

Wood Chemistry

Max L. Flaig, Jens Berger, Philip Wenig, Andrea Olbrich and Bodo Saake*

Identification of tropical wood species in paper: a new chemotaxonomic method based on extractives

<https://doi.org/10.1515/hf-2023-0048>

Received May 17, 2023; accepted October 5, 2023;

published online November 7, 2023

Keywords: CITES; database; extractives; GC-MS; mixed tropical hardwood; pulp and paper

Abstract: The European Deforestation Regulation 2023/1115 (EUDR) prohibits trading of wood and wood products obtained from illegal logging on the EU market. While the identification of solid wood via anatomy, chemistry and genetics has already been established, there is a lack of identification methods for pulp and paper that complement anatomy. This publication presents a newly developed chemotaxonomic method for identifying mixed tropical hardwood (MTH) species in pulp and paper products based on their extractives analyzed with thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS). The measured data was processed and compared to identify marker substances and was then merged into a fingerprint database for identifying MTH species in paper of unknown composition. As database references, fully bleached kraft pulps were produced from 38 anatomically identified wood samples and then cryo-ball milled and extracted successively with *n*-hexane and acetone. The results show that the remaining wood extractives generated from bleached pulps are specific enough to find chemical relevant marker substances to detect MTH species. As chemical composition and anatomy are independent characteristics of wood, this paper makes a completely independent method available, which potentially improves the screening for Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) protected species.

1 Introduction

Most of the time, pulp and paper are products obtained from the renewable raw material wood. Since quality requirements for pulp are simply focused on the individual fiber and not on the wood structure itself, inferior quality and mixed wood can be involved. Thus, this can be connected to the use of wood from clear-cutting of natural forests, such as those in Indonesia for example, for pulp production from so called mixed tropical hardwood (MTH) (Uryu 2008). There is a suspicion for the whole of South-East Asia regarding tropical woods being illegally cleared e.g. to make room for palm oil plantations while their wood is used for pulp production (Hirschberger et al. 2010). Over a period of time the annual pulp and paper production in Southeast Asia and China was steadily increasing, reaching a record high in 2018. Since then, production volume has stagnated at a level close to the all-time high (FAO 2021). Even in industrialized countries such as Canada or Australia (e.g. Tasmania) virgin forests are still being clear-cut for pulp and paper production (Hirschberger et al. 2010).

In order to preserve biodiversity and to protect the rainforest as an effective ecosystem it is necessary to substantiate the suspicion named above by identifying tropical timber species in pulp and paper within global trade specified by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) as proof. In relation to that, the European Timber Regulation (EUTR) from 03 March 2013 prohibited the imports of illegally harvested timber and timber products on EU internal markets (European Union 2010) which also applied to pulp and paper. This regulation was updated by the European Deforestation Regulation (EUDR) on 31 of May 2023 (European Union 2023) placing due diligence onto market participants to take appropriate measures and comply with the standards required. Considering the fact that manufactured paper

*Corresponding author: **Bodo Saake**, Institute of Wood Science, Chemical Wood Technology, University of Hamburg, Haidkrugsweg 1, 22885 Barsbüttel-Willinghusen, Germany, E-mail: bodo.saake@uni-hamburg.de

Max L. Flaig and Jens Berger, Institute of Wood Science, Chemical Wood Technology, University of Hamburg, Haidkrugsweg 1, 22885 Barsbüttel-Willinghusen, Germany

Philip Wenig, Lablicate GmbH, Martin-Luther-King-Platz 6, 20146 Hamburg, Germany

Andrea Olbrich, Johann Heinrich von Thünen Institute (TI), Institute of Wood Research, Leuschnerstraße 91, 21031 Hamburg, Germany

but from processed wood such as dried and oven heated sawn wood and glued wood like window frames and other products (Asif and Channon 2005; Rachmayanti et al. 2009). In this process, the minimal test specimen size for DNA extraction needs to be 1 cm^3 (Schmitz et al. 2019). To analyze charcoal, specific anatomy and NIRS can be used (Nisgoski et al. 2021; Zemke et al. 2020). In contrast to genus/species identification of solid wood, bleached pulps and papers present an even bigger challenge. Throughout the entire manufacturing process – from pulping to the various bleaching stages – many wood constituents are removed and the original DNA is destroyed by the chemicals. NIRS is also influenced by surface roughness, fiber orientation, sample thickness and, most importantly, the number of minor compounds for analysis is limited to a minimum mass fraction of 0.1–0.5 % (Schwanninger et al. 2011). This explains why anatomy was the singular method possible for pulp, paper and fiberboard up to now. Challenging this view, this publication shows how thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS) can provide an even more detailed method for species identification while also providing solutions for some of the difficulties of the anatomical method.

Although anatomical identification via morphological characteristics of wood vessels has already been established (Helmling et al. 2016; Helmling et al. 2018; Ilvessalo-Pfäffli 1995), this method reaches a limit when fibers are modified by refining, a common process in papermaking, which destroys the remaining vessel elements (Helmling et al. 2018). In addition to that, there are wood genera that have only a few distinctive anatomical structural features and can be confused with other genera – even with non-closely related woods (Gasson 2011). A particularly problematic example is the genus *Gonystylus* (Ramin), which is strictly protected under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES Appendix II) (CITES 02/2023). In order to have another independent method complementary to the anatomical method for pulp and paper, the chemotaxonomic method was developed. A major advantage is that this method is not impaired by the mechanical transformation of anatomical features, such as the refining of fibers.

Overall, chemotaxonomy or fingerprint analysis can be used to classify and identify plants based on their secondary plant compounds (Alaerts et al. 2014). The potential of chemotaxonomy in CITES enforcement using different MS techniques was demonstrated many times before. For example, Kite et al. (2010), analyzed solid heartwood of

various *Dalbergia* species successfully using LC-MS. Further, Lancaster and Espinoza (2012) showed how DART-TOFMS coupled with the multivariate data analysis methods PCA and LDL can chemotaxonomically distinguish different solid wood *Dalbergia* species. Cody et al. (2012) classified White Oak and Northern Red Oak with a success rate of 100 % while this rate can vary when used on closely related species (Deklerck et al. 2019). Moreover, Chemotaxonomy using DART-TOFMS can be used to distinguish wild from cultivated solid *Aquilaria* spp. (Espinoza et al. 2014). This shows that the DART-TOFMS method is highly beneficial for solid wood analysis (Dormont et al. 2015; Low et al. 2022). However, in relation to genetics and anatomy it still has a relatively small albeit growing body of collected reference spectra. However, what must be considered is that while previous studies concentrated on solid wood, the chemotaxonomic studies of this paper were performed on the pulp and paper deriving from tropical wood species. A high degree of care is applied during the industrial pulping and bleaching process in order to remove wood extractives as far as possible, without risking poorer paper quality, still ensuring a faultless production. Because this results in low extractive content, it was not clear whether wood species in pulp could be identified by chemotaxonomic methods at the start of this study. Also, the composition of the pulp extractives differs from that of the original solid wood extractives. For this reason, information on chemotaxonomically relevant substances in solid woods, as described by Hegnauer (1986), cannot simply be relied upon for pulp. Nevertheless, this investigation showed that extractives suitable for chemotaxonomic identification can be extracted from pulp in sufficient quantities by n-hexane. In the end, TD-GC-MS analysis of these reference extracts was performed, specific marker substances were found, and a fingerprint database was built up using specialized software.

2 Materials and methods

For the method development, a purchased industrial MTH pulp from Sumatra served as the initial material for the development of suitable grinding and extraction parameters for MTH pulps. Subsequently, 38 single variety solid wood samples of approximately 2 kg each (Table 2) were obtained from different, mostly commercial sources (no documented origin) and generated the reference database of the chemotaxonomic study. Genera and species of these solid wood samples were morphologically identified by wood anatomists with the assistance of the Xylothek at the Thünen Institute, Hamburg, Germany. Then, each individual reference wood was digested and bleached to produce respective kraft pulps for further analysis within the database created.

2.1 Production of the reference pulps

To start with, the solid wood samples were manually cut into 3 x 3 cm wood chips with a self-built semi-automatic chipping machine. Afterwards, they were steamed with saturated steam at 1 bar prior to pulping for 30 min. The kraft pulping was carried out in a program-controlled 7-L M/K digester (M/K Systems INC., Williamstown, USA) with liquor circulation. Depending on the quantity of the raw material, between 550 and 1000 g dry matter wood chips were used per cooking. The liquor to wood ratio was 4:1 (v/w). Total NaOH chemical usage was 22–25 % with a sulfidity (Na_2S) of 35 %. The heating time (t) was 90 min followed by 120 min at a maximum temperature (T_{max}) of 165 °C. The obtained kraft pulps were slot screened to a maximum of 0.15 mm and subsequently bleached in five bleaching stages, adapted from the industrial bleaching practice of the Mercer Stendal pulp mill. The target brightness was 90 % ISO. The oxygen stage (O) was carried out with 12 % consistency (c), 2.5–4.0 % NaOH content and 0.4 % MgSO_4 content for 120 min at 98 °C. The complexing agent stage (Q) was carried out with 3 % consistency and 0.2 % DTPA content for 30 min at 60 °C. The oxygen-enhanced peroxide stage (OP) was carried out with 12 % consistency, 2 % NaOH content, 0.1 % MgSO_4 content and 0.05 % DTPA content for 120 min at 95 °C. The chlorine dioxide stage (D) was carried out with 10 % consistency, 2–2.5 % ClO_2 content, 0.1 % DTPA for 180 min at 70 °C. Lastly, the peroxide stage (P) was carried out with 10 % consistency, 1.25–1.75 % NaOH, 0.1 % MgSO_4 content and 1–2% H_2O_2 content for 120 min at 80 °C.

The described kraft pulping process was based on the process steps and parameters commonly used in the industry. The fibers of the pulps produced are therefore similar to the fibers of industrially produced ECF pulps. A short discussion on the chosen pulping and bleaching parameters is included.

2.2 Grinding

The pulp was ground in a programmed and liquid nitrogen cooled ball mill (CryoMill, Retsch GmbH, Haan, Germany) with a frequency of 25 Hz using a 50 ml stainless steel grinding jar and one grinding ball of 25 mm diameter. Each pulp sample was ground five times for 1 min. Prior to the first grinding cycle, the grinding jar including the pulp was automatically cooled for 2 min. Cooling was applied for 0.5 min between all grinding cycles. The knife mill (M20 universal mill, IKA-Werke GmbH & CO. kg, Staufen, Germany) with a rotating vane blade was used with a speed of 20 000 1/min. The samples were ground three times for 30 s.

2.3 Extraction

The solvents were freshly distilled by rectification from both technical quality *n*-hexane and acetone (VWR International, LLC., Radnor, USA). The quality was analyzed via GC-MS. It is identical to GC quality *n*-hexane and PESTINORM acetone from VWR (VWR International, LLC., Radnor, USA). The extraction was performed by a Soxtherm SOX 6 from the company C. Gerhardt GmbH (Königswinter, Germany), run with the automatic MULTIISTAT controlling system. The cellulose extraction thimbles and the viscose wadding used to plug the thimbles were pre-washed by a short Soxhlet extraction of 5–6 cycles for 60 min

using a self-made azeotrope of *n*-hexane and acetone. The ground pulps of 5 g per extraction thimble were extracted with 140 ml of solvent with the automated self-programmed *n*-hexane extraction program: T-Class 200 °C; hotplate temperature 180 °C; setback interval 3.5 min; setback pulse 2 s; boiling phase 45 min; distilling off A, four intervals; extraction time 3 h; distilling off B, one interval; general distilling off interval time 5 min; total time 4 h 7 min. For acetone the extraction program was equal apart from the setback interval, which was 4.5 min. After extraction, the extracts were adjusted to a volume of 50 ml in volumetric flasks.

