3

An specific case in the classification of woods by FTIR and chemometrics: Discrimination of Fagales from Malpighiales

8 Ara Carballo-Meilan, Adrian Goodman, Mark Baron, Jose Gonzalez-Rodriguez

9 School of Life Sciences of the University of Lincoln, Brayford Pool, Lincoln, LN6 7TS, UK
10 E-mail: jgonzalezrodriguez@lincoln.ac.uk

11 Department of Chemical Engineering of the University of Loughborough, Loughborough, LE11 3TU,
12 UK
13 E-mail: M.A.Carballo-Meilan@lboro.ac.uk, Tel: + 44 1509 265203

Abstract

15 Fourier Transform Infrared (FTIR) spectroscopic data was used to classify wood samples from nine
16 species within the Fagales and Malpighiales using a range of multivariate statistical methods.
17 Taxonomic classification of the family Fagaceae and Betulaceae from Angiosperm Phylogenetic
18 System Classification (APG II System) was successfully performed using Supervised Pattern
19 Recognition techniques. A methodology for wood sample discrimination was developed using both
20 sapwood and heartwood samples. Ten and eight biomarkers emerged from the dataset to
21 discriminate order and family, respectively. In the species studied FTIR in combination with
22 Multivariate analysis highlighted significant chemical differences in hemicelluloses, cellulose and
23 guaiacyl (lignin) and shows promise as a suitable approach for wood sample classification.

Introduction

The polymeric composition of wood

26 Wood is composed mainly of cellulose, hemicellulose, and lignin where the cellulose microfibrils are
27 embedded within a matrix of hemicelluloses and lignin (Sjostrom 1993). Cellulose is the main
28 component of wood and the skeletal polysaccharide of cell walls (Barnett and Jeronimidis 2003).
29 Cellulose chains are aggregated together by inter- and intramolecular hydrogen bonds to form
30 microfibrils with highly ordered (crystalline) and less ordered (amorphous) regions (Martin 2006). On
31 other hand, hemicelluloses are non-cellulosic polysaccharides. The most important hemicelluloses in
32 wood cell walls are xylans and glucomannans (Barnett and Jeronimidis 2003). Lignin is a

33 macromolecule formed by the polymerization of three phenylpropane monomers, it is highly
34 heterogeneous and is the most complex structural component of the wood cell wall. Lignin attaches
35 to cellulose and hemicelluloses by lignin-polysaccharide bonds (Wang et al. 2009); hemicellulose
36 may be associated with both lignin and cellulose. In Spruce (*Picea abies*), it has been shown that
37 xylan is associated with lignin and glucomannan with cellulose (Martin 2006).
38 There is considerable variation in the composition of wood; it varies at all levels from species to
39 species, between cell types, and within the cell wall itself (Martin 2006).The underlying factors
40 controlling wood properties are essentially the result of its chemical composition and their relative
41 proportion and distribution (Barnett and Jeronimidis 2003).

Cross-sectional variation in the chemical composition of wood: Sapwood vs Heartwood

44 The outer region of wood with living parenchyma cells is known as sapwood and the inner region
45 with only dead cells is termed heartwood - this is often, but not always, a dark colour than the
46 sapwood due to the presence of various extractives (Ek et al. 2009; Martin 2006). The precise cause
47 of heartwood formation is not known but it is characterised by the accumulation of polyphenolic
48 substances in the cells and a general reduction in the moisture content of the wood. In addition to
49 its colour heartwood may be more aromatic on account of the extractives. The structure of bark is
50 complicated in comparison with wood (Sjostrom 1993). Its high variability is mainly attributed to its
51 rich composition in extractives (Ek et al. 2009).

52

53 In a tree considerable differences exist in the hemicellulose content and composition between the
54 stem, branches, roots and bark (Sjostrom 1993). Heartwood has higher lignin and lower cellulose
55 content compared to sapwood (Martin 2006; Meinzer et al. 2011). There is a similar trend with tree
56 height. Hemicelluloses seem to be more concentrated in the juvenile regions and usually decrease in
57 mature wood regions (Martin 2006). Differences in lignin composition(S/G ratio) were also found
58 between heartwood and sapwood and with tree height (Barnett and Jeronimidis 2003). Trees with a
59 decrease in lignin have a proportional increase in cellulose (Callow et al. 2006) while the two
60 components of holocellulose (i.e., α-cellulose and hemicellulose) tend to vary directly (Stewart et al.
61 1995).

Infrared spectroscopy

63 Fourier Transform Infrared (FTIR) Spectroscopy is a rapid, noninvasive, high-resolution analytical tool
64 for identifying types of chemical bonds in a molecule by producing an infrared absorption spectrum
65 that is like a molecular "fingerprint" (Shen et al. 2008). However, a single IR spectrum also contains
66 information about the molecular structure and intermolecular interactions among the individual
67 sample components (McCann et al. 2001). Infrared spectroscopy has been extensively applied to the
68 analysis of plant cell walls (Kacuráková et al. 2000). Furthermore, in combination with multivariate
69 analysis, FTIR has been used for the chemotaxonomic classification of flowering plants such as the
70 identification and classification of the genus of *Camellia* using cluster analysis and Principal
71 Component Analysis (PCA) (Shen et al. 2008), the taxonomic discrimination of three families using a
72 dendogram based on PCA (Kim et al. 2004) and the differentiation of plants from different genera
73 using cluster analysis (Gorgulu et al. 2007) (Table 1). In tree species wood has also been analysed
74 using FTIR spectroscopy to characterize lignin in wood (Obst 1982; Takayama 1997), determine

75 changes in composition and structure of wood (Stewart et al. 1995), and characterize softwood and
76 hardwood pulps using Partial Least-Squares analysis (PLS) and PCA (Bjarnestad and Dahlman 2002).
77 Furthermore, the interaction of wood polymers and differentiation of wood species using Partial
78 Least-Squares regression has also been investigated (Åkerholm et al. 2001; Hobro et al. 2010). A
79 recent review by Tsuchikawa (2007) summarizes important findings in the classification of wood
80 using NIR-FTR spectroscopy and multivariate techniques.

