

Chemical relationships of Ambers using ATR-FTIR spectroscopy

L. J. Cotton^{1,2*}, F. Vollrath³, M. D. Brasier¹ and C. Dicko^{3,4*}

¹ *Oxford University, Dept. of Earth Sciences, South Parks road, Oxford, OX1 3AN, UK.*

² *Current affiliation: School of Biological Sciences, The University of Hong Kong, Kadoorie Biological Sciences Building, Pokfulam Road, Hong Kong SAR, China*

³ *Oxford University, Dept. of Zoology, South parks road, Oxford, OX1 3PS, UK.*

⁴ *Current affiliation: Chemistry Department, Pure and Applied Biochemistry, Lund University, Naturevetarvägen 16, 22241 Lund, Sweden*

**Correspondence: lj cotton@hku.hk, cedric.dicko@tbiokem.lth.se*

Abstract

Amber is known for its remarkably well-preserved fossils, less known is amber chemical complexity and its history. Amber is highly variable in both its physical and chemical properties, dependent on factors such as source tree and diagenetic history. Therefore, amber of a given locality has a unique chemical composition. Fourier Transform Infra-Red (FTIR) Spectroscopy is a technique often used to determine the chemical composition and provide a “fingerprint” of amber samples. Here we carry out FTIR analysis of samples from 17 localities from the Americas, Europe and Asia and spanning the Lower Cretaceous to Oligo-Miocene followed by cluster analysis to examine trends in amber chemistry and increase understanding of its formation through time. A detailed analysis of the clustering followed by modelling of the variables of importance suggests that the exocyclic methylene group content and conformation play a major role in explaining the clustering. Other variables, such as

ester and alkyl contents contribute to the identification. Put in a broader perspective, our study indicates the dominant factor in clustering is the age of the amber, followed by locality.

Keywords: amber, FTIR, cluster analysis, exocyclic methylene

Introduction

Fossilised tree resin, otherwise known as amber, has long been prized by humans as gemstone as well as for its remarkable fossils and the quality of their preservation. Amber forms in the terrestrial environment and samples a niche otherwise almost absent from the fossil record. It preserves fine structures, interspecies relationships and even DNA (Cano et al. 1993; Desalle et al. 1993; Poinar et al. 1993) thus creating snapshots in time – even to the extent of inspiring “Jurassic Park” scenarios. Numerous studies have been carried out on fossil inclusions providing a unique window into the evolution and development of arthropod and other terrestrial faunas and floras through time. However, questions remain about the amber resin itself.

Amber has a geological record spanning back to the late Palaeozoic and encompassing localities worldwide. The oldest known amber is Upper Carboniferous in age (Van Bergen et al. 1995), and is exceedingly rare. Small droplets of Triassic amber have also been found within the Italian Dolomites (Schmidt et al. 2006). But the turning point for amber fossilisation seems to have been in the Cretaceous with increasing numbers of specimen being documented. Lower Cretaceous ambers remain rare, known fossiliferous deposits including the Lebanese deposits at Jezzine (Schlee & Dietrich 1970), material from Israel (Nissenbaum 1975) and Jordan (Bandel & Vavra 1981) as well as the Isle of Wight, UK (Jarzembowski 1999) (Nicholas et al. 1993). Amber became more common towards the middle of the Cretaceous with deposits reported from several localities in France (Neraudeau et al.

51 2002; Perrichot 2004; Breton 2007) and Spain (Alonso et al. 2000; Delclos et al.
52 2007; Penalver et al. 2007), and in the Crato formation of Brazil (Martill et al. 2005).
53 Amber occurrences continued to increase in the Upper Cretaceous with a global
54 distribution including sites in the USA, Asia and Europe (Waggoner 1994; Grimaldi
55 1996; Grimaldi et al. 2002; Penney 2002; Katagiri et al. 2013). This continued into
56 the Cenozoic, to which some of the most prolific amber bearing deposits, the Baltic
57 and Dominican, belonged.

