

RESEARCH

Open Access



Multivariate statistical analysis on Japanese traditional papers, *washi*, for evaluation of DNA presence and quality

Minsoo Shin¹ and Toshiharu Enomae^{2*}

Abstract

Traditional papermaking was a significant innovation that intrigued the information-sharing society in a wide range of the globe. The importance of finding the historical background hidden behind the historical traditional paper is critical to understanding societies in history. Advanced molecular biological research is applied to many forensic studies to identify the origin and criminal events. Yet, biological research refrained from the heritage science of traditional papers due to its damage to the samples. The current study explores the Pearson correlation between DNA in traditional paper and non-destructive measurement of characteristics of traditional paper. DNA concentration and purity were found to have a proportional relationship based on the CIEL*a*b* color system. The color of the traditional paper was assessed assessment shows green and yellow rather than red and blue. The DNA concentration demonstrated a positive linear correlation with the peaks of 992 cm⁻¹, 1599 cm⁻¹, and 1628 to 1638 cm⁻¹ region, which were lignin-specific functional groups. The presence of lignin in traditional paper defines the colors green and yellow, which is related to the presence of a cell wall. The quantity of cell walls present is linked to DNA survival during the traditional papermaking process. Furthermore, the partial least square (PLS) regression model is established. The prediction model based on DNA, paper properties, and smoothing + the second derivative preprocessed attenuated total reflection-Fourier-transform infrared (ATR-FTIR) spectral data demonstrated valid purity predictions. The evaluation parameter of the cross-validated prediction model contains a low Root Mean Square Error of Cross-Validation (RMSECV) of 0.091 for DNA purity. The determination coefficient was close to 1. However, the further establishment of a validated DNA extraction methodology is required for the consistency of the prediction model for DNA concentration. The hierarchical clustering analysis demonstrated the agglomerative dendrogram on traditional paper samples. For the dendrogram, the Nasu paper mulberry (originating in northern Japan) was distinguished from another paper mulberry (originating in southwestern Japan) by clustering. By multivariate statistical analysis, the DNA presence and quality were predictable from the Japanese traditional papers. In addition, the clustering analysis provided the possible origin of paper mulberry materials for traditional papers. Based on the preliminary information, the need for DNA analysis on historical traditional paper can be determined.

Keywords DNA, Traditional papermaking, Paper mulberry, Multivariate analysis

Introduction

Background of traditional papers

In a historical context, the values of traditional papers are extremely important. Written records on traditional papers serve as a link between the past and the future. Many studies attribute the origins of old papermaking technology to Cai Lun and China in the first century AD

*Correspondence:

Toshiharu Enomae

enomae.toshiharu.fw@u.tsukuba.ac.jp

¹ Graduate School of Science and Technology, University of Tsukuba, Tsukuba, Japan

² Faculty of Life and Environmental Sciences, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8572, Japan



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

[1]. However, the form of paper existed before Cai Lun developed his own papermaking methods. The *Baquiao* is exposed to the first traditional plant tissue paper, and the Western Han Dynasty's *Baquiao* paper was found in 1957 [1, 2]. Cai Lun is recognized for the dissemination of developed papermaking techniques rather than being the inventor of paper [3]. Even with Cai Lun's paper propagation efforts, the history of traditional paper propagation is vague. The Korean papermaking has been introduced as a modified technique from Chinese papermaking technique. Yet, the written accounts in historical documents remain unclear on how papermaking spread from China to Korea [4]. It is assumed that traditional Chinese papermaking impacted Tibetan and Central Asian papermaking. In Dunhuang, a manuscript of Tibetan records was found. It contains details about invitation notes sent to Chinese imperial princesses from Tibet. This symbol dates back to the tenth century [5, 6]. Papermaking was introduced into Central Asia by the Tang empire between the seventh and eighth centuries, according to Islamicate writers of the ninth century. However, there is a possible historical record that predates the seventh century [7]. According to *Nihon Shoki* (the Chronicle of Japan), a Korean monk named *Damjing* (also known as *Doncho*) brought traditional papermaking to Japan in the seventh century [8]. However, it is debatable because Japanese papermaking techniques already existed in many crafts from the third to the fifth centuries [9]. Paper materials and traditional papers were innovative tools in society. Papermaking technology and the materials it uses had a significant impact in East Asian countries comprising Korea, China, Japan, and Thailand. However, imports of paper mulberry have increased due to the rising cost of paper mulberry cultivation and water usage during traditional papermaking. While historical records verify these imports, scientific evidence remains limited. The hybrid species of paper mulberry, in particular, are not fully understood in the field of plant taxonomy. The availability of genetic information in the traditional paper may prove helpful in the unraveling of mysteries regarding not only the historical propagation of traditional papermaking technology but also the biodiversity of paper mulberry species [10]. Identifying fibers in traditional papermaking has been difficult, especially because paper mulberry with an age less than a year old is used for traditional papermaking. Until recently, the main method for identifying archaeological fibers was microscopy, which could result in distorted interpretations [11]. Also, the concepts of traditional papermaking influenced a wide geographical range. Thus, the historical context behind traditional papermaking is difficult to define with only a written record. Further

scientific approaches to traditional papers are required to provide scientific evidence for the historical context of papermaking propagation in history.

DNA presence of traditional paper

One of the most important aspects of traditional papermaking is the mildness of the procedures. The pulping process in modern papermaking involves mechanical, thermal, or chemical procedures. The wood materials are treated with intense heat, pressure, and strong alkali cooking liquors (kraft or sulfite processing) during the process [12]. In traditional papermaking, the bast fiber substance is treated with reduced heat, pressure, and a weaker alkaline reagent (wood ash or lime) [8, 13, 14]. The mild conditions of traditional papermaking cause less degradation in natural structure of fiber materials. As a consequence, paper used for some traditional documents was found to contain parenchyma tissue-like structures between the fibers [15]. Furthermore, *Abelmoschus Manihot* mucilage is added as dispersion aids during paper production [16]. Individual bast fibers are surrounded by the mucilage components during sheet forming. The original fiber structures and constituents within the fibers are distinctively preserved throughout the papermaking processes until the end of traditional paper production. The previous biodiversity research indicates that, in addition to the high possibility of remaining bast fiber DNA, analysis of traditional papers is required for determining the historical importance of understanding paper material and histories [17]. Regardless of species, the chemical specifications of the paper mulberry vary, including fiber width and length. Because of the surrounding environment, the paper mulberry will have different characteristics that influence traditional paper. Specific characteristics or componential differences can be used as element markers to determine the origin of source materials. However, with the remaining DNA from traditional paper, the precise location and timeline of paper mulberry species will be estimated with more precision.

Scientific research on traditional paper

Due to the historical value of traditional papers, a micro-/non-destructive approach to measurement is important in the heritage sciences. To assess the condition of traditional paper samples, qualitative readings were taken. Previous research examined the properties of traditional paper. Optical lenses and computation tools were used to examine the optical densities and colors [18]. The computed properties are saved in the digital collection for future reference. Mossbauer spectroscopy was used to evaluate the risk of deterioration due to hydrolysis and oxidation [19]. UV-Vis Spectrophotometer, Scanning

Electron Microscopy (SEM), X-ray fluorescence (XRF), Near-Infrared (NIR), Fourier-transform infrared (FTIR), X-ray Photoelectron (XPS), and Raman Spectroscopy were used to determine the chemical and structural properties [20–23]. For example, the fiber characterization in ancient papers was developed by Pyrolysis-Gas Chromatography (Py-GC) and GC/Mass Spectroscopy for analyzing the lignin monomer of lignocellulose, organic and inorganic materials [11, 24–27]. Those identified components and compounds were utilized as chemical markers for assessing historical document source materials and its potential origin.