81

Chemometrics

83 Chemometrics is the science of extracting information from chemical systems with the application of
84 statistical and mathematical methods to chemical data (Gidman et al. 2003) and the most common
85 ones are based on dimensionality reduction via cluster analysis using Principal Component Analysis
86 (PCA) and Discriminant Function Analysis (DFA). Multivariate analysis builds on the application of
87 statistical and mathematical methods, and includes the analysis of data with many observed
88 variables, as well as the study of systems with many important types of variation (Gottlieb et al.
89 2004). Multivariate analysis is an inductive analysis, where hypotheses can be set up after having
90 carried out the computational experiments to identify outlying observations (Gottlieb et al. 2004),
91 clusters of similar observations and other data structures. The biochemical profiles of FTIR from
92 whole cell samples are extremely high-density data sets and, consequently, FTIR data must be
93 analyzed by means of multivariate analysis when multiple whole cell samples are compared (Kim et
94 al. 2004).

Phylogenetic classification of trees

96 Trees belongs to seed-bearing plants (Spermatophytae), which are subdivided into gymnosperms
97 (Gymnospermae) and angiosperms (Angiospermae) (Sjostrom 1993). Based on the APG II System
98 Classification, the fagales are an order of flowering plants, including some of the best known trees
99 and containing 8 families and approximately 30 genera and nearly 1000 species. They belong among
100 the rosids group of dicotyledons. Two of the families of this order are Fagaceae and Betulaceae. On
101 other hand, Malpighiales is one of the largest orders of flowering plants, containing about 16000
102 species, approximately 7.8% of the eudicots. Malpighiales is divided in 35 families by APG II system
103 (APG II 2003). One of them is Salicaceae. Malpighiales and Lamiales are the two large orders whose
104 phylogeny remains mostly unresolved. The order is very diverse and hard to recognize except with
105 molecular phylogenetic evidence.

106 The aim of this work is use the chemical composition of wood extracted from the FTIR data to
107 discriminate wood samples between order and family, taking groups from the current plant APG II
108 classification system.

Materials

110 Branch material was collected from 9 tree species in Lincoln (Lincolnshire, UK). All were members of
111 the rosid clade; 6 species were from the order fagales and 3 from the order malpighiales (**Table 2**).
112 Samples were stored in a dry environment at ambient temperature conditions.

Methods

Sample preparation

Samples, approximately 30 cm in length, were removed from branches from 9 species and then cut into transverse sections between 5 to 15 mm in thickness using a bandsaw. Two of the transverse sections were selected at random to represent each tree; it was assumed that there was no variation in the composition of wood along the branch. The two sections from each tree were taken as replicates of the same measure in the statistical analysis. Spectra were collected for a total of 256 scans with a resolution of 4 cm^{-1} between 4000 and 500 cm^{-1}. Spectra were acquired from six positions across the transverse section from the central pith to the outer bark region including: the pith, a ring close to the pith, a central ring, a ring close to the bark, the bark, and an arbitrary point in the sapwood with no visible rings. A total of 12 spectra were recorded for each tree species. The dataset obtained from a PerkinElmer Spectrum 100 FTIR Spectrometer was integrated by 3500 variables and 108 observations[1].

The data set was processed using Tanagra 1.4.39 software and analysed using PCA, One-Way ANOVA, Stepwise DA method, Partial-Least squares for Classification (C-PLS), Linear Discriminant Analysis (LDA) and PLS-LDA linear models. The dataset was split in 4 groups; each named according to the position in which spectrum was recorded: pith, bark, rings (this included all ring measures) and sapwood dataset. Each group dataset was used as input of a Stepwise DA method and classification functions were computed. Only vibrational spectra of wood samples recorded in the rings group could be used to discriminate between taxon.

Multivariate techniques

The raw spectra of 6 wood species that belong to fagales order and 3 wood species from malpighiales order were statistically analysed; sample sizes are given in **Table 3**.

Wavenumbers from the ring dataset were normalized and any detected outliers were removed. An initial exploratory analysis with PCA was followed by a pattern recognition procedure to distinguish the relations between taxa. PCA is commonly applied to spectroscopy data to study the main variability in the spectra. The factors were firstly rotated by VARIMAX method to facilitate the interpretation and secondly interpreted in terms of their chemical meaning. A One-Way ANOVA analysis was applied over the scores of the rotated principal components from PCA to highlight the most influent chemical features for grouping the data in fagales or malpighiales orders, as well as their family subgroups. Supervised pattern recognition was carried out after the exploratory analysis. Predictor variables, selected using a Stepdisc method, were related to the proposed group structure (APG II system classification) using mathematical linear models. Several linear models were computed and compared: PLS-LDA, LDA and C-PLS. This allowed the subsequent classification of unknown samples. The accuracy of the model was evaluated with validation methods such as Leave One Out (LOO), bootstrap method and an independent test set.

---

[1] Total number of observations: 108. Showed in the present work: 67 (45 from ring dataset and 22 from sapwood dataset)

Results and Discussion

Wood spectra data set

Vibrational spectra of wood samples recorded in the rings are shown for order and family classification in **Fig. 1**. The arrows indicate important bands in the discrimination of wood samples based on the Stepwise DA results. The very intimate mixing of the components makes it difficult to interpret the spectrum and no visual evidence in the spectra of wood can be found to discriminate among the groups (Fig. 1).

Exploratory analysis

A Principal Component Analysis (PCA) mathematical technique was applied to over 60 samples of individual spectra of trees to find the more relevant wavelengths, between the range 4000-500 cm^{-1}, which contribute to sample discrimination between Fagales and Malpighiales order and Fagaceae and Betulaceae family. The data set was standardized to avoid the variance of the variables affecting the principal components analysis. The factor rotated loading (FR) extracted from PCA were used for interpreting the principal components and to determine which variables are influential in the formation of PCs. Loading plots are shown in **Fig. 2** for order and family datasets, respectively. The higher the loading of a variable, the more influence it has in the formation of the factor and vice versa. The output from PCA is shown in **Table 4**.