2.4 Sample preparation

The extracts were characterized by TD-GC-MS. Approximately $90 \pm 5 \mu\text{g}$ of extract was applied into each sample cup by adding the extract solutions. Therefore, a defined volume of 50 μl of the solvents was gently evaporated by being placed under the fume cupboard at room temperature covered with a paper sheet to avoid dust pollution. Afterwards, the concentrations in the extracts were measured and sample cups filled up multiple times until the target mass was reached. After that, the sample cups were introduced into the pyrolysis system for TD-GC-MS analysis.

2.5 TD-GC-MS parameters

The TD-GC-MS analysis was performed using a micro furnace Double-Shot Pyrolyzer (Py-2020iD) equipped with an Auto-Shot Sampler (AS-1020E) both produced by Frontier Laboratories Ltd (Koriyama, Japan). The pyrolysis system was interfaced to a GC-MS (6890/5973N, Agilent Technologies Inc., Santa Clara, USA). The TD temperature for the dried and solid extractives was 325 °C and the interface was set to 330 °C. For pyrolysis of low-density polyethylene (LD-PE) retention index standards the pyrolyzer was set at 500 °C. The GC inlet and the GC-MS interface temperature were kept at 320 °C. A low polarity column (ZB-5HT, Phenomenex Inc., Torrance, USA) of 30 m x 0.25 mm i.d. and 0.25 μm film thickness was used with helium as carrier gas. The split ratio was set to 20:1. A flow rate of 1 ml/min (constant flow) was set for gas chromatographic separation. The signal data rate was 20 Hz, scan frequency was 2.22 scans/second and scan speed (u/s) was 1562 [$N = 2$].

The oven temperature of the GC started at 45 °C, which was held for 2 min and subsequently increased up to 340 °C with decreasing heating rates. This temperature was kept constant for 30 min. The exact configuration of the GC oven temperatures is listed in Table 1 and visualized in Figure 3. The total run time was 134.17 min. For mass spectral detection an 5973N MSD (Agilent Technologies Inc., Santa Clara, USA) was used with electron impact ionization energy of 70 eV. The scanning range for measurement in total ion current (TIC) mode was 29–700 m/z with a threshold of 100.

2.6 Data evaluation

Every dried pulp extract was analyzed at least twice. The multiple chromatograms per extract were visually compared in an overlay before checking its individual peaks. In case of unusually missing or additional peaks, these were screened and its reasons, such as a small

Table 1: GC oven temperature ramps.

Ramp	Rate (°C/min)	Final temperature (°C)	Hold time (min)
Initial		45	2.0
1	10.0	100	
2	5.0	180	
3	2.0	280	
4	1.5	320	
5	5.0	340	30.0

air infiltration during the GC measurement for instance, were analyzed. Those samples were prepared again and measured twice to make sure the GC-MS measurement worked without any disturbances and that the chromatograms were equal. One of the representative chromatograms was further used. The data sets were pretreated and put together in a database of single-variety pulp extract reference chromatograms with the continuously developing database software OpenChrom® (version: 1.4.0.202104211010, Lablicate GmbH, Hamburg, Germany) (Wenig and Odermatt 2010a,b) and its ChromIdent® database tool. Compound identification of single peaks was performed using the NIST Mass Spectral Library (NIST20) (National Institute of Standards and Technology, USA).

Based on the fact that developing a new methodology is significant to this paper and therefore, its result, all key aspects of the methodology development are placed in Section 3 (Results and discussion). For instance, the outcome of comparing different mills for pulp milling and the discussion of suitable solvents for ground pulp extraction are part of the technique development process. Subsequently, the following chapter presents these alongside achieved bleached pulp brightnesses for different genera, their extract amount yield, the pulp extract chromatography data pretreatment and setting up/functioning of the reference chromatogram database.

3 Results and discussion

Supporting the development of the presented general sample preparation method, a broadly MTH kraft pulp of Asian origin seemed most appropriate. The industrial MTH pulp was described in Section 2 (Materials and methods). MTH is a raw material segment with a composition that is subject to wide fluctuations. It can also be defined as a more or less natural mixture of tropical hardwood types. For example, in some countries MTH consists of 15–20 species, but in others MTH can contain up to 100 species (Grützmann 2013). As a result, the densities of mixed tropical hardwoods can vary from 700 kg/m³ to 800 kg/m³. Considering individual species within a forest area, the variation increases even further and can vary from 150 kg/m³ to 1300 kg/m³ (Grützmann 2013). As a kind of by-product, MTH is largely produced during large-scale conversions of natural forests in the tropics into palm oil and timber plantations (Broich et al. 2011).

To establish the pulp extract database, the anatomically determined woods underwent kraft pulping and bleaching processes. According to the current FAO pulp and paper capacity report (FAO 2022) the global capacity of chemical paper grade pulp is 2.2 MT of sulphite pulp compared to 125 MT of kraft pulp. With the kraft pulp procedure, the dominant share of the market is covered at this point of the study. A combination involving oxygen, peroxide, chlorine dioxide, and complexing agents for bleaching kraft pulps appears most relevant. While on the one hand, the specific reaction conditions vary due to factors like different kinds of raw material, mill design and philosophy, on the other hand, the overall exposure to pH and chemicals of extractives, remains relatively consistent. Although bleaching conditions vary, the authors believe that they will not alter the fingerprint of extractives. Therefore, sensible choices for the study's starting point were made.

3.1 Choice of wood species

The tropical forests of Southeast Asia are characterized by high species diversity among the tree species growing there. In cooperation with the NGOs Greenpeace and World Wide Fund For Nature (WWF), a list of 38 particularly relevant and interesting wood species was compiled (Table 2).

The pulp industry likes to use cheap and locally available woods. Therefore, the selection of relevant woods for the reference database was also based on the distribution and availability for the industry and correspondingly good procurement opportunities for the pulp industry. Some of the woods originate from plantations, for example *Hevea brasiliensis* or *Paulownia tomentosa*. They also find use as pulp as they are often co-products.

The list also includes *Acacia mangium* and *Eucalyptus globulis*, which are often legally used for pulp. These serve as a proof that the methods developed in this study can be applied to industrially produced pulps. In addition, these pulps are often found in papers and must be distinguished from tropical woods. The palms *Cocos nucifera* and oil palm *Elaeis guineensis* were included in this study because they can also be used for pulp production and it is anticipated that these palms, which are currently grown in plantations on former rainforest areas, will be increasingly cleared and utilized for pulp production (Onuorah E. O. et al. 2015; Wan Daud and Law 2011; Welling and Liese 2019). Table 2 also shows the achieved ISO brightness and extractive content of every samples bleached pulp.

Table 2: Wood taxa/species and pulp properties. The extractive content means only the *n*-hexane extractives.

No.	Wood taxon	Trade name	Bleached pulp	
			Brightness (% ISO)	Extractive content (%)
1	<i>Acacia mangium</i> Willd.	Acacia	90.9	0.52
2	<i>Alniphyllum pterospermum</i> Matsum.	Mee Dong	90.1	0.20
3	<i>Avicennia marina</i> (Forssk.) Vierh.	Api Api	86.5	0.52
4	<i>Calophyllum</i> spp.	Bintangor	88.9	0.23
5	<i>Canarium</i> spp.	Kedondong	90.2	0.14
6	<i>Castanopsis argentea</i> (Blume) A. DC.	Berangan	83.1	0.58
7	<i>Cocos nucifera</i> L.	Coconut palm	89.3	0.24
8	<i>Cunninghamia lanceolata</i> (Lamb.) Hook.	Chinese fir	82.9	0.63
9	<i>Dendrocalamus latiflorus</i> Munro	Bamboo	86.9	0.24
10	<i>Dipterocarpus</i> spp.	Keruing	85.5	1.19
11	<i>Durio</i> spp.	Durian	89.1	0.17
12	<i>Elaeis guineensis</i> Jacq.	Oil palm	89.6	0.06
13	<i>Eucalyptus globulis</i> Labill.	Eucalyptus	90.1	0.73
14	<i>Fagus sylvatica</i> L.	Beech	89.5	0.04
15	<i>Gonystulus</i> spp.	Ramin	89.0	0.16
16	<i>Heritiera</i> spp.	Mengkulang	87.8	0.07
17	<i>Hevea brasiliensis</i> (Willd. ex A. Juss.) Müll. Arg.	Rubberwood	88.7	0.16
18	<i>Ilex triflora</i> var. <i>kanehirai</i> (Yamamoto) S. Y. Hu	Holly, Kecemang	89.2	0.20
19	<i>Intsia</i> spp.	Merbau	86.8	0.34
20	<i>Koompassia malaccensis</i> Maingay ex Benth.	Kempas	84.5	0.29
21	<i>Lophopetalum</i> spp.	Perupok	89.3	0.11
22	<i>Mangifera</i> spp.	Ambacang, Mango	89.7	0.45
23	<i>Nyssa javanica</i> (Blume) Wangerin	Tupelo, Nyssa	88.8	0.09
24	<i>Palaquium</i> sp.	Niato/Suntai	87.9	0.30
25	<i>Parashorea</i> spp.	Gerutu	82.1	0.34
26	<i>Paulownia tomentosa</i> (Thunb.) Steud.	Paulownia	87.3	0.48
27	<i>Phellodendron</i> sp.	Amur Cork tree	88.4	0.48
28	<i>Pterygota</i> sp.	Koto	91.1	0.04
29	<i>Rhizophora</i> spp.	Red Mangrove	85.8	0.23
30	<i>Schima superba</i> Gardn. & Champ.	Samak, Puspa	86.4	0.13
31	<i>Shorea</i> subg. <i>Anthoshorea</i>	White Meranti	89.4	0.20
32	<i>Shorea</i> subg. <i>Richetia</i>	Yellow Meranti	89.1	0.19
33	<i>Shorea</i> subg. <i>Rubroshorea</i>	Dark/light Red M.	85.0	0.32
34	<i>Shorea</i> subg. <i>Shorea</i>	Bangkirai, Balau	88.6	0.28
35	<i>Swintonia</i> spp.	Merpauh	90.8	0.25
36	<i>Tectona grandis</i> L.F.	Teak	82.8	0.10
37	<i>Terminalia tomentosa</i> Willd.	Limba	90.4	0.14
38	<i>Tetramerista glabra</i> Miq.	Punah	83.6	1.01

According to the authors information, whole debarked logs are used in MTH pulp production and not sawmill residues. The authors therefore assume that heartwood always feeds into pulp production in large proportions and thus the use of heartwood as reference material was decided. Origins are not given in Table 2 since 2 kg of solid wood material were required for pulp production, which was not possible to obtain with a documented origin for most of the woods.