FTIR spectroscopy is an important method for the identification of the material. Especially in heritage sciences, traditional paper samples are applied for the material composition. Based on the past research, the spectral data were evaluated for their functional groups and properties. The attenuated total reflection instrument (ATR-FTIR) is a key measurement in this research among many spectroscopic methods. The infrared beam is reflected and recorded without causing any harm to solid samples. To determine the component of a sample, the functional groups and chemical structures of the sample are detected. The FTIR spectra are preprocessed to remove noise from the sample. Wandering data, uneven datasets, and light scattering during measurement are examples of potential noises. The most important preprocessing for multivariate statistics is chosen from among several methods for analysis and visualization. Multivariate statistical analysis is a mathematical evaluation used to identify and forecast the relationship between numerous variables. Many previous studies gave multivariate statistics in addition to preprocessed spectroscopic data. To evaluate lignocellulosic components in wood species, historical papers, and bark-cloth, the correlation and partial least square (PLS) methods are explained in previous research [20, 22, 28–33]. The specific element marker is evaluated with inductively coupled plasma-atomic emission spectroscopy/mass spectrometry (ICP-AES/MS) and PLS-discriminant analysis for the metal and rare-earth elements remaining in traditional paper in Korea [10]. The Pearson correlation was used in this research to determine a linear relationship between spectra and other properties of paper and DNA concentration and purity. A coefficient ranging from -1 to 1 indicates negative to positive linear relationships. Furthermore, the PLS regression model is created as a prediction model using spectra and paper property data. In this research, the PLS regression model on ATR-FTIR spectra and paper properties is fitted to the DNA concentration and purity prediction model. The Hierarchical Cluster Analysis (HCA) is an unsupervised clustering learning technique for identifying similar clusters of observations. The HCA

is created by combining the collection of FTIR spectra with squared Euclidean distance. For wood identification study, the classification of wood was depicted as a dendrogram [34]. Similarly, a dendrogram will be presented for the cluster of traditional papers. The data obtained through the non-destructive method is interpreted as a potential guideline for handling unknown samples using multivariate statistics. The prediction model will be used as a guide, especially as a traditional paper with a high historical value.

Traditional paper is highly likely to retain DNA within the fibers [35–38]. The past background and propagation of a paper material will be explained using genetic analysis of DNA extracted from traditional papers. Depending on the precision of genetic markers, the DNA would navigate to the origin of traditional paper component. However, biological techniques are avoided in heritage science due to its possible damages to the historical traditional paper. By understanding the molecular biological approach to historical documents, it is essential to have a non-destructive method for screening the possibility of remaining DNA. Therefore, this study evaluated the concentration and purity of DNA using the non-destructive method of ATR-FTIR spectra and the color of traditional paper. Therefore, multivariate statistical analysis was applied to the data to demonstrate its applicability in determining the possibility of DNA presence and distinguishing between traditional paper sources.

Materials and methods

Materials

The *Broussonetia* tree (Paper mulberry) was collected and purchased from Daigo, Ibaraki, Japan. The tree samples were stored at -20 °C for storage before lab-scaled traditional papermaking. Commercial Traditional paper samples (NW1–7) were purchased from *Wagamido*, a Japanese paper shop, Saitama, Japan. Traditional paper samples (X1–3, K–1–3, K+1–3) were provided by Ms. Mashiko Endoh, Shiroishi washi studio. Other traditional paper samples (Ue1, Ue2, Nan1) were collected from the personal collection of Mr. Takashima Akihiko, a technical expert from Historiographical Institute, the University of Tokyo, Tokyo, Japan. All basic information of traditional paper samples was confirmed from the source of samples. The list of basic information of the traditional paper is shown in Table 1.

Lab-scale traditional papermaking

The traditional papermaking was modified into lab-scale traditional papermaking based on the procedure of traditional papermaking techniques [8]. The tree branch sample was steamed at 100 °C for 3–5 h. The soft bark was separated from the inner tree stem. The outer black bark

Table 1 List of traditional papers used in current research

Traditional Paper Name	Source material	<i>Broussonetia</i> Origin	Year of Production	Referring Number	Manufacturer
Uejima-paper			1997–2000	Ue1	
Uejima-paper			1997–2000	Ue2	
Yukyu-Paper	<i>Broussonetia X</i>	<i>Nanto, Toyama</i>	≤ 2017	Nan1	<i>Nanto, Toyama</i>
Kozo-Paper			Meiji Era (1868–1912)	X1	
Kozo-Paper			Edo Era (1603–1868)	X2	
Kozo-Paper			Edo Era (1603–1868)	X3	
Mino-Paper	<i>Broussonetia X</i>	<i>Nasu</i>	≤ 2002	K-1	<i>Mino, Gifu</i>
Sekishu-Paper	<i>Broussonetia X</i>	<i>Shimane</i>	≤ 2002	K-2	<i>Masuda, Shimane</i>
Takano-Paper	<i>Broussonetia X</i>	<i>Hosokawa, Wakayama</i>	≤ 2002	K-3	<i>Otsu, Shiga</i>
Misu-Paper	<i>Broussonetia X + Additive</i>	<i>Kochi</i>	≤ 2002	K+1	<i>Yamato, Nara</i>
Uda-Paper	<i>Broussonetia X + Additive</i>	<i>Kochi</i>	≤ 2002	K+2	<i>Yoshino, Nara</i>
Ho-Letter (High-graded)	<i>Broussonetia X + Additive</i>		≤ 2002	K+3	
Usumino-Paper	<i>Broussonetia X</i>	<i>Nasu</i>	≤ 2019	NW1	<i>Mino, Gifu</i>
Sekishuhan-Paper	<i>Broussonetia X</i>	<i>Shimane</i>	≤ 2019	NW2	<i>Masuda, Shimane</i>
Kaga-hosho (High-grade)	<i>Broussonetia X</i>	<i>Kaga</i>	≤ 2019	NW3	<i>Kanazawa, Ishikawa</i>
Hosokawa-Paper	<i>Broussonetia X</i>	<i>Nasu</i>	≤ 2019	NW4	<i>Ogawa, Saitama</i>
Honmino-Paper	<i>Broussonetia X</i>	<i>Nasu</i>	≤ 2019	NW5	<i>Mino, Gifu</i>
Akinikozo-Paper	<i>Broussonetia X</i>	<i>Kaga</i>	≤ 2019	NW6	<i>Kanazawa, Ishikawa</i>
Kigamikozo-Paper	<i>Broussonetia X</i>		≤ 2019	NW7	<i>Nanto, Toyama</i>

Traditional paper names are translated to English alphabet. Blanks refers to no record, or unclarified information

(*kurokawa*) and green cambium (*aohada*) were scraped off from the white innermost bark (*shirokawa*). The white innermost bark was peeled and digested in 10% sodium carbonate (FUJIFILM Wako Pure Chemical, Japan) in 1 L of distilled water. The digestion condition was controlled at 100 °C for 3–5 h. The digested white bark was rinsed with running distilled water to remove darker spots from the white bark. The white bark was placed to a wooden board, and hand-beaten by the experimenter with a wooden hammer for 1 h. The beaten fibers were dispersed in a water tank. The paper-forming tool called a screen (*su*), which was used to assemble fibers into a paper sheet. After enough fibers were retained on the screen to make paper, the formed sheet was stacked on top of the others. The stacked paper was pressurized with different weights of objects, increasing its consistency over 1 h. The stacked paper sheets were separately attached to a wooden board for drying overnight. Dried paper was stored within a desiccator with silica gel for the removal of humidity.

DNA extraction

DNA extraction from traditional paper samples was conducted by a modified version of the Cetyltrimethylammonium bromide (CTAB) method [28, 29]. The extraction buffer was prepared with 100 mL of 1 M Tris-HCl (Nippon Gene Co., Japan), 289 mL of 5 M

NaCl (FUJIFILM Wako Pure Chemical, Japan), 40 mL of 0.5 M EDTA (Ambion, Japan), 20 g of CTAB (FUJIFILM, Wako Pure Chemical, Japan), and volume up to 1 L with ultra-purified water. For individual DNA extraction, the β -mercaptoethanol (Sigma-Aldrich, USA) was diluted to 1% within the extraction buffer. The concentration of β -mercaptoethanol was increased to replace the use of polyvinylpyrrolidone for toxicity reasons. For each DNA extraction, 500 mg of traditional paper was homogenized with mortar and pestle in 5 mL of prepared extraction buffer. The homogenized samples were incubated at 60 °C for 25 min and treated with 6 mL of 24:1 chloroform isoamyl alcohol (Supelco, Germany). The sample was mixed by inverting it 5 to 10 times. Once the emulsion was observed, the sample was centrifuged at 5000 *G* for 15 min. The supernatant was separated from the precipitated solution. In the acquired clear aqueous solution, 2 mL of 5 M NaCl solution and 8 mL of cold 95% ethanol (FUJIFILM Wako Pure Chemical, Japan) were added to the sample solution. The sample solution was centrifuged at 5000 *G* for 10 min. After centrifugation, the supernatant was removed gently by decantation from the remaining pellet at the bottom. The washing process of the pellet was conducted with cold 80% ethanol, twice. The ethanol was completely removed with a pipette, not to interfere with it. The pellet was dried on a clean bench for 5 to 10 min, avoiding over-drying that

damages the remaining DNA. The pellet was diluted in 300 μL TE (Tris–EDTA) buffer. The TE buffer was prepared at the volume ratio of 1:10 with 1 M Tris–HCl and 0.5 M EDTA. The eluted sample was treated with RNase A (Nippon Gene, Japan), to remove remaining RNA. The Nano-Drop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) was used to analyze the DNA concentration and purity. The ratio of absorbance at 260 nm to that at 280 nm ($A_{260}/280$) indicates the DNA purity. DNA, RNA, and nucleotides show absorption peaks at the wavelength of 260 nm while protein, phenol, and other debris show those at the wavelength of 280 nm. The range for sufficient DNA purity to conduct further analysis is 1.7 to 2.0 in $A_{260}/280$ [30, 31]. For maintaining a consistent condition of DNA extraction, initial extractives were analyzed without further purification of the DNA.