A One-Way ANOVA analysis on the scores from PCA revealed one important factor which contains chemical differences to discriminate fagales from malpighiales orders and two relevant factors from family dataset with chemical information relevant for the classification (Table 5; $p \leq 0.05$).

In the case of order dataset, the third factor selected by the ANOVA is defined by the region 1701-1380 cm^{-1} considering only significant correlations higher than 70% as cut-off value. A detailed band assignment of the factor is given in **Table 6**. This region of the spectrum is probably associated with aromatic vibrations in lignin and cellulose, as well as some amide stretching and C=O vibrations as more significant contributions.

On other hand, in the case of family dataset the regions from 4000-3605 and 2819-1754 cm^{-1} contribute to the formation of the second factor with high significance. The assignment of this wavenumbers could not be made with certainty. Some CH stretching, not described in the literature, and the broad C=O stretching band are generally assigned to this region. The third rotated factor is defined by the wavenumbers from 1629-1179 cm^{-1}. The C-H vibrations of lignin, cellulose and hemicelluloses exhibit characteristic bands between 1511-1377 cm^{-1} and the OH in plane bending normally occurs in the region of 1455-1438 (Marchessault 1962). Bands attributed to guaiacyl around 1290 cm^{-1} (Anchukaitis et al. 2008) and cellulose at 1300 cm^{-1}, as well as non-cellulosic polysaccharides at the region 1214-1179 cm^{-1} are important contributions to the third factor in the case of family dataset. The remaining bands observed in this region are shown in **Table 6**.

Stepwise DA Analysis

10 biomarkers (1742, 1719, 1715, 1696, 1566, 1438, 1384, 1273, 1260 and 950 cm^{-1}) were successfully found discriminating from Malpighiales to Fagales order. In the IR spectra, it is known that these wavenumbers belong to C=O stretching in lignin and hemicelluloses and skeletal vibrations from pectin and cellulose (**Table 7**). The darker cells from **Table 7** indicate the wavenumbers highlighted in the ANOVA test described below. The univariate F-value measures the

variable's total discriminating power without considering how much might be shared by other variables (Klecka 1980). The wavenumbers are arranged in a descendent order based on their F-values. The greater contributor to the overall discrimination in the Stepwise method will show a better F-value (Klecka 1980). With regards to family dataset, 8 biomarkers (2595, 2203, 2055, 1997, 1936, 1928, 1916 and 1896 cm^{-1}) were successfully found discriminating between fagaceae and betulaceae families from fagales order. Differences between groups could not be easily assigned (**Table 7**). Kim et al. (2004) also used some of these wavenumbers when analysing some FTIR spectroscopy data in a taxonomic discrimination.

Classification functions

C-PLS, LDA and PLS-LDA linear models were computed. **Table 8** shows the classification functions with their statistical evaluation for order and family datasets. The coefficients of the classification functions are not interpreted. The darker cells represent the relative importance of the predictor from PLS method (variable important in projection indicator or VIP) in the classification with respect to the other variables (Rakotomalala 2005). Smallest lambda values or largest partial F means high discrimination (Klecka 1980). The significance of the difference was checked using MANOVA and two transformations of its lambda, Bartlett transformation and Rao transformation and there was a significant difference between groups (**Table 9**; P< 0.01). Error rate estimation is presented in **Table 10** using confusion matrix, cross-validation, bootstrap method, Leave-One-Out and external sample tests (test size appears in brakets).The bootstrap value shown in **Table 10** is the higher error obtained by the .632 estimator and its variant .632+. The model seems stable and with a low classification error. LDA gave the lowest error in the classification.

Conclusions

FTIR spectroscopy in combination with PCA and linear model such as LDA were suitable techniques to discriminate wood samples based on the APG II Classification. Ten biomarkers seemed to discriminate fagales from malpighiales order with high significance and low classification error. Concerning to the discrimination of fagaceae from betulaceae families from fagales order, 8 biomarkers were highlighted for sample discrimination. With the inclusion of new wood samples it is likely that new chemical features will emerge. In the context of the present work, multivariate analysis has highlighted the chemical differences in hemicelluloses, cellulose and guaiacyl (lignin) in the order dataset, but that this is less evident in the family dataset. Our results suggest that differences in cell-wall composition and structure can provide the basis for chemotaxonomy of flowering plants.

Acknowledgements

References

Åkerholm M, Salmén L, Salme L (2001) Interactions between wood polymers studied by dynamic FT-IR spectroscopy. Polymer 42:963–969. doi: 10.1016/S0032-3861(00)00434-1

228 Anchukaitis KJ, Evans MN, Lange T, et al. (2008) Consequences of a rapid cellulose extraction
229   technique for oxygen isotope and radiocarbon analyses. Analytical chemistry 80:2035–2041.
230   doi: 10.1016/j.gca.2004.01.006.Analytical

231 APG I (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families
232   of flowering plants: APG II. Botanical Journal of the Linnean Society 141:399–436. doi:
233   10.1046/j.1095-8339.2003.t01-1-00158.x

234 Barnett JR, Jeronimidis G (2003) Wood quality and its biological basis. 226.

235 Bjarnestad S, Dahlman O (2002) Chemical compositions of hardwood and softwood pulps employing
236   photoacoustic Fourier transform infrared spectroscopy in combination with partial least-
237   squares analysis. Analytical chemistry 74:5851–5858.

238 Brinkmann K, Blaschke L, Polle A (2002) Comparison of different methods for lignin determination as
239   a basis for calibration of near-infrared reflectance spectroscopy and implications of
240   lignoproteins. Journal of chemical ecology 28:2483–501.

241 Brunner M, Eugster R, Trenka E, Bergamin-Strotz L (1996) FT-NIR spectroscopy and wood
242   identification. Holzforschung 50:130–134.

243 Callow JA, Andrews JH, Tommerup IC (2006) Advances in Botanical Research, Volume 21. 304.

244 Coates J (2000) Interpretation of infrared spectra, a practical approach. Encyclopedia of analytical
245   chemistry 10815–10837.