3.2 Grinding

For the selection of a suitable mill, the knife mill with rotating vane blade and the cryo ball mill were compared for pretreatment of MTH pulp before the extraction. Differences between the results may indicate that changes occur in the pulp during grinding. This could be caused not only by the different grinding techniques, but also by the temperature differences within the mills. The ball mill is cooled with

liquid nitrogen ($-196\text{ }^{\circ}\text{C}$) and produces a very fine powder, which provides a large surface area and thus good accessibility for the solvent (Figure 2c). It allows the solvent to flow uniformly through the sample and dissolve repeatable amounts of extractives, resulting in reproducible peak areas (Table 3). The knife mill, on the other hand, produces an inhomogeneous fluffy grind that contains both fibrous large particles (Figure 2b), but also very small powdery particles that are not visible in Figure 2b.

The differently ground MTH pulps were extracted with petroleum ether (six samples per mill of 5 g each) and analyzed both quantitatively, by gravimetric determination of the amount of extract, and qualitatively, by the detected number of peaks and peak areas in the TD-GC-MS chromatograms (Table 3). The quantitative results showed

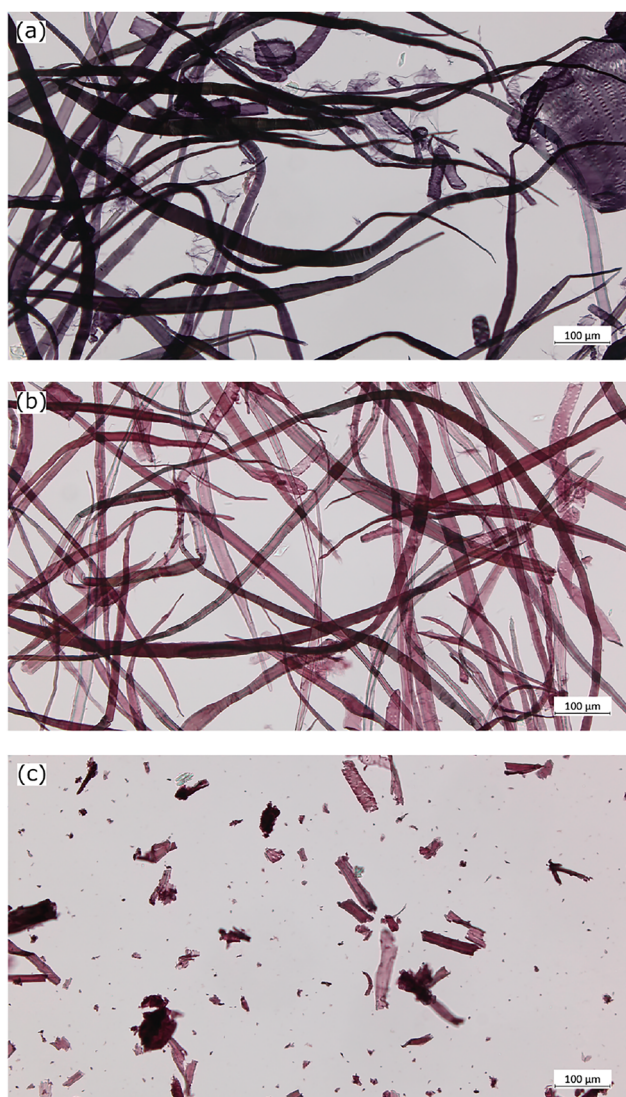


Figure 2: Microscopic images of MTH pulp: (a) raw; (b) knife mill ground; (c) cryo ball mill ground. Staining was done with Alexander Herzberg stain.

Table 3: Comparison of cryo ball mill and knife mill ground petroleum ether extracted MTH pulp. Peak area of 15 biggest peaks per chromatogram, determined by one chosen SIC area for every peak.

	Ball mill		Knife mill	
	Mean	CV (%)	Mean	CV (%)
Number of peaks	104.8	2.4	125.8	3.3
Peak area	2 891 103	12.8	2 571 478	30.1

slightly higher extract amounts of 0.026 % (0.2 % CV) after the first extraction step for knife mill ground pulp in comparison to 0.020 % (2.3 % CV) for ball mill ground pulp. Furthermore, the number of peaks were higher within the knife mill treated pulp. These chromatograms had a mean number of 125.8 peaks (variance 17.1) whereas the ball mill ground pulps had a mean of 104.8 peaks (variance 6.2). The ball mill provided at least twice as good reproducibility of the peak areas from the extract chromatograms: the coefficient of variation (CV) of the areas of the 15 biggest peaks of the ball mill extracts chromatograms is 0.128 in contrast to 0.301 for the knife mill. The higher extract quantity and peak numbers of the knife mill spoke at first sight for this mill. But for the following analytical evaluation and the purpose of this paper the GC-MS measurement precision and the reproducibility of the results are crucial. The function of the fingerprint database relies on reproducible peak areas and proportions. Therefore, the extracts chromatograms of the pulps ground with the ball mill are more suitable for this study. The clearly better reproducibility of the peak areas when using the cryo ball mill could be due to the finer grinding under cold conditions.

3.3 Extraction

The successive extraction was also optimized on basis of the MTH industrial pulp milled with the cryo ball mill. The extract quantity was maximized through the selection of suitable solvents, extraction time and number of extraction cycles. The extract quantities of the individual extraction steps were determined. The extraction process used a combination of nonpolar and polar solvents, including three times petroleum ether (6 samples) followed by two times acetone and stand-alone n-hexane (5 samples) to obtain both fractions of the extracts as exhaustively as possible, as shown by Ponnuchamy et al. (2021) for kraft lignin and Krogell et al. (2012) for Norway spruce bark. Because the third extraction step obtained hardly any measurable extracts (Table 4), this method effectively extracts ground

Table 4: Comparison of the successive extraction with petroleum ether (3 times) followed by 2 times acetone and a sole *n*-hexane extraction. Total and relative extract amounts are determined based on extraction of 20 g of ball milled MTH pulp.

Extraction step	Petroleum ether				Acetone (suc. after pet. ether)				<i>n</i> -Hexane			
	Amount (mg)	Amount (%)	SD (%)	CV (%)	Amount (mg)	Amount (%)	SD (%)	CV (%)	Amount (mg)	Amount (%)	SD (%)	CV (%)
1	4.05	0.020	0.0005	2	10.24	0.051	0.0046	9	4.12	0.021	0.0003	2
2	1.46	0.007	0.0006	8	3.06	0.015	0.0049	32	1.43	0.007	0.0005	7
3	0.77	0.004	0.0014	36								
Sum	6.28	0.031			13.29	0.066			5.55	0.028		

pulps exhaustively with two steps per solvent. Water was not used because it was assumed that water-soluble components were already discharged with the process water during pulping and bleaching. The results showed that most of the absolute extractive amounts were obtained in the first extraction step with either solvent. Table 4 also shows the relative extractives amount in percent and extract yields per solvent and extraction step.

Acetone was found to be quantitatively more effective in extracting pulps than hexane, but high yield does not imply specificity of the extractives. Additionally, the acetone extracts were difficult to measure by GC-MS without derivatization due to their polarity. These polar components can be found in paper making effluents (Björklund Jansson 2005; Holmberg 1999; Örsa and Holmbom 1994; Valto et al. 2012; Willför et al. 2006), meaning they are affected at least in their quantity by pulping and bleaching. It was assumed that acetone mainly extracts polar extractives and hexane the more non-polar, more interesting extractives. The non-polar extractives are least affected by pulping and bleaching, as the process takes place in polar media, as water is used to dissolve the applied chemicals. Therefore, the hexane soluble extractives are more valuable for the database. The focus was on developing a practical method. That is why it was decided to leave the acetone extracts aside and concentrate on the hydrophobic fraction of the extractives. Another reason for the pre-fractionation is reduction of the complexity of the chromatograms. However, one problem with using the solvent petroleum ether is that it is not clearly defined in terms of composition, leading to a risk that extraction conditions may not be identical when changing batches or suppliers. Moreover, the ground pulps were not only cool extracted but boiled in the solvent in the first step. Monitoring the petroleum ethers (boiling range 40–60 °C) boiling temperature during the extraction process showed a temperature rise over time. This supported the assumption that the cool extraction was not carried out with the entire solvent mixture, but only with the less volatile part, because the composition of the petroleum ether presumably changed over time in favor of the higher-boiling fractions. So,

petroleum ether is not perfectly suitable for extraction systems without pressure like Soxtherm or Soxhlet. In contrast, *n*-hexane is a well-defined solvent. The results of using *n*-hexane were found to be similar in terms of efficiency and reproducibility when compared to using petroleum ether (Table 4). Therefore, *n*-hexane was deemed the best choice for this purpose.

During the optimization process, sources of impurities for example in the extraction wadding, the extraction thimbles, the solvent or on glass surfaces and others were identified. They were largely eliminated by establishing a laboratory routine. These measures led to an improved quality of the pulp extracts with a significant reduction of impurities. A chromatogram of a blank extraction of empty and unwashed extraction thimbles and wadding with petroleum ether showing the impurities peaks (Supplementary Figure S5) and a list of identified compounds (Supplementary Table S40) were added to the supplementary material.

3.4 Optimization of TD-GC-MS chromatogram quality

Prior to development of data preprocessing systems, an optimization of the GC-MS analysis was a necessity. In this way, the quality of the chromatograms was increased by defining suitable GC-MS parameters such as the TD temperature of 325 °C and developing a fitting GC oven temperature program for the *n*-hexane pulp extracts. A better chromatographic separation of overlapping or sometimes called co-eluting peaks, which occasionally occur in one-dimensional gas chromatography (Blumberg 2012), was achieved especially of the mainly phytosterol peaks eluting around 280 °C in the range of RT 65–85 min (Figure 3). Their overlap was satisfactorily separated by the slow heating rates of 2.0 and 1.5 °C per minute. This was achieved by using polyethylene analyses as a neutral measure of the effect of the heating ramps. The heating ramps were adjusted so that the distances of the alkane and alkene peaks of the polyethylene were approximately equidistant over the relevant

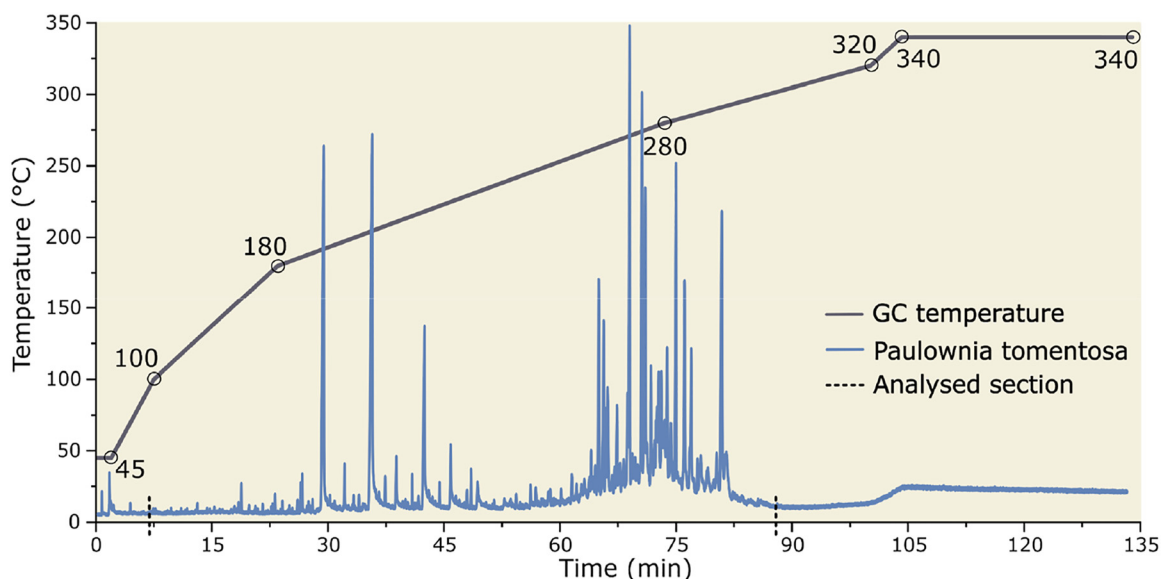


Figure 3: GC oven temperature program and chromatogram of *Paulownia tomentosa* (Thunb.) Steud. pulp extract before preprocessing. For the database references the chromatogram section of RT 7–88 min was analyzed.

chromatogram range instead of increasingly narrow. Thus, a better capture of the information hidden behind coelution in this RT range of the chromatograms was achieved. In order to prevent artefacts of the high-boiling extract components in the following measurements, the final bake-out time needs to be 30 min long. Still the overall measurement time per extract of less than 140 min wasn't unnecessarily extended. The optimized oven temperature program is shown in Figure 3.

