



DATA PAPER

Open Access



# Wood, knots and bark extractives for oak, beech and Douglas fir: a dataset based on a review of the scientific literature

Béatrice Richard<sup>1\*</sup> , Alain Bénard<sup>1</sup> , Stéphane Dumarçay<sup>2</sup>  and Francis Colin<sup>1</sup>

## Key message

Wood\_db-chemistry gathers information on oak, beech and Douglas fir extractives studied between 1950 and 2020. The data can be used by researchers, stakeholders in the forest-wood sector and the chemical industry to make the most of the extractives and their bioactive properties through knowledge of their diversity in the forest resource. Dataset access is at <https://doi.org/10.57745/QZYPUA>. Associated metadata are available at <https://metadata-afs.nancy.inra.fr/geonetwork/srv/fre/catalog.search#/metadata/4f8c07d2-c0f6-4958-8f74-936054a9870a>.

**Keywords** Oak, Beech, Douglas fir, Extractives, Phenolic compounds and terpenoids, Bioactivity

## 1 Background

Wood, knots and bark extractives are the primary and secondary metabolites extracted using unbiased crude extraction procedures aimed at efficiently extracting all or most metabolites in their natural form prior to analysis in the solvents used. Secondary metabolites are metabolic products that are not directly produced by photosynthesis but are, instead, the result of mechanisms by which an organism biosynthesises unique compounds that express the individuality of a species. The secondary metabolites are natural components of plant extracts, referred to as natural products (Dias et al. 2012). Natural products are produced either as a result of the organism adapting to its surrounding environment or to act as a possible defence mechanism against predators to assist in the survival of

the organism. It is the unique biosynthesis of these natural products by the countless number of terrestrial and marine organisms that provides the characteristic chemical structures that possess an array of biological activities (Rowell 2012). The main extractable secondary metabolites are phenolic compounds, terpenoids and alkaloids, present in low concentrations (a small % of total carbon, excluding lignin) in plant tissues, often stored in specialised cells or organs (Tanase et al. 2019). The extracts contain natural bioactive compounds that give them a range of desirable properties for use in health (Weidmann 2012; Mármol et al. 2019; Ferreira-Santos et al. 2020), food (Lourenço et al. 2019; Shah et al. 2014), cosmetics (Zillich et al. 2015) and wood preservation (Singh and Singh 2012; Shirmohammadli et al. 2018).

To combat global warming and to reduce dependence on fossil raw materials, the French public authorities are supporting the development of plant-based chemistry, the seventh principle of green chemistry (Anastas and Warner 1998). By 2030, as a result of the rich and abundant forestry and agricultural biomass of metropolitan France, French authorities hope that 25–30% of basic chemical molecules will be produced from renewable resources. This would be part of an “industrial revival”

Handling editor: Véronique Lesage

\*Correspondence:

Béatrice Richard  
beatrice.richard@inrae.fr

<sup>1</sup> Université de Lorraine, AgroParisTech, INRAE, UMR 1434 Silva, F-54000 Nancy, France

<sup>2</sup> Université de Lorraine, INRAE, EA 4370 USC LERMAB, F-54506 Vandoeuvre-Lès-Nancy, France



© The Author(s) 2024. **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/>.

of plant-based chemicals through the development of synergies between forestry and industry. According to the ADEME/IGN/FCBA: Colin and Thivolle-Cazat (2016) report, the greatest potential for developing forest resources would lie in hardwoods and private forests.

One of the obstacles to the development of extractives derived from forest biomass is the lack of knowledge of their tree-level and resource heterogeneity, as reported by Arbenz and Avérous (2015) for the industrial development of tannin derivatives, and by Zidorn (2018) for the seasonal variation of natural products in European trees.

The question then arises as to where (in terms of tree parts and geo-pedo-climatic situations) the most attractive and/or accessible fractions of forest biomass are from a technical and economic point of view, and which ones could be extracted.

The mobilisation of extractable resources cannot be achieved without the willingness of the many actors involved in forestry and wood-related industries to develop a forest-based chemical industry.