58 Amber from these deposits shows a wide variety of physical properties, varying even
59 for specimens presumed to be of the same age from the same location.. Its chemistry
60 is also complex. The chemical composition of amber is related both to the original
61 resin of the producer tree, and the geological and diagenetic processes to which it has
62 been subjected (Angelini & Bellintani 2005). Given the disparity of amber deposits
63 through both time and geographic distribution it is most improbable that a single tree
64 genus or indeed taxon was responsible for all ambers. It is also likely that some
65 producer trees are now extinct. However, several studies indicate extant conifers, the
66 *Araucariaceae*, as one of the likely sources (Delclos et al. 2004; Poinar & Poinar
67 2008). This family became increasingly abundant in the Cretaceous with a
68 corresponding increase in amber deposits. In addition to taxonomic origin, the
69 chemistry of amber also depends on diagenesis and the geological history of a region.
70 After burial the resin is subjected to heat and pressure on a timescale of millions of
71 years altering its chemistry through additional polymerization as well as isomerisation
72 reactions, crosslinking and cyclisation (Vavra 2009). The specific processes which
73 take place vary regionally and across geological time. Therefore, amber from different
74 localities frequently shows variations in chemical composition (Lambert et al. 2008).

75 Though they are not true minerals due to their lack of regular definable crystal
76 structure, amber specimens can be classified chemically (Anderson 2006) and studies
77 have shown that the majority fall into a small number of categories; Class I to V
78 (Lambert et al. 2008). The most common of these are Class I ambers which have a
79 macromolecular structure based on polymers of labdanoid diterpenes. Class I ambers
80 are then further subdivided into subclasses a-d. Many Northern European types of
81 amber belong to this class, including Baltic amber (Lambert et al., 2008). This class of
82 amber is thought to have evolved early and continued through geological time. Class
83 II ambers are defined by the presence of polymers of sesquiterpenoid hydrocarbons
84 related to cadinene. This class is the second most globally common type of amber,
85 and is particularly common in South-East Asia as well as the western and southern
86 regions of the U.S.A (Lambert et al., 2008). Class III ambers are polystyrenes and
87 occur less commonly; they have only been identified from Germany and the Atlantic
88 coastal plain of the U.S.A (Lambert et al., 2008). Class IV and V ambers are generally
89 rare, composed of cedranes, abietene, pimarane and isopimarane and diterpenoids,
90 and are softer than those of classes I-III. They are therefore not often preserved
91 (Lambert et al., 2008).

92 One of the techniques used in studies of amber is Fourier Transform Infra-Red
93 (FTIR) spectroscopy. This method is commonly used to characterise the organic
94 chemistry of ambers of a certain locality (Nicholas et al. 1993; Alonso et al. 2000;
95 Delclos et al. 2007) and is frequently applied to questions such as identifying the
96 source region of amber archaeological materials (Teodor et al. 2009; Litescu et al.
97 2012) or un-localised amber specimens as well as defining some of the chemical
98 groups present. Infrared spectroscopy has the key advantages that it is relatively
99 simple to carry out and that most molecules present a unique spectrum. The technique

measures the absorption of infrared radiation by the amber, which depends on atomic mass, length strength and force constants of bonds and the overall structure of the material (Russell & Fraser 1994), altogether creating a composite chemical “fingerprint” of the material. Concomitant to the “fingerprint” advantage is the challenge of sorting the spectroscopic information into relevant biological information (Trevisan et al. 2012).

To address this challenge, we carried out FTIR analysis of ambers from a variety of global localities, spanning the Cretaceous and Tertiary, followed by cluster analysis on the most representative parts of the spectra. The aim of this study was not to determine the overall chemistry of the amber but to assess whether this type of study could be used to identify similarities between amber deposits, and potential age-based or regional clustering. We also introduce a modelling method to identify the variable of most importance and to link the observed clusters to chemical groups, amber formation processes and amber evolution through time.