3.5 Preprocessing of GC-MS data

The extractable constituents of the pulps were largely captured by the described conscientious sample preparation. For the following analysis they are available in the form of GC-MS chromatogram data sets. With the help of data preprocessing steps, it was possible to record the chromatographed analytes, i.e. their chemotaxonomic characteristics, and store them in a database. Appropriate pretreatment of the data eliminated or compensated artifacts. The quality of the chromatograms was thus improved and for each peak in the chromatogram, which in turn represents a component of the complex extractive mixture, characteristic properties such as percental peak area and retention index can be determined more reliably. The following nine most important preprocessing steps were found to be crucial for the pulp extract data matrix and the purpose of a subsequent fingerprint database. They need to be performed on the raw GC-MS data files for both references for the

database and unknown pulp/paper samples for comparing against the database.

3.5.1 Rounding of detected mass traces

Each mass trace (m/z) is rounded to nominal masses (Khristanfov and Samokhin 2022). Because the mass axis in the used GC-MS system tends to shift upwards, rounding of decimal places was performed from inclusive -0.3 to exclusive $+0.7$. Whole numbers are important for later trouble-free smoothing, which in turn is important for the peak detection algorithm.

3.5.2 Deleting empty scans

Empty MS scans are those that don't contain any data. They are removed to shrink the dataset size and to speed up the data processing time.

3.5.3 Selecting chromatogram range

The beginning and the end of the recorded raw chromatograms are cut. They are not of interest for this analysis because the front part consists of irrelevant peaks, which don't belong to the pulp extracts but for example arise from solvent residues or the atmosphere. The back part of the chromatogram is also discarded due to mainly noise and column bleeding. Also, to reduce the dataset size and speed up the processing the chromatogram section used for the database was selected from 7–88 min.

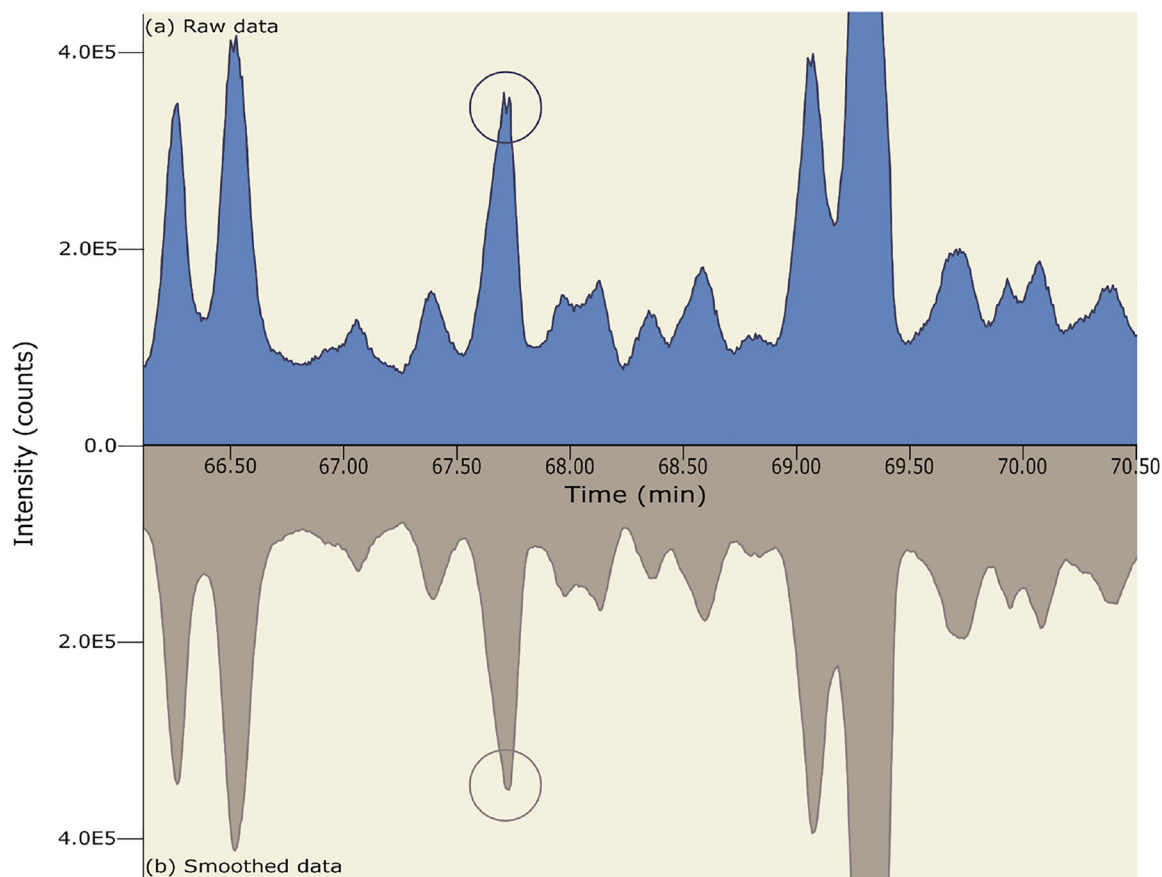


Figure 4: Savitzky–Golay smoothing: TIC chromatogram sections from database reference sample *Paulownia tomentosa* (Thunb.) Steud., RT 66.10–70.50 min: (a) raw data versus (b) smoothed data.

3.5.4 Removing unwanted ions

Mass traces, that do not (only) belong to the pulp extracts but (also) to the atmosphere or the GC column should be avoided. The ions 18, 28, 32, 44 result from the atmospheres gases (water, nitrogen, oxygen, carbon dioxide). 207 and 281 are attributed to column bleed, the loss/decomposition of the stationary phase, at elevated temperatures. Therefore, they are removed from the whole dataset. Also, the mass trace 84 m/z is removed as it occurs regularly and often and is not specific to pulp extractives at all.

3.5.5 Smoothing chromatograms

For a well working peak detection and clean peak integration the whole chromatograms were smoothed. For this purpose, the chromatogram filter Savitzky-Golay smoothing with the settings order = 2 and width = 5 was applied. This filter algorithm was developed by Savitzky and Golay (1964). It is used to smoothen single ion signals or TIC signals and remove electric noise (Wenig 2011). The smoothing was done

per ion, meaning every single m/z was smoothed separately, which is important for the following peak detection via deconvolution. Figure 4 shows how smoothing the raw m/z data also affects the peak shape in the TIC chromatogram. One can see how the blue raw chromatogram in Figure 4a looks unsmoothed in comparison to the smoothed brown TIC signal underneath (Figure 4b). For instance, the three mid-sized peaks in the center (RT 66.5, 67.7, 69.1 min) show rough detector signals and therefore tiny double peaks, which are removed in the smoothed data set below. Choppy signals are risky because they can be falsely detected as multiple peaks on both the ion level and the TIC level.

3.5.6 Deconvoluting/detecting peaks

Multivariate curve resolution by alternate regression (MCR-AR) according to Gerber et al. (2012) was used to decompose the multivariate GC-MS data into individual pure component spectra and to detect the peaks. MCR-AR enables the detection of hidden peaks underneath the baseline noise or other peaks via their single ion traces. Thus, the underlying and

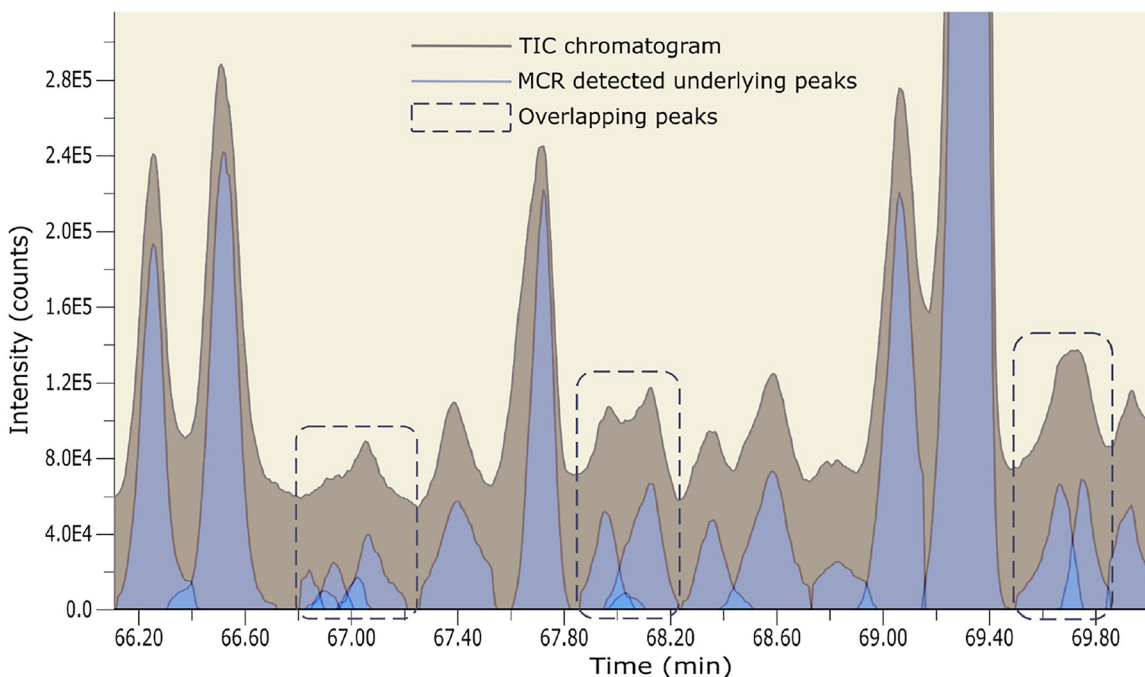


Figure 5: Peak detection using deconvolution (MCR-AR): Section from smoothed database reference sample *Paulownia tomentosa* (Thunb.) Steud., RT 66.11–70.03 min.