We present the Wood\_db-chemistry database developed within and for the needs of the EXTRAFOR\_EST scientific project (2017–2022): “Extractive chemical components from the forests of Eastern France”, with partners from LERMAB (Laboratoire d’études et de recherche sur le matériau bois, University of Lorraine) and the IGN (National Geographic Institute, Direction territoriale Nord-Est France). The aim of the present dataset is to make available and facilitate the dissemination of existing scientific knowledge on wood, knots and bark extractives focusing mainly on secondary metabolites and their bioactivity in *Quercus robur* L., *Quercus petraea* Liebl., *Fagus sylvatica* L. and *Pseudotsuga menziesii* (Mirb.) Franco. The data will be linked to tissues located at different levels of the tree (wood, sapwood, heartwood, knots, bark) and to the industrial biomass resource. In order to make informed decisions about the best use of extractives containing natural products for potential industrial applications, the dataset will also provide knowledge related to their diversity in the forest resource.

## 2 Methods

### 2.1 Bibliographic methodology

The studies presented in this data paper were mainly extracted from published scientific literature reviews retrieved via (i) the scientific literature platforms (Web of Science, Crossref Metadata Search, Europe PubMed Central, Scopus, Hal); (ii) the search engines Google Scholar and Base (Bielefeld Academic Search Engine); (iii) the social network ResearchGate; and (iv) a proximity search of previously obtained resources or review articles.

The main keywords used for queries and searches were as follows: *Fagus sylvatica*, *Quercus robur*,

*Quercus petraea*, *Pseudotsuga menziesii*, extractives, wood extracts, wood, bark, knot, chemical wood, tree, phenolic compounds, terpenoids, polyphenols, tannin, taxifolin, resin acids, heartwood, sapwood, knotwood, natural biocides, bioactivity, antioxidant, fungicide, green extraction, biorefinery, mill, waste biomass. We selected only the studies on healthy trees, conducted on the European indigenous tree species and on Douglas fir as well, which is a species broadly introduced from North America, for which additional studies from North America were analysed.

The 59 articles resulting from this bibliographic search were classified by species (Table 1). The content of each article was previously analysed in order to identify and select the data of interest to be retained for the construction of the database. The items searched for in the articles for each tissue were the extraction process, the solvents used, the yield of the extracts, the chemical nature of compounds analysed in the extract and their relative composition, the bioactivity of the extracts with the name and level result of the test applied, and the origin of the samples collected (forest stand or processing company). Some of the publications contained data on more than one of the four forest species for the same reference.

### 2.2 Structure of the Wood\_db-chemistry database

The chosen method is based on the analysis of adding new tables to the existing object-oriented relational database, Wood\_db, adapted to the PostgreSQL database management system. The method links several tables of values (= variables) organised around objects. The “extracts” table is central. It contains information formally identifying an extract and is linked to other tables that associate an extract with extractives or property information. The metadata of the extracts is managed according to a principle known as “generic measures”, which associates an object (in this case, the extracts) with variables. Each association carries a value (that of the metadata) and, possibly, additional

**Table 1** Metrics from a Wood\_db-chemistry database query for the literature studied

Taxonomy/species	Number of references
<i>Fagus sylvatica</i> L., 1753	9
<i>Pseudotsuga menziesii</i> (Mirb.) Franco, 1950	26
<i>Quercus robur</i> L., 1753	21
<i>Quercus petraea</i> Liebl., 1784	20
<i>Quercus</i> L., 1753	4

information (units, protocol, operators, etc.). All of the new tables created are intended to add a chemistry component to the Wood\_db database. We will refer to this as Wood\_db-chemistry.