246 Ek M, Gellerstedt G, Henriksson G (2009) Wood Chemistry and Wood Biotechnology. 308.

247 Gidman E, Goodacre R, Emmett B, et al. (2003) Investigating plant–plant interference by metabolic
248   fingerprinting. Phytochemistry 63:705–710. doi: 10.1016/S0031-9422(03)00288-7

249 Gorgulu ST, Dogan M, Severcan F (2007) The characterization and differentiation of higher plants by
250   fourier transform infrared spectroscopy. Applied spectroscopy 61:300–8. doi:
251   10.1366/000370207780220903

252 Gottlieb DM, Schultz J, Bruun SW, et al. (2004) Multivariate approaches in plant science.
253   Phytochemistry 65:1531–48. doi: 10.1016/j.phytochem.2004.04.008

254 Heinze T, Liebert T, Koschella A (2006) Esterification of polysaccharides. 232.

255 Hobro A, Kuligowski J, Döll M, Lendl B (2010) Differentiation of walnut wood species and steam
256   treatment using ATR-FTIR and partial least squares discriminant analysis (PLS-DA). Analytical
257   and bioanalytical chemistry 398:2713–22. doi: 10.1007/s00216-010-4199-1

258 Huang A, Zhou Q, Liu J, et al. (2008) Distinction of three wood species by Fourier transform infrared
259   spectroscopy and two-dimensional correlation IR spectroscopy. Journal of Molecular Structure
260   883-884:160–166. doi: 10.1016/j.molstruc.2007.11.061

261 Kacuráková M, Kauráková M, Capek P, et al. (2000) FT-IR study of plant cell wall model compounds:
262   pectic polysaccharides and hemicelluloses. Carbohydrate Polymers 43:195–203. doi:
263   10.1016/S0144-8617(00)00151-X

264 Kim SW, Ban SH, Chung HJ, et al. (2004) Taxonomic discrimination of flowering plants by multivariate
265   analysis of Fourier transform infrared spectroscopy data. Plant cell reports 23:246–50. doi:
266   10.1007/s00299-004-0811-1

267 Klecka WR (1980) Discriminant analysis, Issue 19. 71.

268 Kubo S, Kadla JF (2005) Hydrogen bonding in lignin: a Fourier transform infrared model compound
269   study. Biomacromolecules 6:2815–21. doi: 10.1021/bm050288q

270 Larkin P (2011) Infrared and Raman Spectroscopy; Principles and Spectral Interpretation. 230.

271 Liang CY, Marchessault RH (1959) Infrared spectra of crystalline polysaccharides. II. Native celluloses
272   in the region from 640 to 1700 cm. Journal of Polymer Science XXXIX:269–278.

273 Marchessault R (1962) Application of infra-red spectroscopy to cellulose and wood polysaccharides.
274   Pure Appl. Chem

275 Marchessault RH, Liang CY (1962) The infrared spectra of crystalline polysaccharides. VIII. Xylans.
276   Journal of Polymer Science 59:357–378. doi: 10.1002/pol.1962.1205916813

277 Martin JW (2006) Concise encyclopedia of the structure of materials. 20:512.

278 McCann MC, Bush M, Milioni D, et al. (2001) Approaches to understanding the functional
279   architecture of the plant cell wall. Phytochemistry 57:811–21.

280 Meinzer FC, Lachenbruch B, Dawson TE (2011) Size- and Age-Related Changes in Tree Structure and
281   Function. 510.

282 Mohebby B (2008) Application of ATR Infrared Spectroscopy in Wood Acetylation. J Agric Sci 10:253–
283   259.

284 Mohebby B (2005) Attenuated total reflection infrared spectroscopy of white-rot decayed beech
285   wood. International Biodeterioration & Biodegradation 55:247–251. doi:
286   10.1016/j.ibiod.2005.01.003

287 Nuopponen M (2005) FT-IR and UV Raman spectroscopic studies on thermal modification of Scots
288   pine wood and its extractable compounds. Technology

289 Obst JR (1982) Guaiacyl and Syringyl Lignin Composition in Hardwood Cell Components.
290   Holzforschung 36:143–152. doi: 10.1515/hfsg.1982.36.3.143

291 Pandey KK, Vuorinen T (2008) Comparative study of photodegradation of wood by a UV laser and a
292   xenon light source. Polymer Degradation and Stability 93:2138–2146. doi:
293   10.1016/j.polymdegradstab.2008.08.013

294 Rana R, Langenfeld-Heyser R, Finkeldey R, Polle A (2009) FTIR spectroscopy, chemical and
295   histochemical characterisation of wood and lignin of five tropical timber wood species of the
296   family of Dipterocarpaceae. Wood Science and Technology 44:225–242. doi: 10.1007/s00226-
297   009-0281-2

298 Rana R, Sciences F (2008) Correlation between anatomical/chemical wood properties and genetic
299   markers as a means of wood certification.

300 Revanappa SB, Nandini CD, Salimath PV (2010) Structural characterisation of pentosans from
301   hemicellulose B of wheat varieties with varying chapati-making quality. Food Chemistry
302   119:27–33. doi: 10.1016/j.foodchem.2009.04.064

303 Rhoads CAC, Painter P, Given P (1987) FTIR studies of the contributions of plant polymers to coal
304   formation. International journal of coal geology 8:69–83.

305 Sandak A, Sandak J, Negri M (2010) Relationship between near-infrared (NIR) spectra and the
306   geographical provenance of timber. Wood Science and Technology 45:35–48. doi:
307   10.1007/s00226-010-0313-y

308 Shen JB, Lu HF, Peng QF, et al. (2008) FTIR spectra of Camellia sect. Oleifera, sect. Paracamellia, and
309   sect. Camellia (Theaceae) with reference to their taxonomic significance. plantsystematics.com
310   46:194–204. doi: 10.3724/SP.J.1002.2008.07125

311 Silverstein RM, Webster FX, Kiemle D (2005) Spectrometric identification of organic compounds.
312   502.

313 Sjostrom (1993) Wood chemistry: fundamentals and applications. 293.

314 Stewart D, Wilson HM, Hendra PJ, Morrison IMIM (1995) Fourier-Transform Infrared and Raman
315   Spectroscopic Study of Biochemical and Chemical Treatments of Oak Wood (Quercus rubra)
316   and Barley (Hordeum vulgare) Straw. Journal of agricultural and food chemistry 43:2219–2225.
317   doi: 10.1021/jf00056a047