overlapping peaks can be detected individually as shown in Figure 5. MCR involves a segmentation of chromatograms into non-overlapping minimum 50-scan windows using local minima/maxima. Mass channels are baseline-corrected within these windows by linear interpolation and area subtraction. Each window forms a data cube. This data cube is then unfolded into a matrix $(N \times K) \times L$, where N represents the number of samples, K represents scans, and L represents mass channels. The decomposition obtained from the initial MCR step is used as a starting point for AR. It alternates between deconvoluting chromatographic and mass spectral profiles until convergence. It starts by assuming a single distinct compound (rank = 1) and applies constraints like non-negativity to the solution. The pure component spectra are used as constraints in the regression process. The data is iteratively reprocessed ('MCR-AR max iteration': 50.0), by incrementally increasing the rank by one and calculating solutions with each iteration refining the decomposition. It continues until a pre-defined stopping criterion is met, indicating that the decomposition has reached a stable state. This means that the algorithm has found a solution that best represents the underlying pure components in the mixture (Gerber et al. 2012). This peak detection is very useful for the creation of the fingerprint database, as the complex wood extract samples often contain many similar substances that elute very close to each other in time. The overlaps cannot be completely separated chromatographically and sometimes appear as

"mountains" in the chromatogram (Figure 3, RT 65–85 min). In the following some MCR-AR settings, that were used for best results, are given and explained. The 'Local Maxima/Minima' classifier uses an algorithm that follows the signal from left to right to find local maxima/minima based on the first derivative/slope of the signal. Although this classifier does not search for chromatographic peaks with start/maximum/stop, some kind of start point must still be found internally, to be able to evaluate the parameter 'Signal difference threshold between start and extrema', which was set to 1000.0. To set a starting point it is required that the slope exceeds a certain threshold ('Local Maxima Scan Slope Threshold': 0.001). Then the 'Local Maxima/Minima Consecutive Scans' parameter comes into play: this defines how many scans must remain above this threshold after it has been exceeded for a start point to be set. This helps to filter out the smallest local maxima in the signal noise range. It was set to 1.0. All detailed MCR-AR parameter adjustments including explanations can be found in Supplementary Table S39.

3.5.6.1 Integrating peaks

To determine their peak areas the MCR-AR detected peaks were integrated with the trapezoid peak integrator. The deconvoluted TIC signal was integrated excluding the background. The peak areas were chosen to be put into the database as one of the comparison features rather than the intensity because the intensity and peak shape can vary

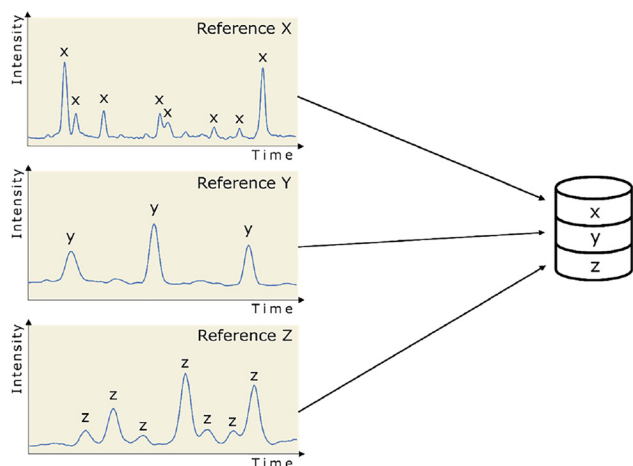


Figure 6: Schematic illustration of the database approach.

whereas the peak areas and proportions to each other, which are used for fingerprinting, stay the same. The peak areas are normalized so that only a percentage value is used. Since this is not a quantitative but a qualitative identification method, the absolute content is not decisive here, thus only the ratios need to be repeatable.

3.5.6.2 High pass ions and peaks

The high pass filters keep all ions or peaks with the biggest intensity or area. For this database the 400 biggest peaks by area per reference chromatogram worked the best and were kept for further analysis. Of every of those peaks, the 150 ions with highest signal intensities were kept. Keeping more than 150 ion traces didn't approve the database performance on random pulp mixtures because in the mixture the amount of every pulp is lower. Therefore, the extracts peaks are smaller and consequently due to the threshold settings the smallest single ions are not detected by the MS anyway.

3.5.6.3 Retention index (RI) calculation

The retention index (RI) is a measure used to characterize the retention time of a compound relative to the elution times of reference compounds. It is dimensionless and represents the relative affinity of a compound for the stationary phase compared to the reference compounds, meaning it converts retention times into system independent constants. An internal RI value is calculated for each peak maximum with LD-PE as standard reference, which is separated under the same conditions as the sample of interest. The RI are determined by relative distances of the compounds of interest to the LD-PE alkene peaks in the chromatogram: Retention Index $RI = 100 \times [n + (N - n) \times (RT(\text{unknown}) - RT(n)) / (RT(N) - RT(n))]$ where n is the carbon number of the n -alkene, $N = n + 1$ and RT = retention time. In every GC-MS

sequences LD-PE RI standards are measured minimum after every 10th measurement. This is how a system-independent index value without time unit could be given, which is not influenced by any kind of RT shift due to column abrasion, cutbacks of the column or a changed column/measurement program. It is very important to calculate the RI for every peak because the RT shift can even on the same GC-MS system be substantial with longer times between measurements and therefore influences the matching accuracy of the database massively if no RI are used.

3.6 Functioning of the database

The pretreated chromatograms contain several pieces of information important for the functioning of the database: deconvoluted peaks consisting of the individual mass traces, their peak areas and RI. These are associated with the peaks. For each reference extract of a pulp produced from a single wood species, the above information results in an individual fingerprint of contained compounds. Although most of the substances are not wood species-specific marker substances, the complex combination of substances with their proportional ratios to each other is unique. This fingerprint information is fed into the database (Wenig and Odermatt 2010a,b), as shown in Figure 6.

When adding a reference chromatogram to the database, each peak within a fixed RI window of ± 10 is compared with all other reference peaks in the same frame. In contrast to the NIST library peak comparison for compound identification by Stein and Scott (1994) for the fingerprint matching the cosine algorithm by Alfassi (2004) is used. The peaks are basically compared via their mass spectral match quality. Using retention indices is essential, otherwise the rate of false positive matches rises due to the similarity of wood extractive peaks, especially of alkanes due to their similar mass fragmentation pattern. Only if the peaks from different reference extracts match 80% or better (Match Factor 80+), they are rated as the same compound and merged and added collectively to the database as **one** library peak. This combined library peak links to all the corresponding references where it originates from. If peaks at the same RT/RI match with a smaller percentage than 80, they are rated to be different peaks and put into the database separately as individual library peaks. The matching parameter was set to a high value of greater than MF 80 because the wood samples contain confusingly similar components.

When matching an unknown mixed species pulp extracts chromatogram for identification against the database, the first step again is to align the unknown chromatogram

with the reference chromatograms in the database using an RI corridor of ± 10 . After that searching for matches in that RI corridor using the cosine algorithm starts. Also, peak areas and ratios (fingerprinting) are compared to find the best matches. The results, based on the similarity between the unknown and reference mass spectra, are ranked/sorted according to the scores, with higher scores indicating stronger matches. These MF scores are penalized by RI distance: the higher the RI (within the determined RI window) are lying apart from each other, the more the MF values are reduced. A minimum MF of 75 for every peak comparison of the unknown against the database reference peaks was chosen somewhat lower, since impure mass traces are expected due to overlaps within the mixed pulp extract sample. In addition, particular mass traces of wood species contained in small amounts in the mixed pulps may be below the mass spectrometers threshold intensity of 100. As soon as a peak of an unknown extract is identified against the database subset ($RI \pm 10$, $MF > 75$), the information is available if this peak occurs only in one or more references. A peak, that occurs only in one reference sample or group is a direct hit/marker peak (Figure 9). Therefore, a relatively high number of marker peaks, compared to the number of total matched peaks assigned to a specific database reference, is an important indicator in the identification of unknown samples. All other peaks are ambiguous hits. The mathematics behind it is quite simple: principally it's just counting the references. A direct hit counts 1. An ambiguous hit, which occurs in two references, counts with 0.5 and so on. The quantity of peaks and number of counting's provide a hint of the unknown extract. When evaluating the comparison results of an unknown sample against the database the reverse similarity index (RSI) and reverse match factor are most important. The unknown pulps are rarely pure. They are mostly mixed samples from different wood species. Therefore, the match against the database is tested in reverse to see how well the fingerprint peaks of the single pure database references match the peaks of the unknown sample mixture, i.e., whether every pure reference can be found in the unknown mixed sample. Since the mixed sample consists of several species, matching the whole unknown mixture sample forwards against each reference of the database results in lower MF, because the entire sample mixture can never be found in one pure reference sample. As an example: a mixture of X, Y and Z is identified by a database consisting of the references X, Y and Z. Reverse matching results: X is 100 % included in the mix, Y is 100 % included in the mix and so is Z. Forward matching results: the whole mixture matches 33 % with each X, Y and Z. The results of a database matching are mean values of the matching of all

individual peaks against each other. The individual peak matching can be represented as in the following Figure 7. Figure 7a shows the mass traces of an unknown deconvoluted peak of an unknown pulp extract mixture in the RT range of 67.50–68.00 min with its maximum at the RI of 3127 in comparison to the mass traces of the *P. tomentosa* database reference peak at the same RI in an inverted overlay format with the legend of all mass traces (m/z) on the bottom right. Figure 7b shows the comparison of the same *P. tomentosa* database reference peak in detail with all its mass traces not over an RT range but a single scan against the NIST20 library. The unknown extract substance is identified as (+)-Sesamin with an MF of 80.3 and a reverse MF of 87.2.

In order to represent the database graphically and to make the dimension of the differences of the individual references visible, a principal component analysis (PCA) was calculated from the database entries with the above-mentioned information. In the score plot of the PCA (Figure 8) the PC 1 (principal component 1) with 10.98 % on the x-axis and the PC 2 on the y-axis with 7.96 % represent only about 19 % of the total database information. This is due to the fact that the constituents of the wood species are often similar and a huge data cloud of more than 6000 peaks times 38 species is difficult to reduce in its complexity to two dimensions. In the 3-dimensional interactive score plot (PC 1, PC 2, PC 3) (Supplementary Figure S3), a total of about 26 % of the database information is represented and the distances are clearer than in 2-D.

Figure 8b shows an enlarged section of the score plot exposing the *Dipterocarpaceae* family. As expected, the close relationship between the *Dipterocarpaceae* family members in particular the *Shorea* subgenera is evident in the chromatograms – they cluster together in the PCA representation. Thus, when identifying one of these species in an unknown sample, a closer look must be taken to ensure that there is no confusion. In order to refine the identification of *Dipterocarpaceae* members, a separate database containing only these very similar species could be built in the future. Possibly, the matching parameters can then be chosen even more narrowly, and thus the distinctiveness of the matching results can be increased.