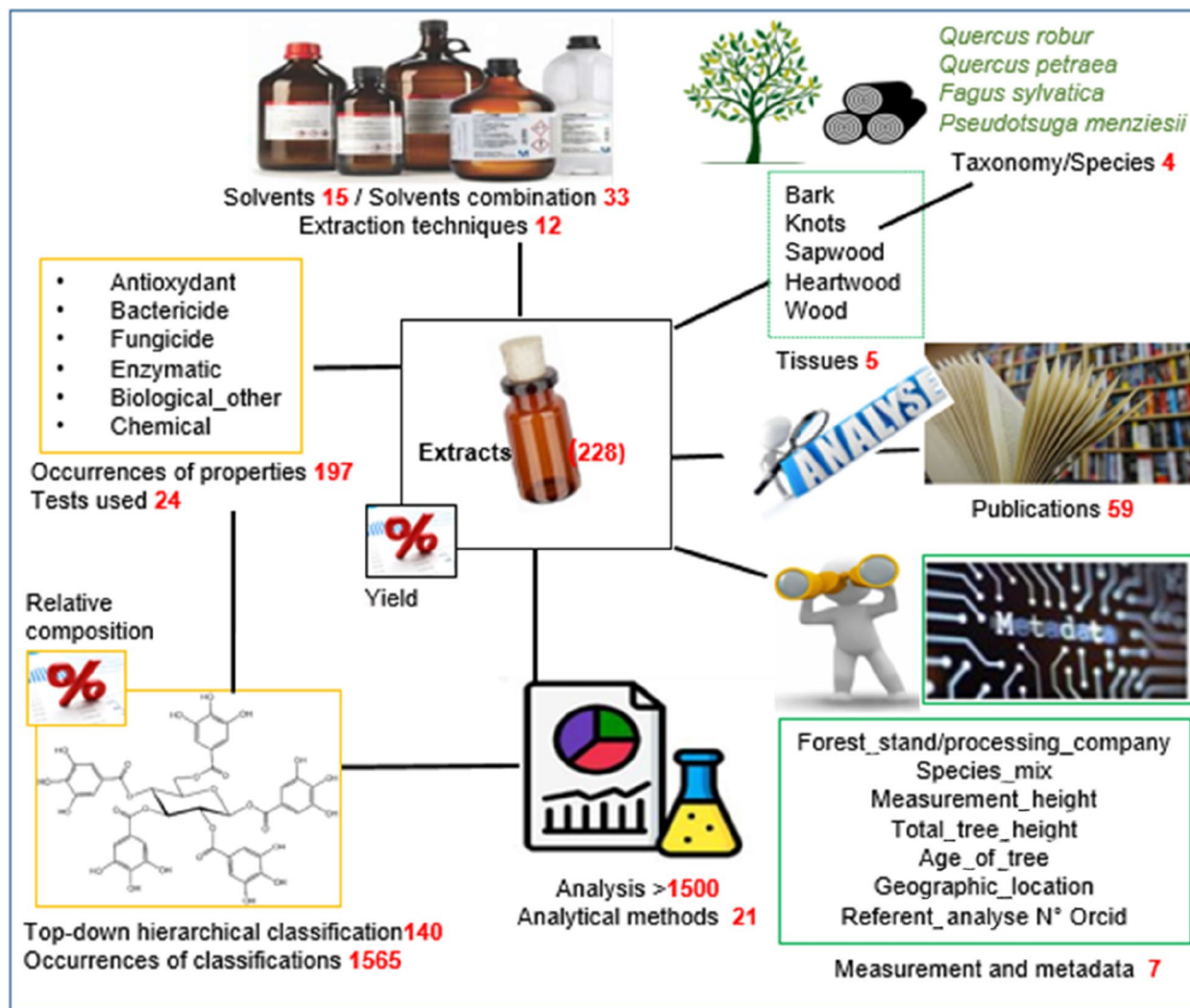
Before the creation of the database and in order to define its structure, the qualified person in charge of the bibliographic analysis was audited by the IT specialist (SIG-DB: Geographic Information System and Database, plateau Silva INRAE Nancy Grand Est). This was done in order to make an inventory of the data of interest to be taken into account (i.e. the chemistry data were limited to quantitative data and possibly qualitative data that may be easily usable by the different users of the database), to classify these data in tables and to define the constraints that ensure the integrity of data within and between tables. Once the structure had been defined, several

exchange file formats were defined and data integration processes were developed with the help of specialised tools (Extract Transform Load: Talend Data Integration). The qualified person entered the data, which were then integrated into the database by the IT specialist. SQL language query tools are used to query the database.

### 2.3 Wood\_db\_chemistry database design

The Wood\_db-chemistry consists of 12 exchange files (one file/table generated) completed for each of the forest species included in the database. An overview of the Wood\_db-chemistry design with a table relationship diagram is provided in Fig. 1. This diagram also shows some database metrics.