318 Stuart B (2004) Infrared spectroscopy: fundamentals and applications. 224.

319 Takayama M (1997) Fourier transform Raman assignment of guaiacyl and syringyl marker bands for
320   lignin determination. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy
321   53:1621–1628. doi: 10.1016/S1386-1425(97)00100-5

322 Tsuchikawa S (2007) A Review of Recent Near Infrared Research for Wood and Paper. Applied
323   Spectroscopy Reviews 42:43–71. doi: 10.1080/05704920601036707

324 Wang S, Wang K, Liu Q, et al. (2009) Comparison of the pyrolysis behavior of lignins from different
325   tree species. Biotechnology advances 27:562–7. doi: 10.1016/j.biotechadv.2009.04.010

326 Wellner N (1998) FT-IR study of pectate and pectinate gels formed by divalent cations. Carbohydrate
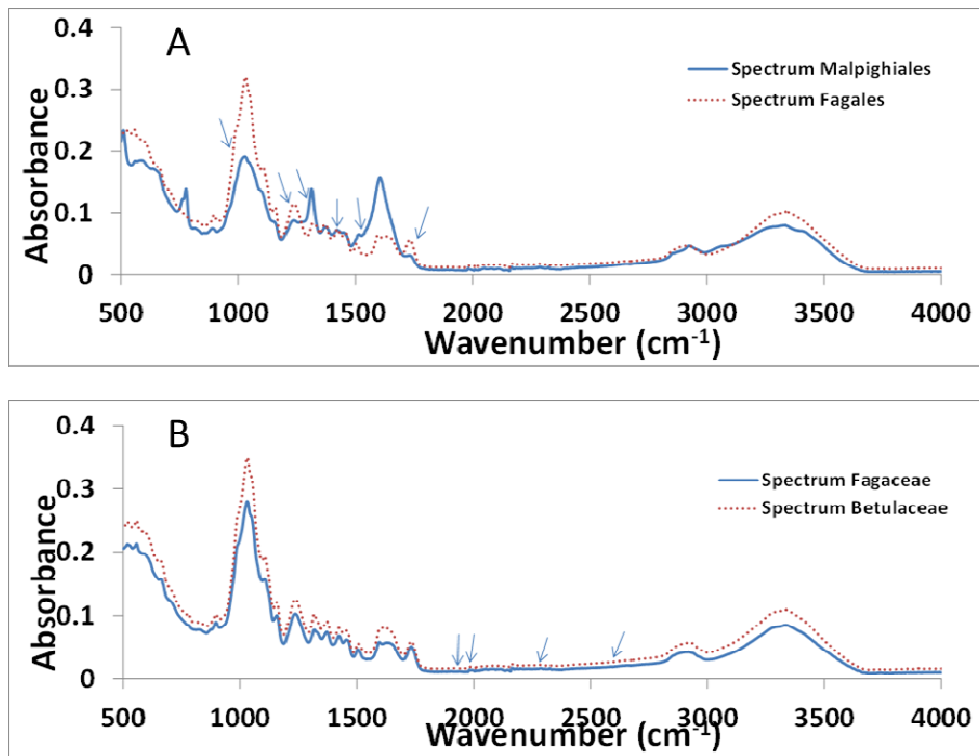327   Research 308:123–131. doi: 10.1016/S0008-6215(98)00065-2

List of Figures

List of Tables

Figures

Fig. 1 Average FTIR spectrum of **(a)** Order: Malpighiales versus Fagales and **(b)** Family: Fagaceae versus Betulaceae samples

349

Fig. 2 PC loading plot showing the correlation plot of the variables with the factor loadings orthogonally rotated from **(a)** Order and **(b)** Family dataset. Average FTIR spectrum is also shown.
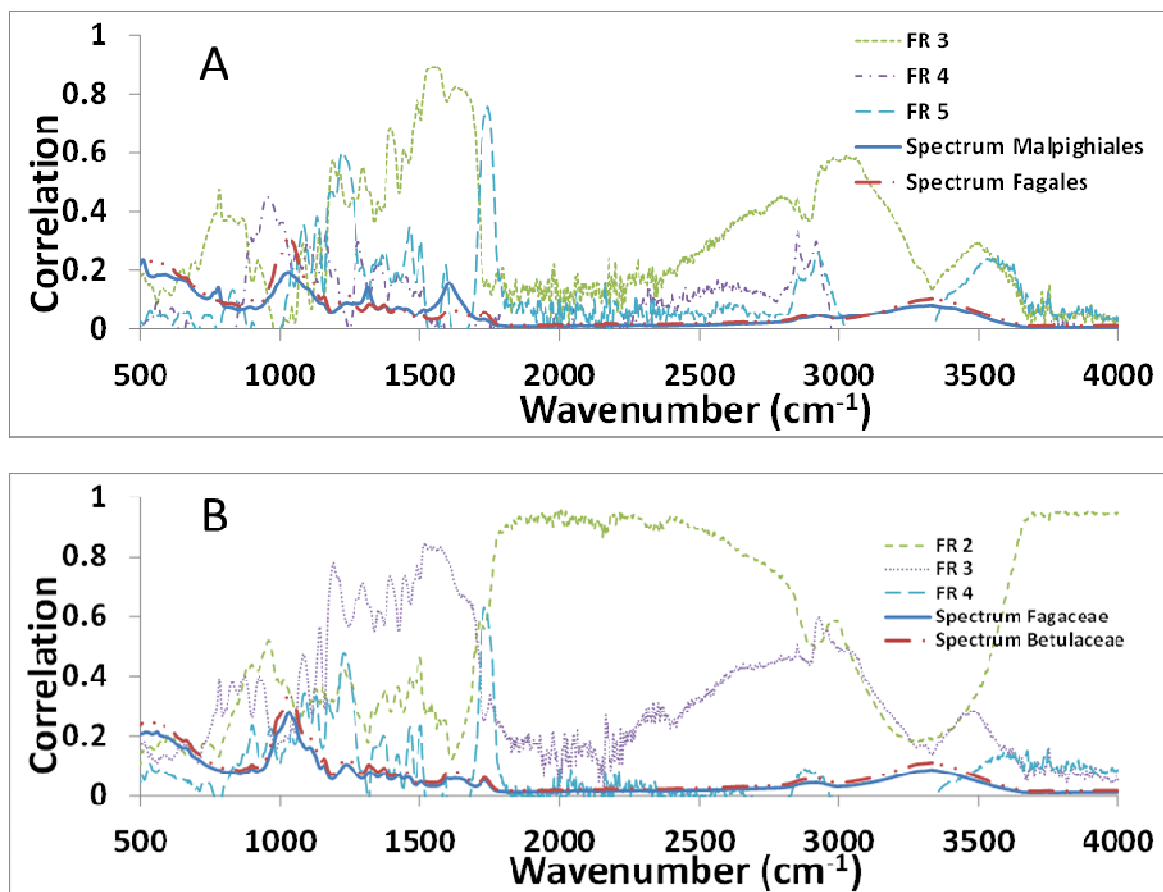
Tables