For each database reference chromatogram, there is a table of recorded library peaks including their RT, RI, peak area, NIST20 identified substances ($MF > 80$) and their 10 biggest ion traces as well as the other database references in which these peaks are also contained (Supplementary Tables S1–S38). The most decisive peaks are the specific marker peaks originating from only **one** reference. The whole database, including all original GC-MS data, can be

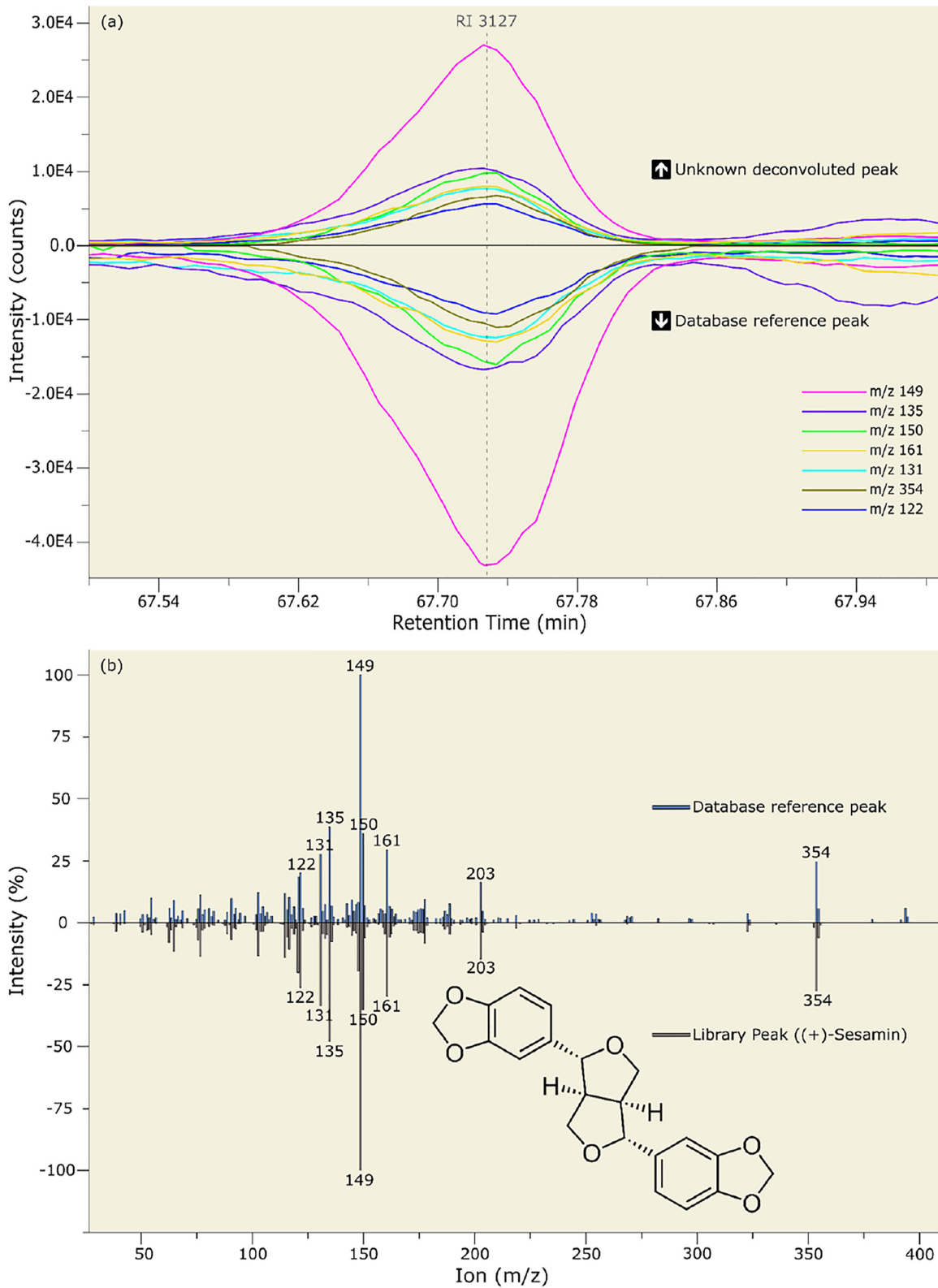


Figure 7: Single peak database comparison: (a) comparison of an unknown deconvoluted peak of an unknown pulp mixture extract against the database library peak from the *Paulownia tomentosa* (Thunb.) Steud. reference pulp extract at RI 3127; (b) comparison of the database reference peak against the NIST20 library peak of the substance (+)-Sesamin.

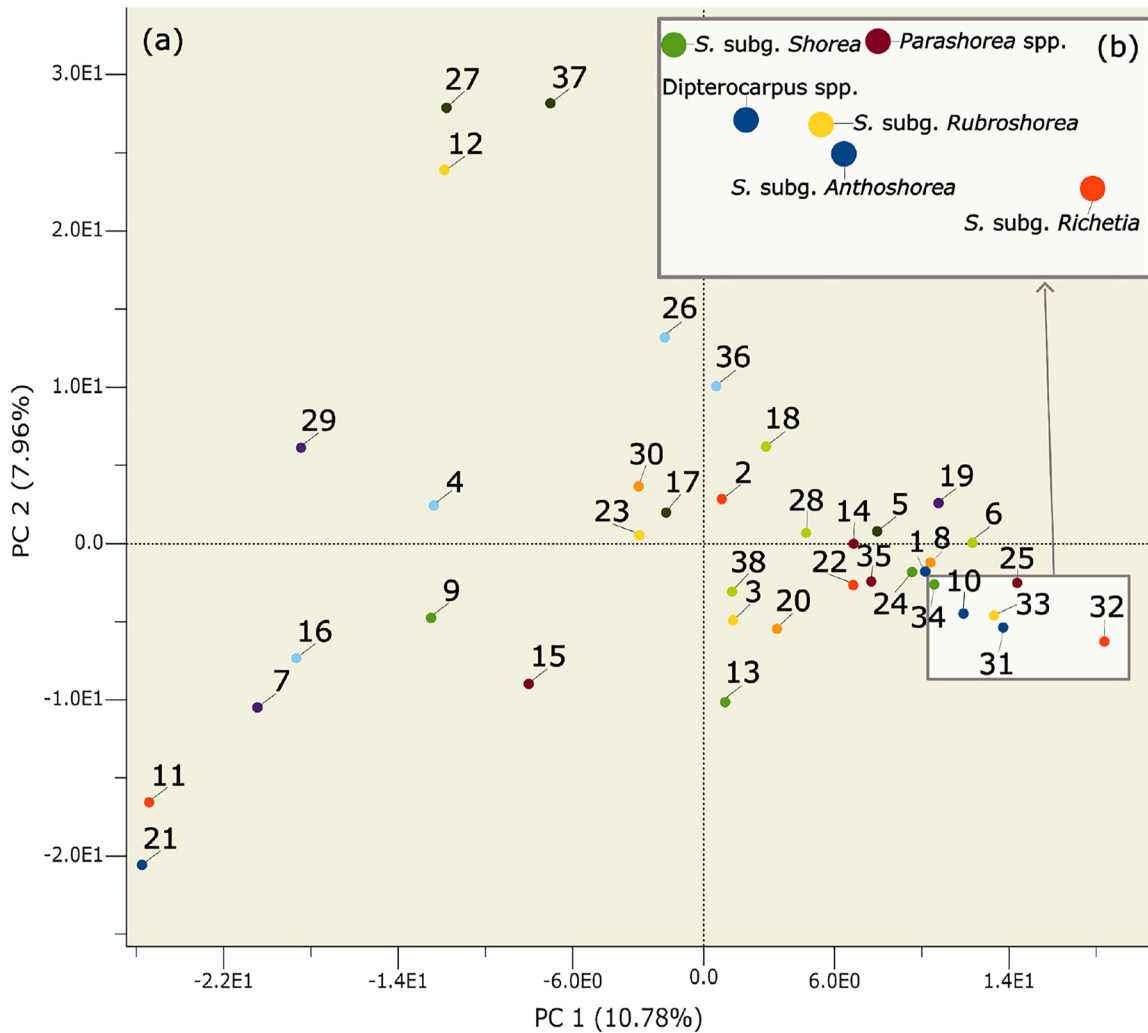


Figure 8: (a) PCA-score plot of the pulp extract database, numbered after Table 2; (b) enlarged section of the *Dipterocarpaceae* family.

found under Supplementary Figure S4. The reference chromatograms of *A. mangium* and *P. tomentosa* pulp extracts are also given as examples in Supplementary Figures S1 and S2.

The comparison of unknown paper extracts chromatograms against the database is supported by a visual comparison tool which shows the unknown chromatogram with unmatched, matched but ambiguous, meaning they are contained in more than one reference, and marker peaks (Figure 9).

When matching an unknown pulp extract chromatogram with the database, several comparison factors and statistics are crucial: the total RSI value, the number of marker peaks/ambiguous peaks and the total matching area which is calculated for every database reference compared to the unknown chromatogram. Ultimately, the observer must incorporate all of these comparison results/statistics

into the decision-making process. An example of an actual species identification in an unknown pulp mixture extract is shown in Figure 10, a screenshot of the software interface. It shows the database query results from a chromatogram of a mixed pulp extract against the reference database. For the database reference of *P. tomentosa* an RSI of 88.7 % is given, 58 marker and 141 ambiguous peaks are matched. Additionally, 10.4 % of the peak area of the unknown chromatogram was identified as *P. tomentosa*. Analyzing these clear results, the authors correctly concluded that *P. tomentosa* is contained in the mixture. *Tectona grandis* (15 matched marker peaks) and *Fagus sylvatica* (22 matched marker peaks) are also contained one-third each in the mixed sample and correctly identified. As the sample was randomly mixed by a colleague, the composition was unknown to the authors until the analysis and identification decisions were made.

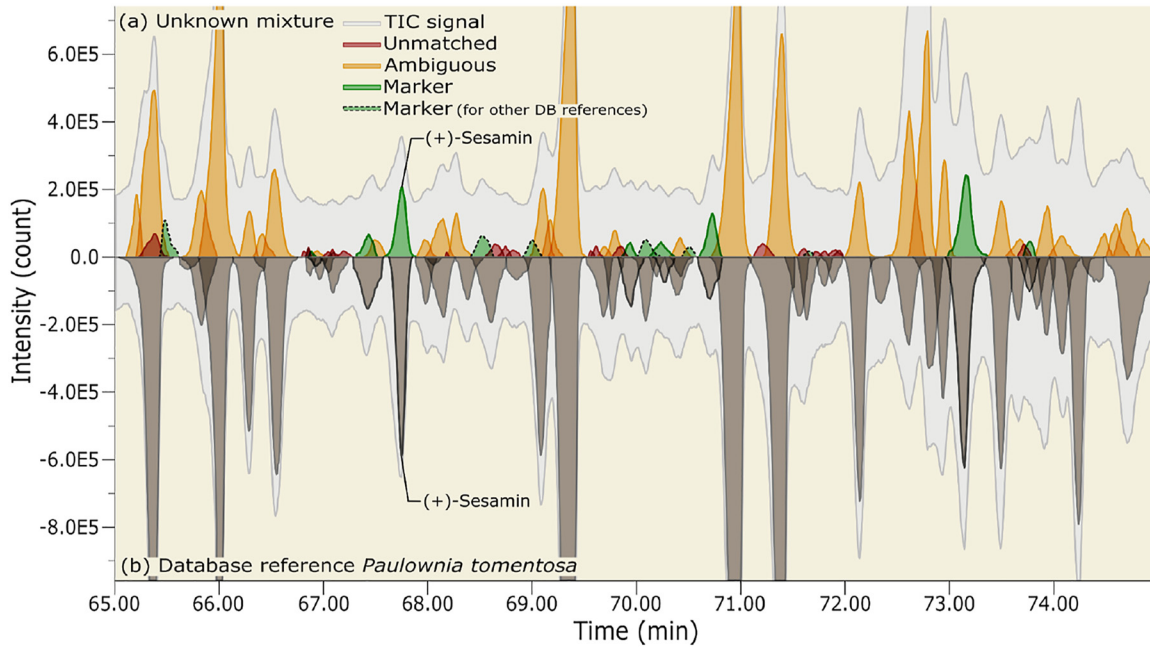


Figure 9: Comparison of an (a) unknown pulp mixture extract chromatogram with (b) the database reference chromatogram of *Paulownia tomentosa* (Thunb.) Steud.