The 59 selected bibliographic references covered the production of 228 extracts. These are the results of the



**Fig. 1** Overview of the design of the Wood\_db-chemistry database. Relationship diagram of tables and metrics

chemical extraction of extractives. The content of these extracts is expressed as a percentage of the dry matter of the tissue. Each extract was obtained by the use of one of the 15 solvents listed, or by a combination of more than one of them. A total of 12 different extraction techniques were used. The chemical characterisation of the analytes in the extract was determined by more than 1500 analyses using 21 different analytical methods. The various extractives identified (1565 occurrences) were grouped into classes, subclasses, families, subfamilies and compounds using a top-down hierarchical classification (140 classifications), usually based on structural and functional similarities and metabolic pathways of production. The yield of chemical compounds is very often given as the equivalent of structurally similar compounds that are available on the market.

The bioactive properties—antioxidant, bactericidal, fungicidal, enzymatic, biological\_other and chemical properties—were mainly established for whole extracts and only a few for single chemical compounds. A total of 197 occurrences of properties were recorded by means of 24 different tests with response levels. Concerning the metadata of the extracts, seven variables of interest were identified during the literature review and were added when the information was available. A complete list of the literature reviewed is available in the data of Wood\_db-chemistry (Richard et al 2023).

### 3 Access to the data and metadata description

The wood\_db-chemistry database (Richard et al. 2023) is available at the Recherche Data Gouv repository with the URL: <https://doi.org/10.57745/QZYPUA>. For metadata description, see <https://metadata-afs.nancy.inra.fr/geonetwork/srv/fre/catalog.search#/metadata/4f8c07d2-c0f6-4958-8f74-936054a9870a>.

The database is open by exporting three files in Comma Separated Values (CSV) format, encoded in UTF8, each generated by a SQL language query tool sent to the database. The contents of the three CSV files proposed accurately reflect the contents of the wood\_db-chemistry database at any given time. By default, Recherche Data Gouv repository offers a version of the deposited files in (tab) format; the separator is a tabulation or space. It is still possible to request a download in the original format, that is CSV. The descriptive information provided in this document relates to the CSV format used at the time of deposit. Using the tools provided, the various pieces of information contained in each of the three files can be filtered and grouped using a spreadsheet programme that manages UTF8 encoding, (i.e. Libre Office Calc).

3.1 The file in the database entitled “Extract\_process\_extractives\_content\_and\_compounds\_analysis”

includes research information on the chemistry of wood, knots and bark extractives.

This information, presented in one column, is divided into five topics:

- Bibliographic information (four columns): author, year, DOI/URL, citation/format RIS
- Focus (two columns): species, tissue
- Extract process (two columns): extraction\_process, extract\_sequence\_number
- Content of extracts (seven columns): sample preparation (tissue preparation, extract conditioning and preparation for analysis), choice of solvent (solvent, sequence of solvent use), extractives yield (extractives\_content\_per\_solvent, summative\_extractives\_content, phenolic extractives)
- Extractives determination (eight columns): chemical compounds (top-down hierarchical classification of extractives, position in the hierarchical classification, chemical extractives name, compound CAS registry number, extractives presence, rate of extractives), analytical used (analytical methodology, comment/analysis)

This file can be used to explore the extractives content and/or phenolic yield of an extract by species, by tissue, by extraction method, by solvent or solvent sequence, by author, by year and by DOI/URL or RIS citation. Information can be accessed on the sequence of solvents used to obtain an extract and the sampling of tissues and their preparation for analysis. It is also possible to consult a comment field, which provides additional information on the extraction conditions.

The file also allows users to explore the chemistry of the extractives using analytical techniques via the position in the top-down hierarchical scheme (class/subclass/family/subfamily/compound), the common name, the rate, the CAS number in the case of a compound and its presence/absence. Text fields provide additional information on the metadata about the conditions of analysis.

3.2 The file in the database entitled “Properties of extracts” specifies the active biological or chemical properties of extracts.

In addition to the focus, the bibliographic information and the extract process, we considered the following topics:

- The properties looked for (antioxidant, bactericide, fungicide, enzymatic, biological other, chemical);
- Their highlighting with the level of response (performed test, the result of the test, comment on the test used).