Measurement of paper properties

The traditional paper samples were characterized from the aspect of the parameters of grammage, thickness, $CIE L^*a^*b^*$, and surface pH. Before measurement, all traditional paper samples were preserved at constant room temperature and desiccator for humidity control. All the parameters were measured by non-destructive methods for traditional paper sample. Each parameter was measured at least three times and averaged.

The grammage of traditional paper was calculated by dividing its weight (g) by the area of the sample. The thickness of traditional paper was measured with a micrometer (TW-21, Tozai Seiki, Japan). The density of the traditional paper was calculated by Eq. (1):

$$\text{Density}(\text{kg/m}^3) = \frac{\text{Grammage}(\text{g/m}^2)}{\text{Thickness}(\mu\text{m})} \times 1000 \quad (1)$$

Color values based on the $CIE L^*a^*b^*$ color system were measured with a spectrophotometer PF7000 (Nippon Denshoku, Tokyo, Japan) according to ISO/CIE standard 11664-4:2019. L^* , a^* , and b^* were measured after calibration with a reference standard and black cavity according to the manufacturer's manual. The scales of each parameter are explained: $L^*=0$ (Black) to 100 (White), a^* =negative (green) to positive (red), and b^* =negative (blue) to positive (yellow) [28]. The pH of the traditional paper surface was measured with a planar pH meter FPH70 (As One, Osaka, Japan).

ATR-FTIR measurement

The FTIR spectra of the traditional paper samples were measured in the ATR measurement mode of Nicolet iS50 (Thermo Fisher Scientific, USA). Before measurement, all traditional paper samples were preserved at constant

room temperature and desiccator for drying. For individual traditional paper samples, three measurements were conducted on different sections of the paper surface. The averaged spectra were used for analysis. The background calibration was conducted against a diamond crystal surface as a reference. The sampling spectra were measured with a total of 36 scans in the frequency range of 4000 cm^{-1} to 400 cm^{-1} at 4 cm^{-1} resolution. The measured spectra were sorted, averaged, and preprocessed with the Spectragryph optical spectroscopy software (Ver.1.2.16.1, F. Menges, Germany) [39].

Multivariate statistics

The multivariate statistical analysis was performed using R software, ver. 4.2.2 by R-studio. Also, the calculation and visualization were generated by the R packages listed [40–47]. The ATR-FTIR spectral data were mean-centered and baselined by the Spectragryph software before analysis. The preprocessing to eliminate the interferences or noises of measurement on the spectral data was adopted to 3400 cm^{-1} to 850 cm^{-1} regions. The preprocessing methods include normalization, Savitzky-Golay smoothing, the first derivative, the second derivative, Standard Normal Variate (SNV), Multiplicative Scatter Corrections (MSC), and a combination of Savitzky-Golay Smoothing + the first derivative, and + the second derivative. From the preprocessed spectral data, the prominent peaks were labeled and identified. The labeled peaks were evaluated as their properties and expected detection by the previous literature [30, 34, 48–58].

The Pearson correlation coefficient (r) between the x and y variables was generated by the following equation:

$$r = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 \sum (y - m_y)^2}}, \quad (2)$$

where, m_x and m_y are the means of x and y , respectively.

As r changes from -1 to 1 , the heatmap was represented as blue to red. The positive correlation (coefficient values near 1) was colored red, and the negative correlation (coefficient values near -1) was colored blue.

A partial least square (PLS) regression model was established to further identify the multilinear relationship from the DNA presence to paper properties and ATR-FTIR spectral data. The PLS regression model utilizes standardized variables within the significant variance of the dataset. The principal component is calculated as maximum variances, which prevents the overfitting of the training model. The least-square model fits the linear regression model and the k -fold cross-validation is applied in the present prediction model. The k -folding cross-validation was performed at $k=10$ folds with R-software. The performance of the model was

assessed by using Root Mean Square Error of Cross-Validation (RMSECV), Coefficient of determination (R^2), and Ratio of Performance to Deviation (RPD) of the model. RMSECV describes the average differences between observed and predicted values within the cross-validated data. As the RMSE values, the average error must be as small as possible to identify a highly validated prediction model. R^2 means the proportion of the response variable to predictor variables. R^2 values are denoted in the 0 to 1 range. An R^2 value higher than and the nearest to 1 will be determined as a stronger relationship between responsible and predictor variables. RPD is a measurement of consistency between the observed and predicted values. Higher consistency of RPD value will determine the most susceptible model for prediction. Additionally, the predicted values of the lab-scale traditional paper were used to validate the model with observed values. The best preprocessed for DNA purity was selected by assessing the validation.

Hierarchical Cluster Analysis (HCA) was calculated by pairwise dissimilarity between the observed values from the collected data of DNA concentration, purity, paper properties, and FTIR-ATR Spectra by Euclidean distance metric. The FTIR-ATR spectra were preprocessed by the Smoothing+the second derivative method. The hierarchical clusters were presented in Ward's minimum variance method to dendrogram by R software. The Ward's minimum variance method was chosen using the highest agglomerative coefficient near 1.

Results and discussion

DNA and paper property

The result of DNA extraction and paper property measurement for traditional paper samples are demonstrated in Table 2. The DNA concentration of the X3 paper sample was significantly higher than the other paper samples. Despite the general trend of traditional paper in a^* of CIE 1976 $L^*a^*b^*$ color space being toward negative (green), the X3 paper sample is the only sample with a positive (red) range. As the result, the extracted DNA of the X3 paper sample is possibly different from other traditional papers. The DNA concentration for NW3 was a negative value, which indicates the transparency for DNA eluted solution was not clear for the Nano-Drop spectrophotometer could not measure correctly. Based on previous research [59, 60], DNA extraction from plant fresh leaves is preferred over the roots or stem of the plant. The dried or frozen bark of a tree contains less DNA and cause difficulties in isolating DNA. Therefore, the anticipated amount of DNA from traditional paper samples was similarly low to DNA available in historical bark-cloth [61]. Moreover, the DNA presence was evaluated by staining the traditional paper with DAPI

(4',6-diamidino-2-phenylindole) and Fast Green staining. The stained traditional paper was observed on confocal laser scanning microscope (Leica TCS SP8). The Additional file 1: Fig. S1 provides the DNA presence with DAPI staining as DAPI strongly binds to AT region of double strand DNA. Also, the Fast Green stains the cellulose, cytoplasm and collagen. Under different ages and unknown preservation conditions, traditional paper samples have various paper properties. L^* (lightness) of CIE 1976 $L^*a^*b^*$ color space ranged from 60 to 85, a^* and b^* coordinates were towards negative (green) positive (yellow), respectively. The measured L^* , a^* , and b^* presented that traditional paper were white, green, and yellow in color. For an indication of chemical content changes, the CIE $L^*a^*b^*$ color measurement is a preliminary screening for observable changes [62]. Additionally, surface pH of traditional paper is an important paper property to be able to or not to observe DNA in a traditional paper surface. One characterization of double-stranded DNA (dsDNA) is its denaturation during the alkaline condition [63]. However, the surface pH of traditional paper is mostly neutral and not severely alkaline. Being a neutral surface, the remaining DNA on traditional paper surfaces is considered as not denatured. Thus, the paper properties are utilized to provide a prediction model for DNA presence and quality within the traditional paper by the multivariate statistical analysis including linear, multilinear, and clustering algorithms.