Table 1 Recent bibliography related of chemical taxonomy of wood using FTIR spectroscopy and Multivariate analysis

Analytical approach	Objective	Conclusions	Ref
FT-IR	Investigate the changes in composition and structure of oak wood and barley straw	Good results in the analysis of plant fibers and cell walls	(Stewart et al. 1995)
FT-IR/ Multiv. analysis	Taxonomic classification of 65 leaf samples of genus Camellia (Ericales order)	Identification and classification of species in the same genus	(Shen et al. 2008)
FT-IR/ Multiv. analysis	Characterization of 5 dipterocarp wood species of malvales order and their aisolated lignins	Measure of the G/S ratio to distinct between species.	(Rana et al. 2009)
FT-IR/ Multiv. analysis	Inter-species competition between a monocotyledon and a dicotyledon by growth analysis	Detecting changes in the global metabolic profiles of plants	(Gidman et al. 2003)

FT-IR/ Multivar. analysis	Analysis of whole cell of leaf samples of 7 higher plant species	· Determination of 3 to 5 biomarkers that discriminated plants in the carbohydrate fingerprint region (1200-800 cm-1) · Discrimination between monocot and dicotyledonous plants · FT-IR reflects phylogenetic relationships between plants · The differences in cell-wall composition and estructure can provide the basis for chemotaxonomy of flowering plants	(Kim et al. 2004)
FT-IR/ Multivar. analysis	Study of polisaccharides and hemicelluloses extracted from plants	The region at 1200-800 cm-1 where shown to be useful for the identification of polysaccharides with different structure and composition	(Kacuráková et al. 2000)
FT-IR/ Multivar. analysis	Differenciation of plants based in their leaf fragments technique applied to 3 different genera: Ranunculus (Ranunculales order), Acantholimon(Caryophyl lales order)and Atragalus(Fabales order)	FT-IR revealed dramatic difference between genera in lipid metabolism, carbohydrate composition and protein conformation. Succeed in molecular characterization and identification of plants.	(Gorgulu et al. 2007)
FTIR/Multivar. analysis	Differentiation of wood species	It is possible to differentiate between samples of wood species of different origins	(Brunner et al. 1996)
FT-NIR	Determination of the origin of several Spruce samples	Trees growing in different locations have different chemical composition and can be distingued by using FTIR and multivatiate analysis	(Sandak et al. 2010)

354   **Table 2** Wood samples names based on APG II System Classification

Order	Family	Genus	Species	Common name
Malpighiales	Salicaceae	Populus Populus Salix	*Populus* *Poplar nigra* *Salix fragilis*	Poplar Black Poplar Willow
Fagales	Betulaceae	Alnus M. Corylus L. Betula L.	*Alnus glutinosa* *Corylus avellana* *Betula pubescens*	Black Alder Hazel Birch
	Fagaceae	Castanea Fagus L. Quercus	*Castanea sativa* *Fagus sylvatica* *Quercus robur*	Sweet Chesnut Beech English Oak

355   **Table 3** Sample size available for chemometric analysis

Rings position	Sample size	
	Order (45)	Family (27)
	fagales:27 malpighiales:18	fagaceae:12 betulaceae:15
	Training set: 37 Test set: 8	Training set: 20 Test set: 7

356   **Table 4** PCA summary of factor loading

	Order			Family		
Factor	Eigen value	% explained	% cumulated	Eigen value	% explained	% cumulated
1	24447.53	69.91	69.91	2559.140	73.10	73.10
2	626.618	17.90	87.81	588.9263	16.82	89.92
3	138.49	3.96	91.76	142.1563	4.06	93.98
4	99.3673	2.84	94.60	66.20825	1.89	95.87
5	66.8034	1.91	96.51	59.16276	1.69	97.56

357 **Table 5** Output from One-Way ANOVA analysis

	Value	Examples	Average	Std-dev	Variance decomposition			Significance level		
					Source	Sum of square	d.f.	Statistics	Value	Proba
FR3	**Fagales**	27	0.0329	0.0661	BSS	0.0729	1			
	Malpighiales	18	-0.0493	0.0938	WSS	0.2631	43	Fisher's F	11.915489	0.001261
	All	45	0	0.0874	TSS	0.336	44			
FR2	**Fagaceae**	12	-9.865	15.3611	BSS	2100.3693	1			
	Betulaceae	15	7.8848	22.4355	WSS	9642.5285	25	Fisher's F	5.445587	0.027957
	All	27	-0.004	21.2521	TSS	11742.898	26			
FR3	**Fagaceae**	12	-7.6125	8.5755	BSS	1250.871	1			
	Betulaceae	15	6.0853	12.455	WSS	2980.7246	25			
	All	27	-0.0026	12.7575	TSS	4231.5956	26	Fisher's F	10.491333	0.003376

358 **Table 6** Band assignments of factor rotated loadings from PCA highlighted in the ANOVA analysis