	Total Extractives content (% of dry mass)	Families of major compounds (% of dry mass)	Major compounds (% of dry mass)	Other compounds
<b>Bark</b>	<p><b>Ethanol/water (50:50) : 12% to 36%</b> (ASE triple cycle renewed solvent 1600 psi 100°C)</p> <p><b>Ethanol : 5.3% to 24.7%</b> (soxhlet)</p> <p>Dichlorométhane :1% to 9% (soxhlet)</p> <p><b>Acetone/water (95:5): 13%</b> (ASE)</p> <p>Water : 1.4% to 7,8% (soxhlet)</p> <p>Methanol : 4.8% (soaking)</p> <p>Methanol thioglycolic acid (50:50) : 1.3% to 10.8%</p> <p>phénolics yield (leaching)</p>	<p><b>Phenolics dimers or oligomers : condensed tanins proanthocyanidin 22.5%</b> (methanol/water 75:25) soxhlet after DCM extraction; 0.4% à 4.1% (ethanol); 0.15% to 0.2% (water); 1.3% to 8% (methanol/thioglycolic acid 50:50); 2.2% (methanol/water 50:50)</p> <p><b>Phenolics dimers or oligomers : hydrolysable tanins ellagitannins 6.3 to 10.2%</b> (ethanol); 1.3% to 3.6% (water)</p> <p><b>Phénoliques monomères : flavonoïdes 3.4% to 4.5%</b> (ethanol); 1.1% to 1.8% (water); 2.6% fresh mass (méthanol)</p> <p><b>Terpenes diterpénoïdes : resins acids and derivatives 0.5%</b> (dichlorométhane)</p>	<p><b>Procyanidine : 3.1%</b> (acetone/water 70:30); 2.2% (méthanol/water 75:25); 0.005% to 0.1% (ethanol/water 50:50)</p> <p><b>Taxifolin : 0.1% to 8.9%</b> (méthanol/thioglycolic acid 50:50)</p> <p><b>Eriodictyol : 0.03% to 4.4%</b> (ethanol/water 50:50)</p> <p><b>Catechin : 0.02% to 1%</b> (ethanol/water 50:50)</p> <p><b>Epicatechin : 0.006% to 1.5%</b> (ethanol/water 50:50)</p> <p><b>Dehydroabietic acid : 0.3%</b> (dichlorométhane)</p>	<p>Dihydrokaempferol, Simplicoside</p> <p>Dehydroabietic acid, Abietic acid, Isopimaric acid</p> <p>Palustric acid, Levopimaric acid, Levopimaric acid</p> <p>Pinoresinol, Lariciresinol, Lignan esters</p> <p>Terpene triterpenoids sterols : beta-Sitosterol, Campesterol</p> <p>Phenolic monomers stilbenes</p> <p>Phenolic monomers acid : caffeic acid</p>
<b>Knot</b>	<p><b>Ethanol : 3% à 21%</b> (ASE triple cycle renewed solvent 1600 psi 100°C)</p> <p><b>Acetone /water (95:5): 10% to 15%</b> (ASE)</p> <p>Acetone : 3%</p>	<p><b>Phenolics monomères : flavonoïdes 0.1% to 4.2%</b> (acetone /water 95:5)</p> <p><b>Phenolics monomères : lignanes 0.6% to 8.1%</b> (acetone/water 95:5)</p> <p><b>Phenolics dimers or oligomers oligolignans 0.6% to 0.8%</b> (acetone/water 95:5)</p> <p><b>Phenolics dimers or oligomers oligolignans dilignans 0.1% to 0.4%</b> (acetone/water 95:5)</p> <p><b>Phenolics dimers or oligomers oligolignans sesquilignans : 0.2% to 0.6%</b> (acetone/water 95:5)</p> <p><b>Terpenes diterpénoïdes : resins acids