ATR-FTIR spectra

The ATR-FTIR spectral data are illustrated in Fig. 1. The characteristic peaks in ATR-FTIR spectral data are listed in Table 3. The fiber materials in paper consist of mainly cellulose, hemicellulose, and lignin, which are also interpreted as cell wall components. From the full spectral region, the peaks are identifiable in the two specific ranges. In the range from 3286 to 3321 cm^{-1} , hydroxy groups of cellulose show absorption [48, 49]. The intermolecular hydrogen bonds of cellulose are extensively projected. In the range from 2898 to 2902 cm^{-1} , the aliphatic C–H stretching of holocellulose is interpreted [50, 51, 64]. The remaining methyl or methylene groups of remaining lignin are found in traditional paper fiber materials. Most peaks are located in the fingerprint region of 1800 cm^{-1} to 850 cm^{-1} for lignocellulosic identification of wood species [30]. The fingerprint region of lignocellulosic absorbance differences is capable to classify and form clusters of wood-based products including traditional papers. The region of 1628–1638 cm^{-1} displays the lignin or cellulose by water absorbance and carboxyl group stretches [34, 52]. The aromatic skeletal or ring vibration is detected at 1599 cm^{-1} from traditional paper [30, 53]. Primary alcohol of methylene group is

Table 2 Result of DNA extraction and paper property measurements

Sample	DNA Concentration (ng/μL)	DNA Purity (A260/280)	Grammage (g/m ²)	Thickness (μm)	Density (kg/m ³)	L*	a*	b*	CIE Color	pH
X1	5.1	2.51	23.7	66.3	357.2	70.7	-1.44	8.69		8.52
X2	15.1	1.27	40.3	90.3	446.4	72.7	-1.77	11.20		8.86
X3	483.7	1.52	34.8	61.7	564.2	71.5	0.38	19.20		7.19
K-1	49.2	1.59	39.6	40.0	990.6	60.7	-1.35	4.25		7.83
K-2	13.7	1.66	28.8	69.7	413.9	70.4	-1.42	6.16		7.68
K-3	6.2	1.18	36.3	110.3	329.3	74.1	-1.41	8.50		7.45
K+1	22.3	1.88	81.3	61.3	1326.1	67.7	-0.77	1.95		9.12
K+2	2.5	0.96	34.0	71.7	474.3	78.7	-1.23	4.89		8.45
K+3	3.5	1.52	74.9	761.0	98.5	85.5	-1.71	9.53		8.90
Ue1	31.1	2.70	28.0	75.0	373.3	67.6	-2.19	9.63		6.73
Ue2	65.2	2.02	38.9	90.7	428.8	70.0	-1.74	5.68		7.96
Nan1	7.2	1.61	40.6	111.7	363.9	78.7	-0.96	3.98		8.23
NW1	24.9	1.79	23.4	40.7	575.4	60.2	-1.10	3.33		8.11
NW2	12.2	1.65	32.3	78.3	412.1	64.7	-0.83	1.76		8.66
NW3	-18.8	2.98	117.0	308.0	379.9	83.0	-1.65	10.08		8.38
NW4	35.7	1.89	24.5	58.7	418.0	62.7	-1.66	5.69		6.80
NW5	126.4	1.69	37.6	93.7	401.9	72.8	-1.05	3.50		8.96
NW6	22.0	1.72	39.6	99.3	398.3	71.3	-1.66	6.98		8.61
NW7	29.4	1.71	89.6	268.3	333.8	81.4	-0.29	6.82		8.24
Lab*	8.5	1.47	50.0	1330.0	35.0	75.7	-1.52	10.17		7.55

* Lab-scale traditional paper

assigned at the peaks of 1418 cm⁻¹ and 1315–1316 cm⁻¹ [48, 49, 54–56]. The strongest peak is demonstrated from 1020 to 1030 cm⁻¹ and assigned to carbonyl group

stretching and deformation [30, 34, 48, 57]. The carbonyl group valence vibration of lignin is detected at 992 cm⁻¹

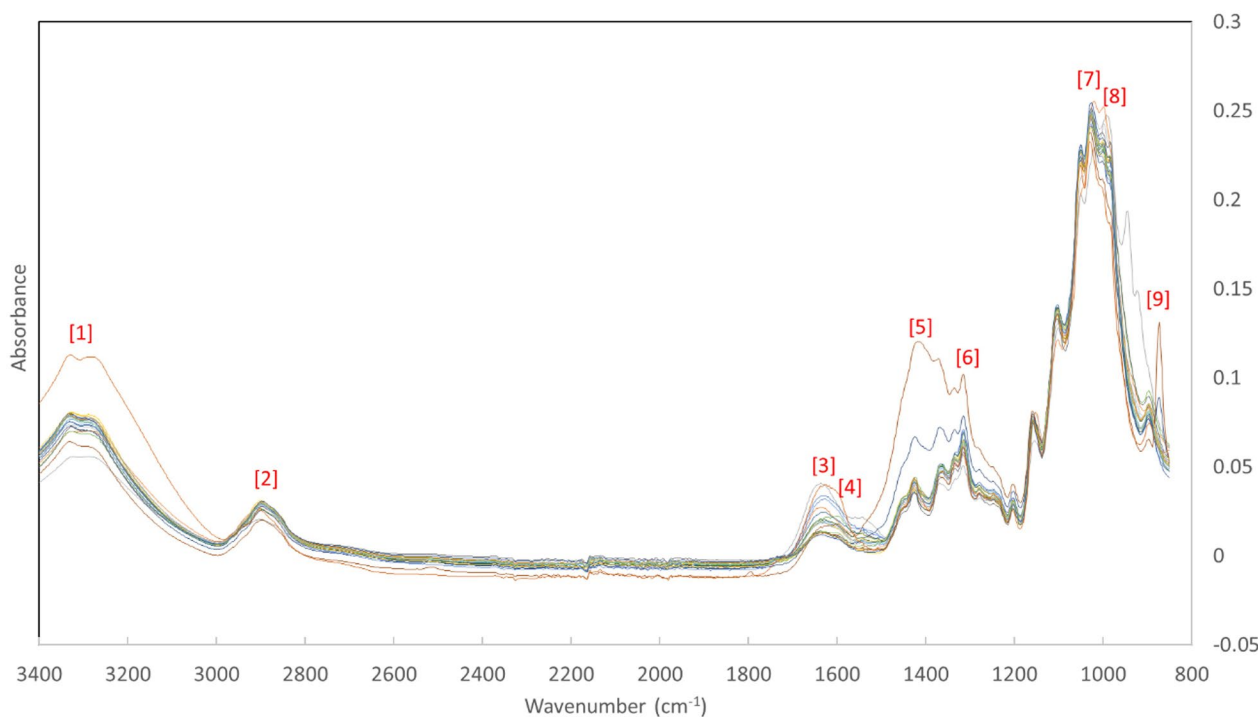


Fig. 1 ATR-FTIR Spectra in the range of 3400–850 cm^{-1} with labeled peaks

Table 3 Characteristic of peaks in the ATR-FTIR spectra, including according to literature data

Peak number	Absorption range	Interpretation	Assignment	References
1	3286–3321	Cellulose	O–H intermolecular bonds	[48, 49]
2	2898–2902	Cellulose and Hemicellulose	aliphatic C–H stretching	[50, 51, 64]
3	1628–1638	Lignin or Cellulose	O–H absorption, COO– Stretching	[34, 52]
4	1599	Lignin and Flavone	Aromatic skeletal or ring vibration	[30, 53]
5	1418	Starch, Lignin, Cellulose	CH ₂ scissoring	[49, 54]
6	1315–1316	Xyloglucan, Cellulose I and II	CH ₂ symmetric bending, CH ₂ rocking vibration, CH ₂ wagging	[48, 54–56]
7	1020–1030	Cellulose I and II, Polysaccharide, Lignin	C–O stretching, C–O deformation	[30, 34, 48, 57]
8	992	Lignin	C–O valence vibration	[49, 55]
9	874	Pinene (Terpene)	Terminal methylene group absorption	[58]

[49, 55]. Pinene remains in the traditional paper and has methylene groups assigned at 874 cm^{-1} [58].

