.105503.

Pettersson, E., Boman, C., Westerholm, R., Bostro¨m, D., Nordin, A., 2011. Stove performance and emission characteristics in residential wood log and pellet combustion, part 2: wood stove. Energy Fuels. 25, 315–323. https://doi.org/10.1021/ef1007787.

Ren, M., Wang, R., Yang, Y., 2018. Identification of the proto-inkstone by organic residue analysis: a case study from the Changle Cemetery in China. Heritage Science. 6, 19. https://doi.org/10.1186/s40494-018-0184-3.

Ren, M., Guan, L., Wang, N., Xu, C., Cui, Y., He, D., Wang, C., Yang, Y., 2022. Investigation of the manufacture development of early Chinese ink in the Western Han dynasty. Archaeometry 64, 1213–1225. https://doi.org/10.1111/arcm.12763.

Rogge, W.F., Hildemann, L.M., Mazurek, M.A., Cass, G.R., 1998. Sources of fine organic aerosol. 9. Pine, Oak, and synthetic log combustion in residential fireplaces. Environ. Sci. Technol. 32, 13–22. https://doi.org/10.1021/es960930b.

Shen, G., Wang, W., Yang, Y., Zhu, C., Min, Y., Xue, M., Ding, J., Li, W., Wang, B., Shen, H., Wang, R., Wang, X., Tao, S., 1994. Emission factors and particulate matter size distribution of polycyclic aromatic hydrocarbons from residential coal combustions in rural Northern China. Atmos Environ 44 (2010), 5737–5743.

Simoneit, B.R.T., 2002. Biomass burning — a review of organic tracers for smoke from incomplete combustion. Appl. Geochem. 17, 129–162. https://doi.org/10.1016/S0883-2927(01)00061-0.

Simoneit, B., Rogge, W., Lang, Q.-Y., Jaffe, R., 2000. Molecular characterization of smoke from campfire burning of pine wood (Pinus elliottii). Chemosphere Global Change Sci. 2, 107–122. https://doi.org/10.1016/S1465-9972(99)00048-3.

E´. Trombert, Les manuscrits chinois de Koutcha : fonds Pelliot de la Biblioth`eque nationale de France, Institut des hautes ´etudes chinoises du Coll`ege de France, Paris, France, 2000. https://gallica.bnf.fr/ark:/12148/bpt6k255699b (accessed December 22, 2021).

Walter, M.N., 1998. Tokharian Buddhism in Kucha: Buddhism of Indo-European Speakers in Chinese Turkestan before the 10th century CE. University of Pennsylvania, Philadelphia.

Wei, S., Fang, X., Yang, J., Cao, X., Pintus, V., Schreiner, M., Song, G., 2012. Identification of the materials used in an Eastern Jin Chinese ink stick. J. Cult. Herit. 13, 448–452. https://doi.org/10.1016/j.culher.2012.02.001.

Weimer, S., Alfarra, M.R., Schreiber, D., Mohr, M., Pr´evo^t, A.S.H., Baltensperger, U., 2008. Organic aerosol mass spectral signatures from wood-burning emissions: Influence of burning conditions and wood type. J. Geophys. Res. Atmos. 113 https://doi.org/10.1029/2007JD009309.

Yao, N., Zhan, X., Ma, Q., Wei, S., 2021. Characterization and identification of chinese historical rubbings preserved in wuyuan museum by pyrolysis-gas chromatography/mass spectrometry. Heritage Science. 9, 116. https://doi.org/10.1186/s40494-021-00596-z