and derivatives 0.2% to 1.9%</b> (hexane)</p> <p><b>Terpenes sesquiterpénoïdes : Juvabiones 0,3% to 0,5%</b> (hexane); 0.1% to 0.3% (acetone /water 95:5)</p>	<p><b>Taxifolin : 0.1% to 4.2%</b> (acetone/water 95:5)</p> <p><b>Cyclolariciresinol (isolariciresinol) : 0.07% to 5.5%</b> (acetone/water 95:5)</p> <p><b>Secoisolariciresinol : 0.03% to 1.4%</b> (acetone/water 95:5)</p> <p><b>Nortrachelogenin: 0.08% to 0.6%</b> (acetone/water 95:5)</p> <p><b>Lariciresinol : 0.009% to 0.5%</b> (acetone/water 95:5)</p> <p><b>7-Hydroxymatairesinol : 0,02% to 0,2%</b> (acetone /water 95:5)</p> <p><b>Todolactol A : 0.05% to 0.3%</b> (acetone/water 95:5)</p> <p><b>Dihydrotodomatuic acid : 0.51% to 0.54%</b> (hexane)</p>	<p>Dihydrokaempferol</p> <p>Matairesinol, Isohydroxymatairesinol, Alpha- conidendrin, Pinoresinol</p> <p>Phenolics dimers or oligomers oligolignans sesterlignans</p> <p>Palustric acid, Isopimaric acid, Abietic acid, Levopimaric acid, Dehydroabietic acid, Sandaracopimaric acid, Neoabietic acid, Thunbergol</p> <p>Terpene triterpenoids stérols : beta-Sisterol, Campesterol, Sitostanol</p>
<b>Heartwood</b>	<p><b>Toluene/ethanol (66:33) then ethanol : 13.56%</b> (soxhlet)</p> <p><b>Acetone/water (70:30) : 2% to 9%</b> (soaking)</p> <p>Diethyl ether : 2.1% of dry mass extractibles free</p> <p>Methanol/water (70:30): 2.4% (phénolics yield)</p> <p>Methanol/water (80:20): 1.4% (phénolics yield)</p> <p>Acetone :1.5%</p>	<p><b>Phénoliques dimers or oligomers : condensed tanins proanthocyanidin. 0.01%</b> (methanol/water 80:20)</p> <p><b>Phénoliques monomères : flavonoïdes : 2.6% to 3.2%</b> (acetone/water 95:5)</p> <p><b>Phenolics monomères : Lignanes : 0.2%</b> (acétone/water 95:5)</p> <p><b>Phenolics dimers or oligomers oligolignans : 0.005% to 0.2%</b> (acétone/water 95:5)</p> <p><b>Phenolics dimers or oligomers oligolignans dilignans : 0.04%</b> (acetone/water 95:5)</p> <p><b>Phenolics dimers or oligomers oligolignans sesquilignans : 0.1% to 0.14%</b> (acetone/water 95:5)</p> <p><b>Phenolics dimers or oligomers oligolignans sesterlignans : 0.005% to 0.024%</b> (acetone/water 95:5)</p> <p><b>Terpenes diterpénoïdes : resins acids and derivatives: 0.09% to 0.13%</b> (hexane); 0.3% of dry mass extractibles free (diethyl ether)</p> <p><b>Terpenes sesquiterpénoïdes :Juvabiones : 0.03% à 0.05%</b> (hexane)</p>	<p><b>Taxifolin : 5% to 9%</b> (acetone/water 70:30); 2.5% to 3% (acetone/water 95:5); 1.5% (methanol/water 70:30)</p> <p><b>Dihydrokaempferol : 0,03% à 0,19%</b> (acetone/water 95:5)</p> <p><b>Cyclolariciresinol : &lt;0.005% à 0.2%</b> (acetone/water 95:5)</p>	<p>Procyanidine , [5-5]-Bisdihydroquercetin, Procyanidin B (dimer 1, dimer2, dimer3), Procyanidin C (dimer1, dimer2, dimer3)</p> <p>Pinocembrin, Dihydrokaempferol, Catechin, Epicatechin Taxifolin-3'-O-glucoside, Eriodictyol, Quercetin-3'-O-glucoside</p> <p>Matairesinol, Secoisolariciresinol, Todolactol A, Alpha- conidendrin, 7-Hydroxymatairesinol, Lariciresinol, Nortrachelogenin, Pinoresinol, Lignan-O-glucoside</p> <p>Palustric acid, Isopimaric acid, Abietic acid, Levopimaric acid, Neoabietic acid, Dehydroabietic acid, Thunbergol, Sandaracopimaric acid</p> <p>Terpene triterpenoids sterols : beta-Sitosterol, Campesterol, Sitostanol, Cycloartenol, Stigmastanol, 24-Methylenecycloartanol</p>
<b>Sapwood</b>	<p><b>Acetone/ water (70:30) : 2.5% à 9%</b> (ASE 10 bars 100°C)</p> <p>Diethyl ether : 0.8% of dry mass extractibles free</p> <p>Methanol/water (70:30): 0.7% (phénolics yield)</p>	<p><b>Phénoliques monomères : flavonoïdes &lt;0.005%</b> (acetone/water 95:5)</p> <p><b>Phenolics monomères : Lignanes 0.01% à 0.02%</b> (acetone/water 95:5)</p> <p><b>Phenolics dimers or oligomers oligolignans dilignans : 0.007% to 0.01%</b> (acetone/water 95:5)</p> <p><b>Phenolics dimers or oligomers oligolignans sesquilignans : 0.007% to 0.009%</b> (acetone/water 95:5)</p> <p><b>Phenolics dimers or oligomers oligolignans sesterlignans : 0.014% to 0.019%</b> (acetone/eau 95:5)</p> <p><b>Terpenes diterpénoïdes : resins acids and derivatives: 0,2% of dry mass extractibles free (diethyl ether); 0.1% to 0.15%</b> (hexane)</p>	<p>Pinocembrin, Dihydrokaempferol, Catechin, Epicatechin, Taxifolin-3'-O-glucoside, Eriodictyol, Quercetin-3'-O-glucoside</p> <p>Matairesinol, Secoisolariciresino, Todolactol A, alpha- Conidendrin, 7-Hydroxymatairesinol, Lariciresinol, Nortrachelogenin, Pinoresinol, Lignan-O-glucoside</p> <p>Palustric acid, Isopimaric acid, Abietic acid, Levopimaric acid, Neoabietic acid, Dehydroabietic acid, Thunbergol, Sandaracopimaric acid</p> <p>Terpene triterpenoids sterols : beta-Sitosterol, Campesterol, Sitostanol, Cycloartenol, Stigmastanol, 24-Methylenecycloartanol</p>	<p>Taxifolin, Catechin, Epicatechin, Eriodictyol, Pinocembrine, Quercetin-3'-O-glucoside, Dihydrokaempferol</p> <p>Pinoresinol, Lignan-O-glucoside, Cyclolariciresinol, Lignan-O-glucoside; alpha-Conidendrin, 7-Hydroxymatairesinol, Lariciresinol, Matairesinol, Nortrachelogenin, Secoisolariciresinol, Todolactol</p> <p>Abietic acid, palustric acid, isopimaric acid, neoabietic acid, levopimaric acid</p> <p>Terpene triterpenoids sterols : beta-Sitosterol, Campesterol, Campestanol, Sitostanol, Stigmastanol, 24-methylenecycloartanol</p>