The absorption peaks of traditional paper display differently by the source of materials. Japanese traditional papers are produced from many materials, such as paper mulberry, hemp, Japanese *gampi*, *mitsumata*, and so on. This research focuses on the *Broussonetia* species, which is paper mulberry. There are materials unknown contained among *Broussonetia* fibers with other major additives in traditional paper. By spectral

data of macromolecular polymers, the DNA purity was predicted with a higher accuracy. The lignocellulosic residues and humic substances are acknowledged for their inhabitant effect on DNA extraction [65]. With the possible detection of several components within spectral data, the prediction model of DNA quality was developed precisely. The traditional paper sheets were identified and clustered according to the ATR-FTIR spectra, paper properties, and DNA presence.

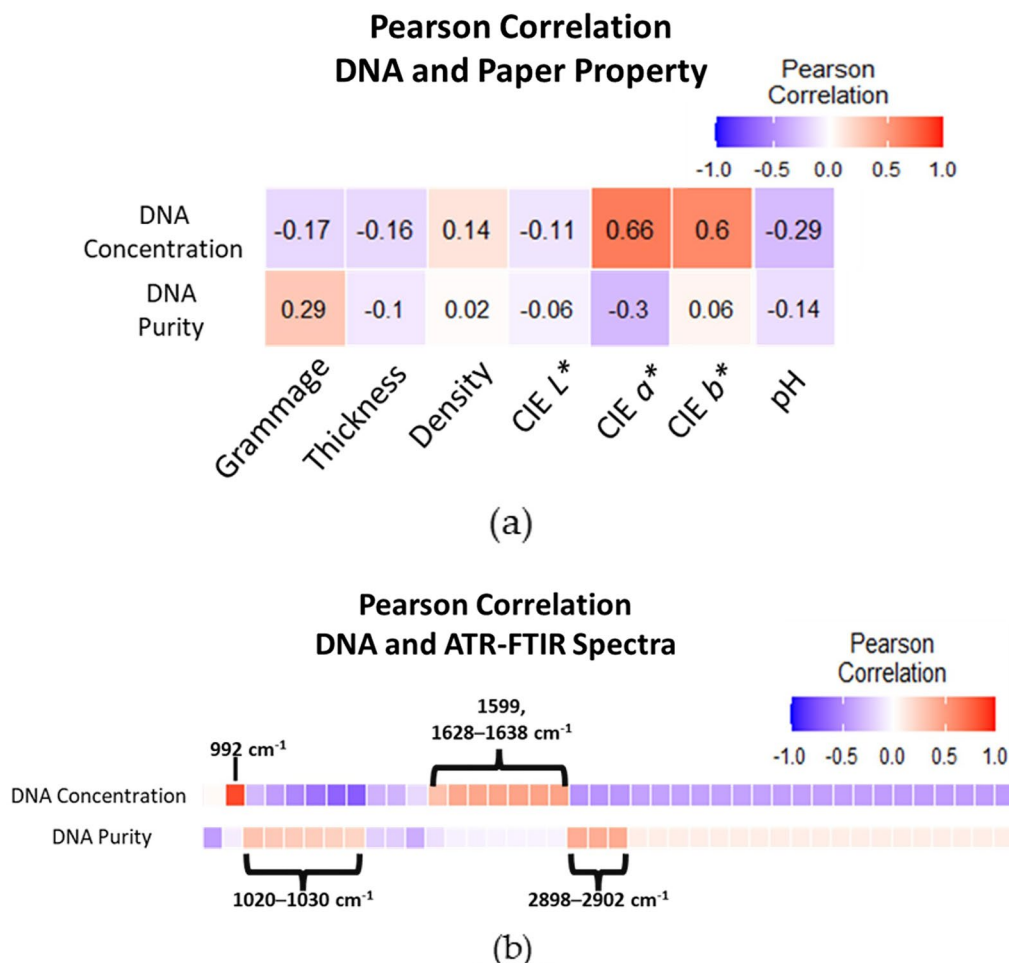


Fig. 2 Heatmap of r between DNA properties (concentration and purity) and paper properties (a) and that between DNA properties and absorption spectral patterns of ATR-FTIR spectra (b)

Pearson correlation coefficient

R was calculated between DNA and paper properties as Additional file 1: Table S1 shows. The correlation was demonstrated by the heatmap in Fig. 2a. PCC determines the linear relationship between two variables; negative, positive, or none. The DNA concentration had a PCC of 0.663 and 0.601 with CIE a^* and CIE b^* , respectively. The positive correlation was demonstrated in red. The pH of the traditional paper also displayed a slight correlation negatively as colored blue. Yet, DNA purity does not provide any significant linear correlation to paper properties. The highest correlation was with CIE a^* parameter at -0.298 . Despite the significance, the CIE a^* parameter to measure the color of green to red demonstrated a positive linear correlation to DNA concentration and a negative linear correlation to DNA purity. Moreover, the surface pH of traditional paper was recognized as an important characteristic to be correlated to DNA presence. With a relatively narrow range of parameters at neutral pH 7 to

8, the DNA concentration and purity were not strongly correlated.

r values were calculated between DNA concentration or purity on one side and measured parameters of ATR-FTIR labeled peaks in Additional file 1: Table S2 on the other. The correlation is demonstrated by the heatmap in Fig. 2b. The DNA concentration to ATR-FTIR spectral data was mostly negatively correlated with identified peaks. The peaks of 992 cm⁻¹, 1599 cm⁻¹, and 1628–1638 cm⁻¹ regions were positively correlated. The correlation was highest at the peak of 992 cm⁻¹, which has a coefficient of 0.857. Referring to Table 3, the peaks are identified as lignin-specific functional groups. According to the Pearson correlation model, the lignin functional groups were positively correlated with DNA concentration. This connects to the colors of CIE a^* that was positively correlated to DNA concentration. The remaining lignin will produce the color green on the surface of the traditional paper. In addition, the fingerprint region of

1020–1030 cm^{-1} and 1628–1638 cm^{-1} constantly had a significant correlation to DNA concentration. The linear correlation between the spectral data and DNA purity is not strong. The highest correlation coefficient between DNA purity and spectral data was shown in the region of 2898–2902 cm^{-1} . The coefficient was near 0.42, meaning positive correlation. On the other hand, the DNA concentration in the region of 2898–2902 cm^{-1} displayed a negatively correlated coefficient near -0.45 . According to the Pearson correlation, the methyl or methylene groups of holocellulose fibers were positively correlated to DNA purity and negatively correlated to DNA concentration. The presence of methyl groups was influential in DNA purity and concentration. DNA methylation occurs as the methyl group is added to the DNA structure and changes the genes. In the environment of papermaking, DNA methylation will not occur due to the lack of specific enzymes or the temperature conditions. Yet, the detection of methyl groups in Nano-drop spectrophotometers will influence the DNA concentration and purity values. Non-destructive measurement of paper properties and ATR-FTIR spectra applied to DNA concentration and purity explicates possible linear correlations. Using IR qualitative data and spectrophotometers, this may preliminarily hypothesize a link between the presence of DNA and the chemical components of traditional paper. In addition, the yellow and green colors of traditional paper may indicate a higher content of lignin, which could preserve the DNA of source materials by remaining in the cell wall.

Partial least square regression modelling

The PLS regression model was established based on the non-destructive paper properties and preprocessed ATR-FTIR spectral data. The multilinearity correlation between the DNA properties and the paper properties or spectral data was k-fold validated at $k=10$. For assessment parameters, RMSECV, R^2 , and RPD were calculated. The smoothing+the second derivative preprocessing method was most suitable for the DNA purity prediction model. Despite the low determination coefficient less than that determined in the second

derivative preprocessing method, other assessment parameters of the lowest RMSECV and highest RPD provided enough creditability to the smoothing+the second derivative preprocessed spectra prediction model. The DNA purity prediction model demonstrated a highly accurate parameter of low RMSE value of 0.91, good data fitting with a determination coefficient of as high as 0.964, and a consistent prediction power of 5.38 according to Table 4. The predicted value of DNA purity was 1.49 and the observed DNA purity for the lab-scale traditional paper was 1.47. The predicted value was only 0.02 higher than the observed value. Additionally, this study was conducted for the first time to present the extraction of DNA from traditional paper. Therefore, further development of DNA extraction methodology and a larger sample portion would provide higher accuracy to the PLS regression model with smoothing+the second derivative preprocessing method.