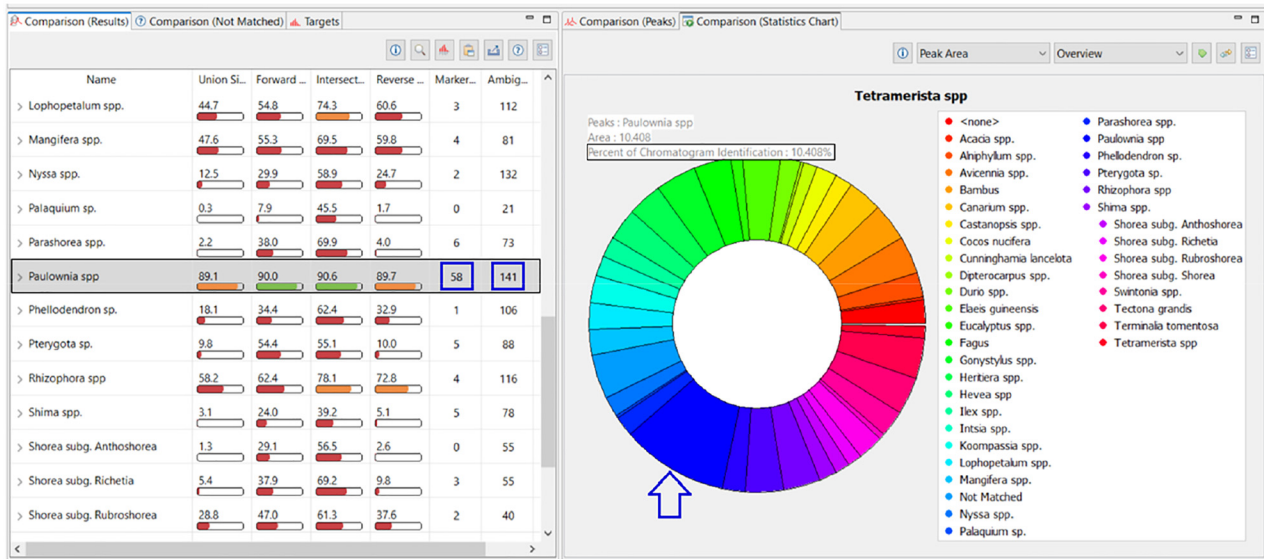


Figure 10: Database query results from a chromatogram of a mixed pulp extract (containing *Paulownia tomentosa*) against the reference database: screenshot of the software interface.

3.7 Limitations

The authors acknowledge that limitations exist in the studies ability to comprehensively represent all possible variations within tree species. The researchers have taken these limitations into account.

Due to the requirement of 2 kg of material for pulp production, it was not possible to obtain samples with documented origin for most of the woods. As a result, there could be variations within tree species/genera that the current database is not capable of identifying, particularly in cases where trees are grown under extreme conditions.

In conclusion, the inability to obtain samples with a documented origin for most of the wood genera/species, along with the practical constraints on the number of samples analyzed, constitute limitations of the study. They highlight the future need for expanding the database to account for variations in tree species grown in diverse environments.

4 Conclusions

The objective of this research was to develop a chemotaxonomic technique for identifying tropical wood species in pulp and paper using a chemical fingerprint database. As a result, it enabled the use of chemotaxonomy based on extractives and thus, a new method for extracting and analyzing the chemical components of wood pulps. The grinding and extraction methodology used was selected on the basis of two important criteria. Firstly, it should ensure a good extraction of pulps. This means that a large number of chemotaxonomically relevant extractives is obtained ensuring sufficiency for the next steps of the process. Secondly, the methodology should be simple and reproducible so that a large number of prospective samples can be processed quickly and reliably. With these criteria in mind, the combination of grinding with the cryo-ball mill and the extraction with the Soxtherm extractor using *n*-hexane was chosen. Although the extractable content did not give the highest quantity in all tested options, many extracts were captured and the reproducibility was of high standard. Moreover, such a combination is an advantageous application since its grinding and extraction processes are partly automated.

Furthermore, all extractives within this method are measured in a standardized manner with the same temperature program by TD-GC-MS. This ensures preprocessing the obtained data continuously in the same way using a standardized batch process. A database for pulp extract chromatograms has been built up from 38 reference chromatograms. When matching an unknown pulp extract chromatogram with the database, several factors such as the RSI value and the number of marker peaks as well as ambiguous peaks are crucial. It shall be added that the ChromIdent[®] software tool for matching unknown samples with the database will be further developed and improved in the future making this method even more reliable.

New wood species are constantly coming into focus and gain relevance due to increased use in paper production and in connection to new classifications of CITES protection statuses. Thus, the current chemotaxonomic *n*-hexane extractives database needs to be extended on a regular basis. The larger the reference database, the more powerful the

method will become. In perspective, its performance will be tested in a blind test of pulp samples with unknown composition and on other GC-MS systems to validate the method. Additionally, the potential of the acetone extracts will be included for increased resolution of critical species. Future investigations will also focus on chemotaxonomic variations among different provenances of wood. This is essential for the protection of natural forests, as no other method can distinguish between pulp from plantations and pulp from natural forests yet. In general, the database could even be extended to include wood products such as particle board, which is laborious for anatomists to work with, thus widening the usage of this method. All in all, although more work can and should develop this method even further, it already contributes to supporting the EUDR and sustainable forestry by adding new possibilities for its protection.

Acknowledgments: PD Dr. habil. Jürgen Odermatt, who initiated the research for this paper and passed away suddenly in 2019, deserves special thanks. The authors would also like to thank Birte Buske, who supported with laboratory work, as well as Othar Kordsachia, Nils Grützmann and Alina Wassink for their exploratory work on pulping.

Research ethics: As the research does not involve the use of humans or animals, the Declaration of Helsinki does not apply. Therefore, the research ethics statement is not applicable.

Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: The author states no conflict of interest.

Research funding: This work was funded by the German Environmental Foundation (DBU) (Grant number: AZ 34295/01) in connection with the project “Detection of Tropical Hardwood in Paper – Chemotaxonomy and Anatomy for the Identification of Mixed Tropical Hardwood”. The funding organization played no role in study design; analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

Data availability: The raw data can be obtained on request from the corresponding author.

References

- Alaerts, G., Pieters, S., Logie, H., van Erps, J., Merino-Arevalo, M., Dejaegher, B., Smeyers-Verbeke, J., and Vander Heyden, Y. (2014). Exploration and classification of chromatographic fingerprints as additional tool for identification and quality control of several *Artemisia* species. *J. Pharm. Biomed. Anal.* 95: 34–46.