**Fig. 3** Example of the formatting of the data extracted from the Wood\_db-chemistry database in response to the query: "What are the summary extractive contents, the families of major chemical compounds and the respective contents of these and other compounds present in Douglas fir for different tissues: bark, knots, heartwood and sapwood?" The results of the natural products are listed in different colours: condensed tannins (light blue), hydrolysable tannins (dark blue), flavonoids (brown), lignans (purple), stilbenes and acid (orange), oligolignans (pink), diterpenoids (light green), sesquiterpenoids (dark green) and triterpenoids (black)

by an analyst identified in the database by its name and Orcid number. The articles under review are referenced in the database (citation/RIS format), making it easy to go back and forth in the bibliographic analysis, if necessary. The storage of the data from the knowledge synthesis on wood and bark extractives in a database means that the

bibliographic analysis can be maintained over the long term. Unlike other less robust methods, traceability to published data is maintained. The database integration of the files resulting from the bibliographic analysis is also carried out by specialised staff using integration tools (Extract Transform Load, Talend Data Integration). This

is first done on a pre-production server and then on the production server once the necessary corrections have been made.

## 5 Reuse potential and limits

Wood\_db-chemistry data are stored in structured files in open csv format, which can be easily processed using any tool or language (R, Python, etc.) capable of handling this format. There are numerous and simple ways to query the data, as shown in Fig. 2. This database is a tool designed to help access scientific bibliographic knowledge on wood extractives of trees, which can also help to develop research hypotheses from existing knowledge. Data on wood and bark extractives from four tree species are available, but the structure of the database allows it to be extended to include other species. Researchers can contact us to define collaborative projects. These projects can include questions other than the answers to the three published exports.

Consultation of the Wood\_db-chemistry dataset is also an opportunity to raise awareness of the chemistry of woody plants among forestry and wood industry professionals.

- One of the results of the ExtraFor Est project, which aims to promote the use of forest biomass to supply industries, has been the construction of a decision-making tool designed to gather knowledge, to model the sectors and to analyse the feasibility of a regional forest-chemicals sector. This tool brings together data from the <https://extraforest.ign.fr> and AF Filière <https://dev.terriflux.com> utilities and records carbon data using CAT software (Pichancourt et al. 2021). Data from the Wood\_db-chemistry database can be used as input for this decision-making tool by integrating it into the website utility <https://extraforest.ign.fr>, along with the National Forest Inventory (IFN) and infradensity data, in order to estimate the quantities of extractable material present in the standing resource and/or removed annually from the bark, node and wood for a given geographical area.
- The dataset can be used as input variables for the SimCop-Qual simulator developed in the ExtraFor Est project. It tests the hypothesis that silviculture influences the amount of extractive compounds in a stand and is based on the existing growth simulator SimCop. Examples of data input variables to the SimCop-Qual tool—rate of extractives for “taxifolin” from the tissue “knot” and species “*Pseudotsuga menziesii*”—will give the tool output as the result of estimated amounts of taxifolin from the *Pseudotsuga menziesii* knots as a function of forest dynamics.

- In addition to the “market study” service provided by the IAR competitiveness cluster (Now Bioeconomy For Change B4C) for the ExtraFor Est project, the dataset can be consulted to obtain information about the antioxidant, bactericidal and fungicidal properties of extracts for which numerous market segments can be envisaged. These include the fields of cosmetics, human and animal nutrition (nutraceuticals and natural preservatives) and agriculture (bio-control and biostimulant).

Beyond the ExtraFor Est project, the extension of this database is planned with future partners, for the bark and wood extracts of various species, currently under international contract as part of the WoodChem Valley initiative (CRITTbois Epinal, France) and the Interreg ExtraBark project (Valbiom, Wallonia, Belgium), for the development of the wood chemistry industry. The joint creation of this database is planned as an extension and development of the existing Wood\_db-chemistry structure. In addition, the integration of data for the species *Abies alba* and *Picea abies* is already underway.

The perspective of this work is also to actually test the database, which is positively perceived by the future partners, with regard to the proven and defined needs for output of the results, in particular, the construction of the data sheets. This will lead to work on restructuring the database by creating new tables according to the questions to be answered. The remodelling of tables will involve new integration processes. Developing user interfaces for querying and even feeding the database would be beneficial to facilitate its use. Wood\_db-chemistry is a young database revisable by adding articles of interest if these are not yet available. The database can also be updated to keep research data current.

Disseminating this dataset will make information more available and help identify potential markets for wood, knots and bark extractives as a new way to add value to forest biomass.

### Acknowledgements

We are grateful to Philippe Gerardin from Université de Lorraine, INRAE, EA 4370 USC LERMAB, F-54506 Vandoeuvre-lès-Nancy, France, for constructive discussions throughout this work.

### Code availability

Not applicable.

### Authors' contributions

Conceptualisation: designed the study—BR, AB (leader) and FC, SD (supporting). Methodology: (i) collected the bibliographical references, carried out the analysis, extracted the relevant data and filled the relevant data in the exchange files—BR (leader); (ii) integrated the exchange files into the database—AB (leader); (iii) for the hierarchical classification of the chemical compounds—SD (supervisor). Formal analysis and investigation: built the Wood\_db-chemistry database and defined its structure and the exchange files—BR, AB (leader). Writing—original draft preparation: BR (leader), AB (equal), drafted the paper. Writing—review and editing: BR. Funding acquisition: FC, leader of