FR	v (cm^{-1})	Literature assignments and Band origin	Ref.
Order			
3	1701-1380	1645-1612 C-O stretching of conjugated or aromatic ketones, C=O stretching in flavones	(Hobro et al. 2010; Huang et al. 2008)
		1640-1630 water	(Liang and Marchessault 1959; Marchessault and Liang 1962; Revanappa et al. 2010)
		1650, 1555 C=N and N-H stretching from amides, mainly proteins	(Gorgulu et al. 2007)
		1610-1600, 1513-1502 aromatic skeletal vibration lignin	(Bjarnestad and Dahlman 2002; Hobro et al. 2010; Huang et al. 2008; Kubo and Kadla 2005; Marchessault 1962; Pandey and Vuorinen 2008; Rana et al. 2009; Rana and Sciences 2008; Stewart et al. 1995; Wang et al. 2009)
		1427,1425,1420 aromatic ring vibration combined with C-H in-plane deformation lignin	(Kubo and Kadla 2005; Rhoads et al. 1987; Wang et al. 2009)
		1455 OH in-plane bending cellulose	(Heinze et al. 2006; Liang and Marchessault 1959)
		1430 CH2 bending cellulose	(Liang and Marchessault 1959; Rhoads et al. 1987)
		1428-1416 CH2 scissors vibrations in cellulose, aromatic skeletal vibration	(Heinze et al. 2006; Hobro et al. 2010)
Family			
2	4000-3605	X-H stretching	(Stuart 2004)
	2820-2573	C-H stretching	(Stuart 2004)
	2565-1774	C=O stretching	(Heinze et al. 2006)
	2350-2340	CO2	(Hobro et al. 2010)
	1773-1771	xylan	(Huang et al. 2008; Mohebby 2005)
3	1511-1377	1511-1502 aromatic ring vibration lignin	(Bjarnestad and Dahlman 2002; Hobro et al. 2010; Huang et al. 2008; Marchessault 1962; Pandey and Vuorinen 2008; Rana et al. 2009; Rana and Sciences 2008; Stewart et al. 1995; Wang et al. 2009)
		1470-1455 aromatic C-H deformation asymmetric in -CH3 and -CH2- pyran ring in	(Hobro et al. 2010; Kubo and Kadla 2005; Marchessault 1962; Pandey and Vuorinen 2008;

Best ν (cm⁻¹)	Band assignment	Reference
	lignin and xylan	Rana and Sciences 2008; Rhoads et al. 1987; Wang et al. 2009)
	1427,1425,1420 aromatic ring vibration combined with C-H in-plane deformation lignin	(Kubo and Kadla 2005; Rhoads et al. 1987; Wang et al. 2009)
	1428-1416 CH2 scissors vibrations in cellulose, aromatic skeletal vibration	(Heinze et al. 2006; Hobro et al. 2010)
1376-1195	1330-1315 C-O of Syringyl ring	(Hobro et al. 2010; Kubo and Kadla 2005; Mohebby 2005; Obst 1982; Pandey and Vuorinen 2008; Rana et al. 2009; Rana and Sciences 2008; Rhoads et al. 1987; Wang et al. 2009)
	1600 C-O antisymmetric stretching glucuronic acid (xylan)	(Marchessault and Liang 1962)
	1375,1374,1372 C-H bending (CH3 skeletal) cellulose	(Liang and Marchessault 1959; Rhoads et al. 1987)
	1350,1245, 1215 OH in-plane bending xylan	(Marchessault and Liang 1962)
	1336,1335 CH2 and C-OH in-plane bending cellulose	(Liang and Marchessault 1959; Rhoads et al. 1987)
	1318,1317 CH2 wagging crystalline cellulose	(Liang and Marchessault 1959; Rhoads et al. 1987)
	1270-1268 C=O of Guaiacyl ring vibration in lignin	(Hobro et al. 2010; Huang et al. 2008; Marchessault 1962; Nuopponen 2005; Obst 1982; Rana et al. 2009; Rana and Sciences 2008; Rhoads et al. 1987; Wang et al. 2009)

359 **Table 7** Best discriminator variables based on a Stepwise estimation from order and family datasets

Best ν (cm⁻¹)	F-value	Percentage in each Rotated Factor (FR)	Band assignment	Reference
Order				
1742	F:35.50 p:0.0000	FR5 (57%), FR2 (30%)	1740-1730 C=O stretching in acetyl groups of hemicelluloses (xylan/glucomannan)	(Åkerholm et al. 2001; Bjarnestad and Dahlman 2002; Gorgulu et al. 2007; Marchessault 1962; McCann et al. 2001; Mohebby 2008; Mohebby 2005; Rana et al. 2009; Stewart et al. 1995)
1384	F:35.23 p:0.0000	FR1 (49%), FR3 (37%)	C-H bending deformation polysaccharide band (cellulose and hemicellulose)	(Marchessault 1962; Marchessault and Liang 1962; Mohebby 2005; Pandey and Vuorinen 2008; Rana and Sciences 2008)
1696	F:24.32 p:0.0000	FR3 (47%), FR1 (33%)	C=O stretching	(Coates 2000; Silverstein et al. 2005)
1719, 1715	F:19.57 p:0.0001 F:9.38 p:0.0040	FR5 (38%), FR2 (30%) FR2 (30%), FR5 (28%)	1711 C=O stretch (unconjugated) in lignin	(Hobro et al. 2010)
950	F:9.86 p:0.0031	FR1 (64%), FR4 (20%)	C-H ring pectin	(Kacuráková et al. 2000; Wellner 1998)
1566	F:9.77 p:0.0033	FR3 (79%), FR1 (18%)	1555 Amide II (C=N and N–H stretching): mainly proteins	(Gorgulu et al. 2007) 21