- Alfassi, Z.B. (2004). On the normalization of a mass spectrum for comparison of two spectra. *J. Am. Soc. Mass Spectrom.* 15: 385–387.
- Asif, M.J. and Channon, C.H. (2005). DNA extraction from processed wood: a case study for the identification of an endangered timber species (*Gonystylus bancanus*). *Plant Mol. Biol. Rep.* 23: 185–192.
- Björklund Jansson, M. (2005). *Birch extractives in kraft pulp washing. STFI-Packforsk report no 141*. STFI-Packforsk, Stockholm, Sweden.
- Blumberg, L.M. (2012). Metrics of separation performance in chromatography. Part 2. Separation performance of a heating ramp in temperature-programmed gas chromatography. *J. Chromatogr. A* 1244: 148–160.
- Broich, M., Hansen, M.C., Potapov, P., Adusei, B., Lindquist, E., and Stehman, S.V. (2011). Time-series analysis of multi-resolution optical imagery for quantifying forest cover loss in Sumatra and Kalimantan, Indonesia. *Int. J. Appl. Earth Obs. Geoinf.* 13: 277–291.
- CITES (2023). Convention on international trade in endangered species of wild Fauna and Flora: Appendix I, II & III, Available at: <https://cites.org/eng/app/appendices.php> (Accessed 15 April 2022).
- Cody, R.B., Dane, A.J., Dawson-Andoh, B., Adedipe, E.O., and Nkansah, K. (2012). Rapid classification of White Oak (*Quercus alba*) and Northern Red Oak (*Quercus rubra*) by using pyrolysis direct analysis in real time (DART™) and time-of-flight mass spectrometry. *J. Anal. Appl. Pyrolysis* 95: 134–137.
- Degen, B., Blanc-Jolivet, C., Stierand, K. and Gillet, E. (2017). A nearest neighbour approach by genetic distance to the assignment of individual trees to geographic origin. *Forensic Sci. Int.: Genet.* 27: 132–141.
- Deklerck, V., Mortier, T., Goeders, N., Cody, R.B., Waegeman, W., Espinoza, E., van Acker, J., van den Bulcke, J., and Beeckman, H. (2019). A protocol for automated timber species identification using metabolome profiling. *Wood Sci. Technol.* 53: 953–965.
- Deklerck, V., Lancaster, C.A., van Acker, J., Espinoza, E.O., van den Bulcke, J., and Beeckman, H. (2020). Chemical fingerprinting of wood sampled along a pith-to-bark gradient for individual comparison and provenance identification. *Forests* 11: 1–13.
- Dormontt, E.E., Boner, M., Braun, B., Breulmann, G., Degen, B., Espinoza, E., Gardner, S., Guillery, P., Hermanson, J.C., Koch, G., et al. (2015). Forensic timber identification: it's time to integrate disciplines to combat illegal logging. *Biol. Conserv.* 191: 790–798.
- Espinoza, E.O., Lancaster, C.A., Kreitals, N.M., Hata, M., Cody, R.B., and Blanchette, R.A. (2014). Distinguishing wild from cultivated agarwood (*Aquilaria* spp.) using direct analysis in real time and time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 28: 281–289.
- European Union (2010). Regulation (EU) No 995/2010 of the European Parliament and of the Council of 20 October 2010 laying down the obligations of operators who place timber and timber products on the market: Legislation 295 with EEA relevance. *Off. J. Eur. Union* 53: 23–34.
- European Union (2023). Regulation (EU) No 2023/1115 of the European Parliament and of the Council of 31 May 2023: on the making available on the Union market and the export from the Union of certain commodities and products associated with deforestation and forest degradation and repealing Regulation (EU) No 995/2010. *Off. J. Eur. Union* 150: 206–247.
- FAO (2021). FAOSTAT - forestry production and trade, Available at: <https://www.fao.org/faostat/en/#data/FO> (Accessed 12 September 2022).
- FAO (2022). Pulp and paper capacities. Survey 2021–2026, Available at: <https://www.fao.org/3/cc1985t/cc1985t.pdf> (Accessed 16 August 2023).
- Gasson, P. (2011). How precise can wood identification be? Wood anatomy's role in support of the legal timber trade, especially cites. *IAWA J.* 32: 137–154.
- Gerber, L., Eliasson, M., Trygg, J., Moritz, T., and Sundberg, B. (2012). Multivariate curve resolution provides a high-throughput data processing pipeline for pyrolysis-gas chromatography/mass spectrometry. *J. Anal. Appl. Pyrolysis* 95: 95–100.
- Grützmann, N. (2013). *Herstellung und Bleiche von Zellstoffen aus Tropenholz: als Beitrag zur Identifizierung in Zellstoff und Papier*. Diplomarbeit, Universität Hamburg.
- GTTN (2020). Scientific methods for taxonomic and origin identification of timber. In: *Global timber tracking network*. European Forest Institute and Thünen Institute, Available at: <https://www.researchgate.net/publication/342003654> (Accessed 15 May 2023).
- Hegnauer, R. (1986). Phytochemistry and plant taxonomy – an essay on the chemotaxonomy of higher plants. *Phytochemistry* 25: 1519–1535.
- Helmling, S., Olbrich, A., Tepe, L., and Koch, G. (2016). Qualitative and quantitative characteristics of macerated vessels of 23 mixed tropical hardwood (MTH) species: a data collection for the identification of wood species in pulp and paper. *Holzforschung* 70: 839–844.
- Helmling, S., Olbrich, A., Heinz, I., and Koch, G. (2018). Atlas of vessel elements: identification of Asian timbers. *IAWA J.* 39: 249–352.
- Hirschberger, P., Jokiel, D., Plaep, C., and Zahnen, J. (2010). Tropenwaldzerstörung für Kinderbücher. In: *Eine Analyse des Buchmarktes in Deutschland*. WWF, Deutschland, Available at: <https://www.wwf.de/fileadmin/fm-wwf/Publikationen-PDF/wwf-kinderbuchstudie-2009.pdf> (Accessed 12 September 2022).
- Holmberg, M. (1999). Pitch and precipitate problems. In: Gullichsen, J. and Paulapuro, H. (Eds.), *Papermaking chemistry: book 4*. Fapet Oy, Helsinki, pp. 223–239.
- Iivessalo-Pfäffli, M.-S. (1995). *Fiber atlas: identification of papermaking fibers*. Springer Verlag, Berlin/Heidelberg.
- InsideWood (2004 onwards). *Database. Published on the internet*. NC State University, Available at: <https://insidewood.lib.ncsu.edu/> (Accessed 19 April 2023).
- Khrisanfov, M. and Samokhin, A. (2022). A general procedure for rounding m/z values in low-resolution mass spectra. *Rapid Commun. Mass Spectrom.* 36: 1–6.
- Kite, G.C., Green, P.W.C., Veitch, N.C., Groves, M.C., Gasson, P.E., and Simmonds, M.S.J. (2010). Dalnigrin, a neoflavonoid marker for the identification of Brazilian rosewood (*Dalbergia nigra*) in CITES enforcement. *Phytochemistry* 71: 1122–1131.
- Krogell, J., Holmbom, B., Pranovich, A., Hemming, J., and Willför, S. (2012). Extraction and chemical characterization of Norway spruce inner and outer bark. *Nord. Pulp Pap. Res. J.* 27: 6–17.
- Lancaster, C. and Espinoza, E. (2012). Analysis of select *Dalbergia* and trade timber using direct analysis in real time and time-of-flight mass spectrometry for CITES enforcement. *Rapid Commun. Mass Spectrom.* 26: 1147–1156.
- Low, M.C., Schmitz, N., Boeschoten, L.E., Cabezas, J.A., Cramm, M., Haag, V., Koch, G., Meyer-Sand, B.R., Paredes-Villanueva, K., Price, E., et al. (2022). Tracing the world's timber: the status of scientific verification technologies for species and origin identification. *IAWA J.* 1: 1–22.
- Lowe, A.J. and Cross, H.B. (2011). The application of DNA methods to timber tracking and origin verification. *IAWA J.* 32: 251–262.
- Ng, C.H., Lee, S.L., Tnah, L.H., Ng, K.K.S., Lee, C.T., Diway, B., and Khoo, E. (2017). Geographic origin and individual assignment of *Shorea platyclados* (Dipterocarpaceae) for forensic identification. *PLoS One* 12: 1–18.
- Nisgoski, S., Gonçalves, T.A.P., Sonsin-Oliveira, J., Ballarin, A.W., and Muñiz, G.I.B. (2021). Near-infrared spectroscopy for discrimination of

- charcoal from Eucalyptus and native Cerrado species. Contribution to a database for forestry supervision. *For Sci.* 67: 419–432.
- Onuorah, E.O., Nwabanne, J.T., and Nnabuike, E.L.C. (2015). Pulp and paper making potentials of *elaeis guineensis* (oil palm) grown in south east, Nigeria. *World J. Eng.* 12: 1–12.
- Örsa, F. and Holmbom, B. (1994). A convenient method for the determination of wood extractives in papermaking process waters and effluents. *J. Pulp Pap. Sci.* 20: 361–366.
- Ponnuchamy, V., Gordobil, O., Diaz, R.H., Sandak, A., and Sandak, J. (2021). Fractionation of lignin using organic solvents: a combined experimental and theoretical study. *Int. J. Biol. Macromol.* 168: 792–805.
- Rachmayanti, Y., Leinemann, L., Gailing, O. and Finkeldey, R. (2009). DNA from processed and unprocessed wood: factors influencing the isolation success. *Forensic Sci. Int.: Genet.* 3: 185–192.
- Richter, H.G. and Dallwitz, M.J. (2000 onwards). Commercial timbers: descriptions, illustrations, identification, and information retrieval, in English, French, German, Portuguese, and Spanish. Version: 9th April 2019. Available at: <https://www.delta-intkey.com/wood/index.htm> (Accessed 19 April 2023).
- Savitzky, A. and Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36: 1627–1639.
- Schmitz, N., Beeckman, H., Cabezas, J.A., Cervera, M.T., Espinoza, E., and Fernandez-Golfin, J. (2019). *The timber tracking tool infogram. Overview of wood identification methods' capacity.* Global Timber Tracking Network, GTTN secretariat and European Forest Institute and Thünen Institute, Available at: <https://www.researchgate.net/publication/332223451> (Accessed 15 May 2023).
- Schmitz, N., Beeckman, H., Blanc-Jolivet, C., Boeschoten, L., Braga, J., Cabezas, J.A., Chaix, G., Crameri, S., Degen, B., Deklerck, V., et al. (2020). *Overview of current practices in data analysis for wood identification. A guide for the different timber tracking methods. A guide for the different timber tracking methods.* Global Timber Tracking Network, GTTN secretariat and European Forest Institute and Thünen Institute, Available at: <https://www.researchgate.net/publication/342281916> (Accessed 15 May 2023).
- Schwanninger, M., Rodrigues, J.C., and Fackler, K. (2011). A review of band assignments in near infrared spectra of wood and wood components. *J. Near Infrared Spectrosc.* 19: 287–308.
- Sieburg-Rockel, Jördis and Koch, Gerald (2020). Identification of wood species used in particleboard production. *IAWA J.* 41: 751–760.
- Silva, D.C., Pastore, T.C., Soares, L.F., de Barros, F.A, Bergo, M.C., Coradin, V.T., Gontijo, A.B., Sosa, M.H., Chacón, C.B., and Braga, J.W. (2018). Determination of the country of origin of true mahogany (*Swietenia macrophylla* King) wood in five Latin American countries using handheld NIR devices and multivariate data analysis. *Holzforschung* 72: 521–530.
- Snel, F.A., Braga, Jez W.B., Da Silva, D., Wiedenhoef, A.C., Costa, A., Soares, R., Coradin, Vera T.R., and Pastore, Tereza C.M. (2018). Potential field-deployable NIRS identification of seven *Dalbergia* species listed by CITES. *Wood Sci. Technol.* 52: 1411–1427.
- Stein, S.E. and Scott, D.R. (1994). Optimization and testing of mass spectral library search algorithms for compound identification. *J. Am. Soc. Mass Spectrom.* 5: 859–866.
- Tsuchikawa, S. and Kobori, H. (2015). A review of recent application of near infrared spectroscopy to wood science and technology. *J. Wood Sci.* 61: 213–220.
- Uryu, Y.e. a. (2008). *Deforestation, forest degradation, biodiversity loss and CO₂ emissions in Riau, Sumatra, Indonesia. Technical report, WWF.* Indonesia. Available at: https://files.worldwildlife.org/wwfcomprod/files/Publication/file/2i5fo619lz_WWF_Indo__27Feb08__Riau_Deforestation___English.pdf (Accessed 12 September 2022).
- Valto, P., Knuutinen, J., and Alén, R. (2012). Overview of analytical procedures for fatty and resin acids in the papermaking process. *Bioresources* 7: 6041–6076.
- Wan Daud, W.R. and Law, K.-N. (2011). Oil palm fibers as papermaking material: potentials and challenges. *Bioresources* 6: 901–917.
- Watkinson, C., Gasson, P., Rees, G., and Boner, M. (2020). The Development and use of isoscapes to determine the geographical origin of *Quercus* spp. in the United States. *Forests* 11: 1–21.
- Welling, J. and Liese, W. (2019). Wood, bamboo and palm wood: similarities and differences in research and technology development. In: *By-Products of palm Trees and their applications, 2018/12/15.* Materials Research Forum LLC, Aswan, Egypt, pp. 83–87.
- Wenig, P. (2011). Post-optimization of Py-GC/MS data: a case study using a new digital chemical noise reduction filter (NOISERA) to enhance the data quality utilizing OpenChrom mass spectrometric software. *J. Anal. Appl. Pyrolysis* 92: 202–208.
- Wenig, P. and Odermatt, J. (2010a). Efficient analysis of Py-GC/MS data by a large scale automatic database approach: an illustration of white pitch identification in pulp and paper industry. *J. Anal. Appl. Pyrolysis* 87: 85–92.
- Wenig, P. and Odermatt, J. (2010b). OpenChrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data. *BMC Bioinf.* 11: 1–9.
- Wheeler, E.A. (2011). InsideWood – a web resource for hardwood anatomy. *IAWA J.* 32: 199–211.
- Willför, S., Hemming, J., and Leppänen, A.-S. (2006). *Analysis of extractives in different pulps: method development, evaluation, and recommendations.* Report B1-2006. Laboratory of Wood and Paper Chemistry, Turku, Finland.
- Windhagen, K., Moldenhauer, T., Burkard, A., and Geiger, G.A. (2019). *Paper - annual report. Ein Leistungsbericht (Annual Report).* Verband Deutscher Papierfabriken e. V, Bonn, Germany.
- Zemke, V., Haag, V., and Koch, G. (2020). Wood identification of charcoal with 3D-reflected light microscopy. *IAWA J.* 41: 478–489.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/hf-2023-0048>).