The DNA purity prediction model is demonstrated in the scatter plot of Fig. 3. The observed predicted values are not widely dispersed along the determination coefficient line with a determination coefficient of 0.964. According to the assessment parameters on cross-validated PLS regression mode, the prediction models are suitable for the initial estimation of DNA presence in the traditional paper. By the initial preliminary evaluation, the decision may establish whether to conduct genetic information through destructive DNA extraction or not.

Table 4 PLS validation assessment parameters of smoothing+2nd derivative preprocessing methods for DNA concentration and DNA purity

Smoothing+2nd derivative preprocessing method	RMSECV	R^2	RPD	Lab-made paper validation	
				Observed	Predicted
DNA purity	0.091	0.964	5.38	1.47	1.49

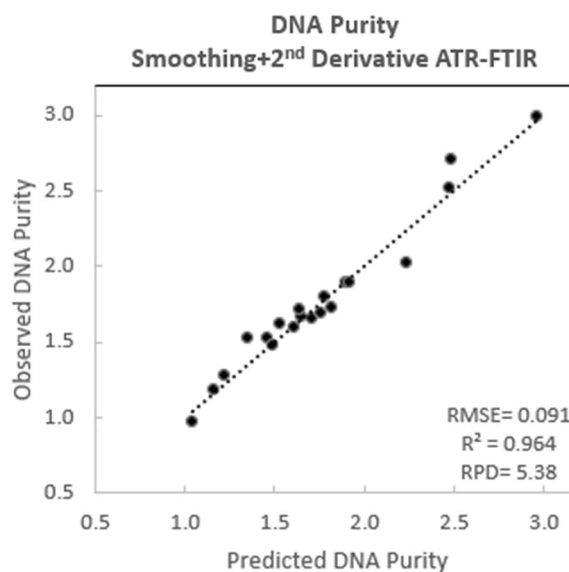


Fig. 3 Relationships between predicted and observed DNA purity both with smoothing+the second derivative preprocessed ATR-FTIR data. Based on the PLS regression model

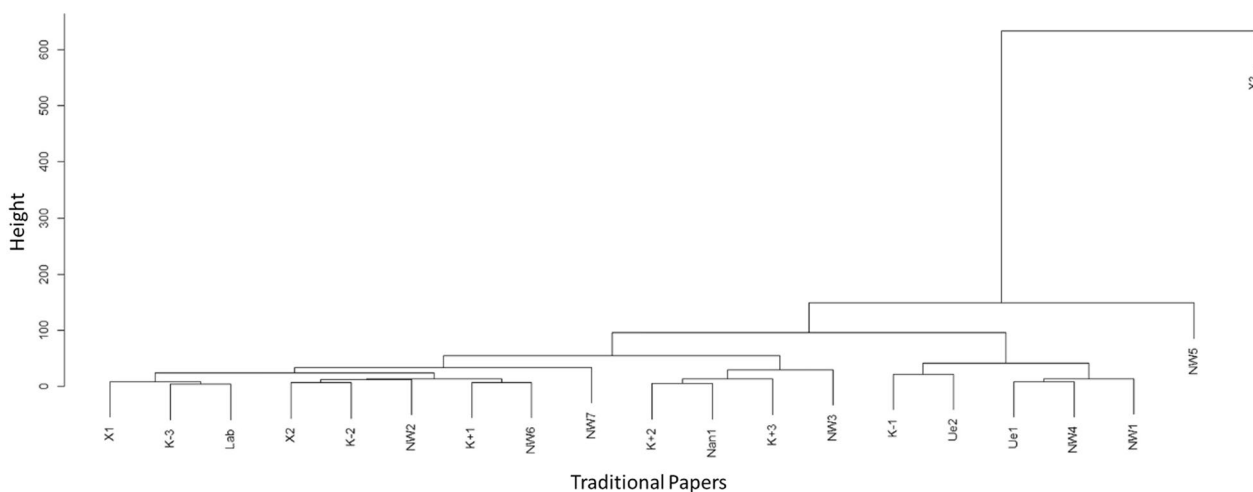


Fig. 4 Hierarchical clustering analysis (HCA) dendrogram of traditional papers

Hierarchical clustering analysis

The hierarchical clustering algorithm was based on bottom-up methods of agglomerative approach, in which individual variables are merged with similar variables as increasing height. The Euclidean distance is calculated based on optimized linkage criterion and individual variables. The optimized linkage criterion was Ward's minimum variance method, in which the agglomerative coefficient was highest as described in Additional file 1: Table S4. The dendrogram of traditional paper samples is demonstrated in Fig. 4 based on DNA concentration and purity, paper properties, and smoothing+the second derivative preprocessed ATR-FTIR spectral data. X3 paper was the farthest sample from other clusters. X3 traditional paper was the outlier with excessively high DNA concentration and had different paper properties of CIE a^*b^* . By the dendrogram, X3 paper was difficult to evaluate with current samples, then excluded. An optical microscope was used to evaluate the possible contamination of mold in the X3 traditional paper with high DNA concentration. Additional file 1: Fig. S2 shows the X3 traditional paper sample with a potential contaminant, while the other traditional paper samples only have clear fiber structures. Ue1 and Ue2 papers had unknown origin and no written record was identified. The only information given to the experimenter was that the two papers were produced by the same craftsman. By dendrogram, Ue1 and Ue2 occur in different clusters but biologically closely related. Ue1 was in the same cluster as NW4, which was used to produce traditional paper in *Ogawa, Saitama* from *Nasu* Paper mulberry tree. Ue2 was in the same cluster as K-1 used to produce traditional paper in *Mino, Gifu* from *Nasu* Paper mulberry tree. Therefore, it is highly possible that

Ue1 and Ue2 were produced with *Nasu* Paper mulberry (originated in the northeastern region of Japan) by the same craftsman but with different techniques or additives. Traditional papers K-1, Ue2, Ue1, NW4, and NW1 are grouped as traditional paper made from *Nasu* Paper mulberry species. Other traditional papers are clustered in a mixture of *Kaga, Toyama, Shimame, and Kouchi* Paper mulberry trees (originated from the southwestern region of Japan). Despite a variety of traditional paper-making procedure and preservation conditions, the clustering method based on non-destructively measured paper properties, ATR-FTIR spectra, and DNA provide sufficient scientific background regarding the origin of the paper mulberry tree on a large scale within Japan. The results of the multivariate statistical analysis provide an expectation of a novel approach to the cultural heritage of traditional papers. The preliminary screening for the possibility of DNA presence and quality leads to the decision on the need for DNA analysis. For the future research, the optimization of DNA extraction in micro-destructive methods and interpretation of genomic sequenced information will provide highly accurate clustering analysis.

Conclusions

The biological approach to historical traditional paper analysis was not an option in heritage research due to its destructiveness on the valuable paper samples. The current study examined the DNA presence and interaction within the traditional paper and non-destructive measurement. The result of this research provides sufficient evidence for a preliminary exploration of the need for DNA analysis for historical traditional paper. The linear correlation is structured between DNA

properties such as concentration and purity and *CIE L*a*b** color metric system. The overall color measurement of traditional paper indicates green and yellow instead of red and blue. This observation connects with ATR-FTIR spectral data. The peaks of 992 cm^{-1} , 1599 cm^{-1} , and $1628\text{--}1638\text{ cm}^{-1}$ regions are positively correlated to DNA concentration. The labeled peaks were assigned to lignin-specific functional groups. The availability of lignin in the traditional paper determines the color of green and yellow, which is connected to cell wall existence. The higher amount of cell wall is positively correlated to DNA survivability through the traditional papermaking process. The PLS regression prediction model is established for DNA purity. The preprocessing method of spectral data is optimized to the smoothing + the second derivative method. The assessment parameter for the cross-validated prediction model includes low RMSECV at 0.091 for DNA purity, respectively. The determination coefficient was near the perfect fitting value of 1. With the established PLS regression model, it is possible to acquire explicit DNA presence and quality from the traditional paper sample in the preliminary screening. Despite further development on the validity of the DNA extraction method, the prediction model will evaluate the fast preliminary screening of DNA presence and purity within the traditional paper. The HCA dendrogram demonstrated the clusters of similar traditional papers, based on the measurement of paper properties, such as ATR-FTIR spectra and DNA concentration and purity. The HCA provided general identification of the source origin of paper mulberry material. The clustering largely separated *Nasu* paper mulberry (originated in north-eastern Japan) from another paper mulberry (originated in southwestern Japan). Through the multivariate statistical analysis, the scientific evidence on traditional paper identification based on written records is justifiable. With the proposed novel non-destructive method analysis on DNA presence and quality, further sequenced information of DNA and genetic markers will enhance the understanding of historical heritages and its anthropological values. By establishing the library of genetic information, the genomic analysis of each traditional sample will clarify the location and timeline of material sources. The taxonomical tree will be established to support the information. The research involved in DNA and traditional papermaking would generate a highly valid analysis of genetic information on traditional papers. The interpretation of genomic information will lay out and untangle the puzzled history of traditional papermaking propagation in a wide geographical spectrum.