1273	F:9.03 p:0.0047	FR1 (56%), FR3 (19%)	1270-1268 C=O of Guaiacyl ring vibration in lignin	(Hobro et al. 2010; Huang et al. 2008; Marchessault 1962; Nuopponen 2005; Obst 1982; Rana et al. 2009; Rana and Sciences 2008; Rhoads et al. 1987; Wang et al. 2009)
1438	F:6.15 p:0.0181	FR1 (50%), FR3 (35%)	aromatic skeletal vibrations combined with CH2 symmetrical bending mode of hydroxymethyl in crystalline cellulose	(Åkerholm et al. 2001; Anchukaitis et al. 2008; Bjarnestad and Dahlman 2002; Marchessault 1962; Rana and Sciences 2008)
1260	F:5.62 p:0.0235	FR1 (46%), FR5 (22%)	1270-1268 C=O of Guaiacyl ring vibration in lignin C-O (ester linkages)	(Hobro et al. 2010; Huang et al. 2008; Marchessault 1962; Nuopponen 2005; Obst 1982; Rana et al. 2009; Rana and Sciences 2008; Rhoads et al. 1987; Wang et al. 2009)
Family				
2203	F:32.50 p:0.0000	FR2 (83%), FR1 (11%)	C=O stretching	(Stuart 2004)
1936	F:23.49 p:0.0001	FR2 (89%), FR1 (6%)	-C=C=CH2 stretching	(Larkin 2011)
2595	F:22.08 p:0.0001	FR2 (71%), FR3 (40%)	S-H stretch	(Coates 2000; Silverstein et al. 2005)
1997	F:16.32 p:0.0007	FR2 (88%), FR1 (6%)	Allenes, ketenes, isocyanates, isothiocyanates	
1916	F:12.76 p:0.0016	FR2 (86%), FR1 (7%)	2nd overtone CONH O-H stretch	(Brinkmann et al. 2002)
2055	F:11.17 p:0.0033	FR2 (87%), FR1 (9%)	-NCS isothiocyanate or transition metal carbonyl	(Coates 2000)
1928	F:11.00 p:0.0033	FR2 (87%), FR1 (6%)	1st overtone P-OH C=O stretch	(Brinkmann et al. 2002)
1896	F:9.05 p:0.0075	FR2 (87%), FR1 (7%)	transition metal carbonyl	(Coates 2000)

360 **Table 8** Classification functions for Betulaceae from order and family models

	Classification functions			Statistical evaluation				
Descriptors	PLS-LDA	PLS	LDA	VIP	Wilks L.	Partial L.	F(1,5)	p-value
Order								
1742	24.5771	0.6958	47.0101	1.0621	0.100925	0.348437	63.57873	0.00000
1719	-25.2183	-0.7140	1.5717	1.0803	0.035177	0.999686	0.01067	0.91835
1715	-14.1490	-0.4006	-76.1243	0.8381	0.082211	0.427755	45.48466	0.00000
1696	14.6019	0.4134	42.9788	0.9556	0.050833	0.6918	15.14716	0.00044
1566	5.6683	0.1605	6.2828	1.0177	0.035861	0.980616	0.67207	0.41804
1438	-16.4909	-0.4669	-27.7396	0.9082	0.045834	0.767255	10.31382	0.00288
1384	-15.7865	-0.4469	-31.3282	0.7091	0.040389	0.870684	5.04976	0.03123
1273	29.7072	0.8410	90.2861	1.1234	0.066927	0.525439	30.70779	0.00000
1260	-2.0519	-0.0581	-41.9669	0.7396	0.040981	0.858109	5.62201	0.02355
950	4.7910	0.1356	4.8394	1.3844	0.037743	0.931713	2.49193	0.12369
constant	-10.7958	0.0000	-20.5787	-				
Family								
2595	30.705272	0.685431	51.63692	1.709	0.039933	0.31961	23.41697	0.00052
2203	-44.716878	-0.998211	-106.72309	1.1159	0.058034	0.219921	39.01791	0.000063
2055	3.987204	0.089006	39.623592	0.5761	0.018318	0.69673	4.78805	0.051132
1997	-29.4359	-0.657095	-58.627727	1.1446	0.022989	0.555184	8.81324	0.012775
1936	-15.100637	-0.33709	-164.08719	0.8776	0.049105	0.259913	31.32187	0.000161
1928	12.499757	0.279031	80.408548	0.5761	0.024477	0.521421	10.0962	0.008801
1916	31.208287	0.69666	148.63912	0.7796	0.047677	0.267698	30.09116	0.00019
1896	23.15418	0.516869	43.218349	0.6944	0.015201	0.839605	2.1014	0.175069
constant	-10.367996	-0.001317	-28.186321	-				

361 **Table 9** MANOVA analysis

Stat	Order		Family	
	Value	p-value	Value	p-value
Wilks' Lambda	0.0352	-	0.0128	-
Bartlett -- C(7)	127.2116	0.0000	61.057	0.0000
Rao -- F(7, 75)	93.2842	0.0000	106.359	0.0000

362 **Table 10** Validation of the order and family models with ring and sapwood samples

Order

Sapwood sample test (12)

Method	Confusion matrix (%)	Cross-validation	Bootstrap	Leave-One-Out	Independent Test set
PLS-LDA	0.0000	0.0000	0.0191	0.0000	0.0000
PLS	0.0000	0.0400	0.0487	0.0175	0.0833
LDA	0.0000	0.0000	0.0025	0.0000	0.0000

Rings sample test (8)

Method	Confusion matrix(%)	Cross-validation	Bootstrap	Leave-One-Out	Independent Test set
PLS-LDA	0.0000	0.0000	0.0148	0.0000	0.0000
PLS	0.0000	0.0000	0.0216	0.0000	0.0000
LDA	0.0000	0.0000	0.0016	0.0000	0.0000

Family

Sapwood sample test (10)

Method	Confusion matrix(%)	Cross-validation	Bootstrap	Leave-One-Out	Independent Test set
PLS-LDA	0.0000	0.0333	0.0809	0.0811	0.3000
C-PLS	0.0000	0.0333	0.0658	0.0811	0.3000
LDA	0.0000	0.0333	0.0809	0.0811	0.3000

Rings sample test (7)

Method	Confusion matrix(%)	Cross-validation	Bootstrap	Leave-One-Out	Test set
PLS-LDA	0.0000	0.0000	0.0107	0.0000	0.0000
PLS	0.0000	0.0000	0.0107	0.0000	0.0000
LDA	0.0000	0.0000	0.0026	0.0000	0.0000

363