Material and Methods

A total of 68 specimens were analysed from 17 different amber localities (see Table 1), including the Americas, Asia and several sites in Europe and spanning Berriasian (~145.0 Ma) to Oligo-Miocene (33.9 – 5.3 Ma). Where there was enough material, up to 4 replicates of the same specimen were run (see supplementary Table 1 with details of samples).

Small samples of the amber specimens were removed (~1-2mm) and ground to a fine powder using an agate pestle and mortar. Samples were then run on a Nicolet Thermofisher 6700 spectrometer; with a single bounce attenuated total reflection (ATR) smart orbit TM accessory using a diamond crystal. The absorption spectra were recorded with a resolution of 4cm⁻¹, using 64 scans accumulation in the spectral range

of 7000-400cm⁻¹. The approach chosen took advantage of the rapid sample and data acquisition made possible by using an ATR accessory. The wavenumber range chosen covers the mid-infra-red to part of the near infrared, allowing for interpretation and classification of the amber spectra.

Data analysis

For analysis, a classical multivariate approach was used, consisting of five steps:

a. Data pre-processing. The spectra were derived to obtain the second derivative and smoothed. To prevent any bias in the principal component analysis the processed data (102 samples) was mean centred.

b. Reduction of factor (wavenumber) dimensionality using a principal component analysis (PCA). The processed data is used and PCA transforms the original set of variables to a new set of uncorrelated variables (PCs). This is a data reduction process and the first few PCs will typically account for >95% variance. Here we used the correlation matrix and the NIPALS (non-linear iterative partial least squares) algorithm to implement the PCA. The first 9 PCA accounted for 98.93% of the variance. Henceforth we used 9 PCA.

c. Discriminant function analysis (DFA) on the independent PC basis. DFA has *a priori* information based on spectral replicates and uses this to minimise within group variance and maximise between group variance. The dataset was therefore randomly divided into three groups: training, validation and test groups. The DFA was calculated using a standard Euclidean distance criterion and cross-validation. The first 6 discriminant vectors accounted for 98.14% of the variances in the training and validation group. Henceforth we used 6 discriminant vectors.

d. For clarity, the discriminant clustered data were organised using a dendrogram (cluster observation), and a similarity matrix was constructed from the DFA space.

150 The dendrogram was drawn using complete linkage and Euclidian distance settings.
151 The similarity level decided for the group clustering. The dendrogram is analyzed as
152 follows: at each step, two clusters are joined. The corresponding similarity levels, the
153 identification number of the new cluster, the number of observations in the new
154 cluster, and the number of clusters were recorded in the amalgamation table. The
155 amalgamation continues until there is just one cluster. The amalgamation steps show
156 that the similarity level increases monotonously until it increases suddenly to a new
157 level and continues to monotonously increase. The sudden change in similarity
158 indicates the number of clusters reasonably sufficient for the final partition.

159 e. To determine the variable of importance underlying the classification, we used a
160 genetic algorithm discriminant function analysis. The processed spectra from step a)
161 were used directly. The genetic algorithm was setup as follows: 10 variables, 10
162 individuals per gene, 100 runs, cross-over 0.8, mutation rate 0.4, insertion rate 0.8,
163 maximum factor 8, and maximum generation 100. The spectral dataset was divided
164 randomly in three: training, validation and test groups. The genetic algorithm used the
165 eigenvalues computed by the DFA as an approximation for the F-ratio, with the GA
166 fitness function configured to minimize the inverse sum for the eigenvalues for the
167 training and validation dataset (Jarvis & Goodacre 2005).

168 For step a, b, c and e the analysis were run using Pychem (Jarvis et al. 2006). For step
169 d) the analysis was run using Minitab 17 (Minitab, Inc. www.minitab.com). For all
170 steps, the analyses were performed over the whole spectra, the mid infrared (500 –
171 4000 cm^{-1}), and the near infrared (4000 – 7000 cm^{-1}). Only the mid infrared results
172 are presented. The near infrared region (above 4000 cm^{-1}) had featureless spectra with
173 high similarities between samples (Figure S1 - supplementary materials). PCA

exploratory analysis in the near-infrared yielded no usable discrimination of the ambers.