Abbreviations

DNA	Deoxyribonucleic acid
CTAB	Cetyltrimethylammonium bromide
CIE	Commission Internationale de l'Éclairage
ATR-FTIR	Attenuated Total Reflection-Fourier Transform Infrared
PLS	Partial least square
SNV	Standard normal variate
MSC	Multiplicative scatter correction
RMSECV	Root mean square error of cross-validation
RPD	Ratio of performance to deviation
HCA	Hierarchical clustering analysis

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40494-023-01065-5>.

Additional file 1: Table S1. Pearson correlation coefficient between DNA and paper properties. **Table S2.** Pearson correlation coefficient between DNA and ATR-FTIR spectral data. **Table S3.** PLS validation assessment parameters of different preprocessing methods for DNA purity. **Table S4.** Agglomerative coefficient of different clustering. **Figure S1.** Confocal laser fluorescence microscopy on DNA presence in traditional paper. **Figure S2.** Optical microscopy on paper sample.

Acknowledgements

The authors would like to thank Mr. Akihiko Takashima, Historiographical institute, the University of Tokyo for providing the traditional paper samples. The authors are grateful to Dr. Friedrich Menges for the use of Spectragryph with basic tools for FTIR-ATR spectra preprocessing.

Author contributions

MS wrote the manuscript, collected materials, performed all experiments, analyzed all datasets and figures, and acquired funds. TE provided material, acquired funds, reviewed the manuscript, and supervised the overall project. All authors read and approved the final manuscript.

Funding

This research was supported by Yoshida Manabu-Kinen Foundation, Japan (2023).

Data availability

Not applicable.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 18 July 2023 Accepted: 9 October 2023

Published online: 24 October 2023

References

- Gao ZY. A research on early spread of traditional hand-made paper in East Asia Countries. *Adv Mater Res*. 2011;174:537–40.
- Zhibin M. The spirit of our times embodied in the Chinese history textbook for junior high schools in the nine-year mandatory education (People's Education Publishing House Edition). *Chin Educ Soc*. 1999;32:55–61. <https://doi.org/10.2753/CED1061-1932320655>.
- Gao Qi, Jiang Dayong, Meng Xiangwu. Analysis of the ancient papermaking technology from Qiliang Village, Shaanxi Province. In: 2020 international conference on big data application & economic management

- (ICBDEM 2020). Francis Academic Press; 2020. https://webofproceedings.org/proceedings_series/article/artid/16129.html. Accessed 6 Mar 2023.
- Song M. The history and characteristics of traditional Korean books and bookbinding. *J Inst Conserv*. 2009;32:53–78. <https://doi.org/10.1080/19455220802630743>.
 - Almogi O. The written word and its media within the Tibetan Culture Sphere. 2016.
 - Tsundru J. Preserve and development of the manufacturing technology of traditional Tibetan Paper. Lhasa; 2010.
 - Solberg J. The papermaking tradition of Central Asia. Iowa: University of Iowa; 2018.
 - Seki M. Database of traditional papermaking centers in East Asian Regions. *Senri Ethnol Stud*. 2013;85:61–81.
 - Karuno H. Shifu: a traditional paper textile of Japan. *Text Soc Am Symp Proc*. 2016. <https://digitalcommons.unl.edu/tsaconf/977>. Accessed 6 Mar 2023.
 - Go IH, Jo AH, Jeong SL, Heo TY, Cho K, Choi TH. Predictive model of geographical origin discrimination of paper mulberry and handmade paper using ICP-AES/MS and multivariate statistical analysis. *J Cult Herit*. 2021;49:222–8.
 - Han B, Yang Y, Wang B, Jiang H, Sablier M. Rapid identification of bast fibers in ancient handmade papers based on improved characterization of lignin monomers by Py-GC×GC/MS. *Cellulose*. 2023;30:575–90.
 - Robert J. C. Chemical analysis in the paper formation process. 1996. <https://arch.neicon.ru/xmlui/handle/123456789/1708860>. Accessed 6 Mar 2023.
 - Yum H. Traditional Korean Papermaking: analytical examination of historic Korean papers and research into history, materials and techniques of traditional papermaking of Korea. 2009;
 - Avataneo C, Sablier M. New criteria for the characterization of traditional East Asian papers. *Environ Sci Pollut Res*. 2017;24:2166–81.
 - Hant Z, Kida K, Handa M, Inaba M. Effect of Cooking method on wet tensile strength of Kozo paper. *紙工技協誌*. 2020;74:928–35.
 - Yum H. History and function of dispersion aids used in traditional East Asian Papermaking. *J Inst Conserv*. 2011;34:202–8.
 - Kuo W-H, Liu S-H, Chang C-C, Hsieh C-L, Li Y-H, Ito T, et al. Plastome phylogenomics of *Allaenanthus*, *Broussonetia* and *Malaisia* (Dorsteniaceae, Moraceae) and the origin of *B. kazinoki*. *J Plant Res*. 2022;135:203–20.
 - Goltz D, Attas M, Young G, Cloutis E, Bedynski M. Assessing stains on historical documents using hyperspectral imaging. *J Cult Herit*. 2010;11:19–26.
 - Dzinavatonga K, Bharuth-Ram K, Medupe TR. Mössbauer spectroscopy analysis of valence state of iron in historical documents obtained from the National Library of South Africa. *J Cult Herit*. 2015;16:377–80.
 - Yan Y, Wen C, Jin M, Duan L, Zhang R, Luo C, et al. FTIR spectroscopy in cultural heritage studies: non-destructive analysis of Chinese Handmade Papers. *Chem Res Chin Univ*. 2019;35:586–91.
 - gorassini A, Calvini P, Baldin A. Fourier transform infrared spectroscopy (FTIR) analysis of historic paper documents as a preliminary step for chemometrical analysis. In: Italy: CMA4CH 2nd Mediterranean meeting on multivariate analysis and chemometry for cultural heritage and environment. 2008. p. 47–8.
 - de Palaminy L, Daher C, Moulherat C. Development of a non-destructive methodology using ATR-FTIR and chemometrics to discriminate wild silk species in heritage collections. *Spectrochim Acta A Mol Biomol Spectrosc*. 2022;270:120788.
 - Sakamoto S, Okada Y. Paper analysis and paper history from ancient Chinese paper to Japanese washi. In: 2013 International Conference on Culture and Computing. 2013. pp 51–6. <http://ieeexplore.ieee.org/document/6680330/>. Accessed 6 Mar 2023.
 - Han B, Niang J, Rao H, Lyu N, Oda H, Sakamoto S, et al. Paper fragments from the Tibetan Samye Monastery: clues for an unusual sizing recipe implying wheat starch and milk in early Tibetan papermaking. *J Archaeol Sci Rep*. 2021;36:102793.
 - Han B, Daheur G, Sablier M. Py-GC×GC/MS in cultural heritage studies: an illustration through analytical characterization of traditional East Asian handmade papers. *J Anal Appl Pyrol*. 2016;122:458–67.
 - Han B, Vial J, Sakamoto S, Sablier M. Identification of traditional East Asian handmade papers through the multivariate data analysis of pyrolysis-GC/MS data. *Analyst*. 2019;144:1230–44.
 - Han B, Vilmont L-B, Kim H-J, Lavédrine B, Sakamoto S, Sablier M. Characterization of Korean handmade papers collected in a Hanji reference book. *Herit Sci*. 2021;9:96. <https://doi.org/10.1186/s40494-021-00570-9>.
 - Mo J, Zhang W, Fu X, Lu W. Near infrared spectroscopy to evaluate change in color and chemical composition in heat-treated bamboo. *J Infrared Spectrosc*. 2020;28:214–23.
 - Zhou G, Taylor G, Polle A. FTIR-ATR-based prediction and modelling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. *Plant Methods*. 2011;7:1–10.
 - Javier-Astete R, Jimenez-Davalos J, Zolla G. Determination of hemicellulose, cellulose, holocellulose and lignin content using FTIR in *Calycophyllum spruceanum* (Benth.) K. Schum. and *Guazuma crinita* Lam. *PLOS ONE*. 2021;16:e0256559. <https://doi.org/10.1371/journal.pone.0256559>.
 - Di Tullio V, Doherty B, Capitani D, Miliani C, Greco E, Ciliberto E, et al. NMR spectroscopy and micro-analytical techniques for studying the constitutive materials and the state of conservation of an ancient Tapa barkcloth from Polynesia, is. Wallis. *J Cult Herit*. 2020;45:379–88.
 - Song SP, Zhang H, Lang Q, Pu JW. Application of near-infrared spectroscopy to predict paper properties of Acacia. *Adv Mater Res*. 2011;236–238:1372–8.
 - Wertz JH, McClelland AA, Mayer DD, Knipe P. Modeling chemical tests and fiber identification of paper materials using principal component analysis and specular reflection FTIR data. *Heritage*. 2022;5:1960–73.
 - Sharma V, Yadav J, Kumar R, Tesarova D, Ekielski A, Mishra PK. On the rapid and non-destructive approach with wood identification using ATR-FTIR spectroscopy and chemometric methods. *Vib Spectrosc*. 2020;110:103097.
 - Moncada X, Payacán C, Arriaza F, Lobos S, Seelenfreund D, Seelenfreund A. DNA extraction and amplification from contemporary Polynesian Bark-Cloth. *PLoS ONE*. 2013;8:e56549. <https://doi.org/10.1371/journal.pone.0056549>.
 - Muhammad AL, Guang-Ning Y, Norman FW, Bruce IR. A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Mol Biol Rep*. 1994;12:6–13.
 - Piskata Z, Servusova E, Babak V, Nesvadbova M, Borilova G. The quality of DNA isolated from processed food and feed via different extraction procedures. *Molecules*. 2019;24:1188.
 - Jorgez CJ, Dang DD, Simpson JL, Lewis DE, Bischoff FZ. Quantity versus quality: optimal methods for cell-free DNA isolation from plasma of pregnant women. *Genet Med*. 2006;8:615–9.
 - Menges F. Spectragryph—optical spectroscopy software. 2023. <http://www.effemm2.de/spectragryph/>. Accessed 6 Mar 2023.
 - Team R RC. A language and environment for statistical computing 2015. Vienna: R Foundation for Statistical Computing; 2018.
 - Schloerke B, Crowley J, Cook D, Briatte F, Marbach M, Thoen E, et al. GGally: extension to 'ggplot2'. R package version 1.4.0. *R Found Stat Comput*. 2018;
 - Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. *J Open Source Softw*. 2019;4(43):1686.
 - Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. Cluster: cluster analysis basics and extensions. 2012.
 - Kassambara A, Mundt F. factoextra: extract and visualize the results of multivariate data analyses. 2017. <https://CRAN.R-project.org/package=factoextra>. Accessed 16 Mar 2023.
 - Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*. 2015. <https://doi.org/10.1093/bioinformatics/btv428>. Accessed 16 Mar 2023.
 - Liland KH, Mevik B-H, Wehrens R, Hiemstra P. pls. Partial least squares and principal component regression. 2011. <https://CRAN.R-project.org/package=pls>. Accessed 16 Mar 2023.
 - Kuhn M, Wing J, Weston S, Williams A, Keefer C, et al. Classification and regression training. 2021. <https://CRAN.R-project.org/package=caret>. Accessed 16 Mar 2023.
 - Popescu C-M, Popescu M-C, Singurel G, Vasile C, Argyropoulos DS, Willfor S. Spectral characterization of eucalyptus wood. *Appl Spectrosc*. 2007;61:1168–77. <https://doi.org/10.1366/000370207782597076>.
 - Schwanninger M, Rodrigues JC, Pereira H, Hinterstoisser B. Effects of short-time vibratory ball milling on the shape of FT-IR spectra of wood and cellulose. *Vib Spectrosc*. 2004;36:23–40.

50. Vartanian E, Barres O, Roque C. FTIR spectroscopy of woods: a new approach to study the weathering of the carving face of a sculpture. *Spectrochim Acta A Mol Biomol Spectrosc.* 2014;136PC:1255–9.
51. Jiang S-F, Sheng G-P, Jiang H. Advances in the characterization methods of biomass pyrolysis products. *ACS Sustain Chem Eng.* 2019;7:12639–55. <https://doi.org/10.1021/acssuschemeng.9b00868>.
52. Wang X, Ren H. Comparative study of the photo-discoloration of moso bamboo (*Phyllostachys pubescens* Mazel) and two wood species. *Appl Surf Sci.* 2008;254:7029–34.
53. Zhang F-D, Xu C-H, Li M-Y, Chen X-D, Zhou Q, Huang A-M. Identification of *Dalbergia cochinchinensis* (CITES Appendix II) from other three *Dalbergia* species using FT-IR and 2D correlation IR spectroscopy. *Wood Sci Technol.* 2016;50:693–704. <https://doi.org/10.1007/s00226-016-0815-3>.
54. Fahey LM, Nieuwoudt MK, Harris PJ. Predicting the cell-wall compositions of *Pinus radiata* (radiata pine) wood using ATR and transmission FTIR spectroscopies. *Cellulose.* 2017;24:5275–93. <https://doi.org/10.1007/s10570-017-1506-4>.
55. Canteri MHG, Renard CMGC, Le Bourvellec C, Bureau S. ATR-FTIR spectroscopy to determine cell wall composition: application on a large diversity of fruits and vegetables. *Carbohydr Polym.* 2019;212:186–96.
56. Alonso-Simón A, García-Angulo P, Mérida H, Encina A, Álvarez JM, Acebes JL. The use of FTIR spectroscopy to monitor modifications in plant cell wall architecture caused by cellulose biosynthesis inhibitors. *Plant Signal Behav.* 2011;6:1104–10. <https://doi.org/10.4161/psb.6.8.15793>.
57. Uar G, Staccioli CG, Stoll M. Chemical composition and ultrastructure of a fossil wood from the genus of ancestral sequoia. *Holz Als Roh- Werkst.* 1996;6:411–21.
58. Schulz H, Baranska M. Fruits and vegetables. In: Sun D, editor. *Infrared spectroscopy for food quality analysis and control.* Amsterdam: Elsevier; 2009. p. 321–53.
59. Novaes RML, Rodrigues JG, Lovato MB. An efficient protocol for tissue sampling and DNA isolation from the stem bark of Leguminosae trees. *Genet Mol Res.* 2009;8:86–96.
60. Irfan M, Ting ZT, Yang W, Chunyu Z, Qing M, Lijun Z, et al. Modification of CTAB protocol for maize genomic DNA extraction. *Res J Biotechnol.* 2013;8:41–5.
61. Peña-Ahumada B, Saldarriaga-Córdoba M, Kardailsky O, Moncada X, Moraga M, Matisoo-Smith E, et al. A tale of textiles: Genetic characterization of historical paper mulberry barkcloth from Oceania. *PLoS ONE.* 2020;15:e0233113. <https://doi.org/10.1371/journal.pone.0233113>.
62. Mononen K, Jääskeläinen A-S, Alvila L, Pakkanen TT, Vuorinen T. Chemical changes in silver birch (*Betula pendula* Roth) wood caused by hydrogen peroxide bleaching and monitored by color measurement (CIELab) and UV-Vis, FTIR and UVR spectroscopy. *Holzforschung.* 2005;59:381–8. <https://doi.org/10.1515/HF.2005.063/html>.
63. Wang X, Lim HJ, Son A. Characterization of denaturation and renaturation of DNA for DNA hybridization. *Environ Health Toxicol.* 2014;29:e2014007.
64. Md Salim R, Asik J, Sarjadi MS. Chemical functional groups of extractions, cellulose and lignin extracted from native *Leucaena leucocephala* bark. *Wood Sci Technol.* 2021;55:295–313. <https://doi.org/10.1007/s00226-020-01258-2>.
65. Costa ÂMA, Santos AO, Sousa J, Rodrigues JL, Gudiña EJ, Silvério SC, et al. Improved method for the extraction of high-quality DNA from lignocellulosic compost samples for metagenomic studies. *Appl Microbiol Biotechnol.* 2021;105:8881–93. <https://doi.org/10.1007/s00253-021-11647-7>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