Results

Locations and samples

Figure 1 shows the geographical location of the samples used in this study. Table 1 summarises the samples and their location of origin. Figure 2 shows the samples on the geological time scale.

Interpretation of FTIR-ATR spectra

Figure 3 illustrates a typical FTIR-ATR spectrum (BH-Bexhill Sussex, UK) with regions of interests. Not shown here is the region between 4000 and 7000 cm^{-1} . Figure 3A can be divided in to three regions of interest with figure 3B making the fourth region. Table 2 summarises the most commonly found peaks and their associated vibration and chemical groups. Although the chemical structure of ambers is very complex, all the FTIR spectra of ambers present similar features (Figure S1, supplementary materials). Consequently, even when ages and origins differ, ambers can be unambiguously differentiated by their FTIR spectra. Moreover, FTIR spectra of ambers show well-resolved absorption bands. The direct comparison of all the spectra is not trivial; we therefore use classification methods to cluster the ambers but also to determine which absorption bands are underlying the clustering of the ambers.

FTIR classification of amber by PCA-DFA

The resultant similarity dendrogram is shown in Figure 4 (see materials and methods for details). The dendrogram consists of two branches (red and blue), which split into four main clusters. The first branch (blue) splits into two main clusters, a larger one containing nine of the localities – all samples of Cretaceous Kuji amber, the Bexhill amber and the Isle of Wight amber, and a smaller cluster consisting only of the

Lebanese and Cooden Beach ambers. The second branch (red) similarly splits into a larger and smaller cluster. The larger cluster consists of three of the French amber localities, the Oligocene Kuji amber and the Baltic amber samples. The smaller cluster consists only of the Tonnay Charente and the Dominican ambers, which are of lowermost Upper Cretaceous and Oligo-Miocene age respectively.

A detailed analysis of the number of clusters using the amalgamation method suggested only two clusters for the PCA-DFA (figure 5), the red and blue branches in figure 4. The poor resolving power of the PCA-DFA could be attributed to the high variable collinearity (Jarvis & Goodacre 2005; Elliott et al. 2007). To help improve the DFA, we performed a genetic algorithm based DFA (GA-DFA).

FTIR classification of amber by GA-DFA- variable of importance analysis

The first step was to estimate the optimum number of variables needed for the GA-DFA to explain the variance in the dataset. Figure 6 shows a plot of the GA-DFA fitness function (see materials and methods) as a function of increasing number of variables. Two points are worth noting: (i) the fitness function showed a maximum value at 17 variables, our optimum; and, (ii) beyond 17 variables the fitness function decreased. This last point most likely explained the poor performance of the direct DFA when the number of variables was not controlled.

Using the same approach as the direct DFA, we can use the GA-DFA to produce a dendrogram (Figure 7). The amalgamation analysis (figure 5) suggested 10 clusters in this scenario.

Figure 8 shows a frequency plot for the variables of importance (see materials and methods for details). Briefly, the x-axis has the same units as the infrared spectra under investigation (i.e. wavenumber). Each wavenumber is considered as a variable (though highly correlated) the analysis reveals which of these wavenumbers

contribute to the explained variance in the PCA-DFA using the genetic algorithm. The y axis gives the occurrence (importance) of those wavenumbers. The higher the bar, the more important is the wavenumber in explaining the variance.

The main vibrations (highest frequency in figure 8) underlying the discrimination were 3048, 920, 1054, 2979, 2910, 662, 2787, 1170, 987, 1830 and 1430 cm^{-1} . Using table 2 known chemical group vibrations, we could identify that the variables of importance that explained the dendrogram were the amounts of exocyclic methylene, esters and alkyl groups.

The non-parametric Kruskal-Wallis statistical test was used to assess the differences in the median signals for the ambers. For each of the identified FTIR peaks of importance, at least one of the amber species had a median value that was statistically different from the rest ($p < 0.00$ for all plots in figures 9-11).

The signal intensities at selected wavenumber values are plotted in Figures 9, 10 and 11, as notched box-whisker plots, from which the isolate-level and species-level differences in the median signals' intensities at the representative wavenumber can be discerned. The lower and upper lines of the 'box' are the 25th and the 75th percentiles of the sample, the box limits indicating the interquartile range for each setting, with the horizontal bar representing the median for the isolate/species. The 'whiskers' (lines extending above and below the box) show the range (the maximum and minimum values), excluding outliers (values of >1.5 times the interquartile range). A plus sign outside the whiskers indicates an outlier in the data. The notches in the box are a graphic confidence interval about the median. A side-by-side comparison of two notched plots can be considered as a graphical equivalent of a t-test. Note that the comparison is made between notches and not the overlap between whiskers. Clear

differences in the signal intensities for the different isolates and species can be noticed at the plotted wavenumber value.

For example, in figure 10 we could confirm, as previously reported (Guiliano et al. 2007; Teodor et al. 2009), that the band at 1170cm^{-1} (the “Baltic shoulder”) significantly differentiates the Paleogene Baltic amber (BAL) from all the other ambers.

Discussion

The final dendrogram based on the FTIR spectra of the ambers (Figure 7) shows an initial split into two branches, with one branch containing the majority of the ambers. However, there appears to be little clear age or locality based trend to the clustering. Within the larger branch is a larger (blue) cluster and several smaller clusters (green, purple, black, red). The larger cluster contains a mix of French, British, Japanese, Baltic and American ambers, with ambers from the same region not branching next to each other. Likewise, the blue cluster contains both some of the earliest Cretaceous amber (BH) and some of the youngest (Bal). The smaller clusters showing Japanese Kuji amber from the Late Cretaceous branching next to, but not with, late Cretaceous French ambers. Kuji and Dominican amber clustering together despite a large age gap. The smaller branch contains only four ambers: the Oligocene Kuji amber, Late Cretaceous Kuji amber (KujiHO) Cooden Beach amber (CB) and Isle of Wight amber (IoWHA), with the Oligocene Kuji samples clustering most closely with the Early Cretaceous Cooden Beach material. The dendrogram combines all data for comparison, though the complexity of amber chemistry may also act to cloud differences between particular wavenumbers of importance. Therefore, more detailed analysis of the wavenumber of importance, such as the Kruskal-Wallis test shown in

figures 9, 10 and 11, may better help to determine specific links between chemical and geological processes.

The multivariate analysis confirmed that the exocyclic methylene (3048 cm^{-1}) signal was the dominant factor explaining the variance in the DFA, simply because all the ambers have the exocyclic methylene. This was not sufficient to differentiate the ambers. Figure 9 shows for 3048 cm^{-1} that most of the ambers are not significantly different from one another except for BH.

Exocyclic methylene content has been directly linked to age of ambers (Lambert et al. 1985; Grimaldi 1995; Kimuara et al. 2006), but factors other than age (such as salinity, temperature and pressure) may explain the variation in the exocyclic methylene content. Brody et al. (Brody et al. 2001) used Fourier-transform Raman spectroscopy in an attempt to link chemical features to geological variations. They found that the direct analysis of the vibrational spectra was not sufficient to draw any conclusions. Our approach, using the variable of importance, allows for a more intuitive analysis of the factors underlying amber differences (figures 9-11).

Concomitant to the 3048 cm^{-1} band, signals at 920 and 1054 cm^{-1} did contribute to the dendrogram classification. In figure 9 at 920 cm^{-1} , BAL and DOM significantly differ from the other ambers. One could speculate that the 920 contribution is partly linked to amber age. In figure 9 at 1054 cm^{-1} , the differences were not directly linked to any amber formation factors.

Careful analysis of the exocyclic methylene group suggested that single bonds between carbon and hydrogen in the exocyclic group correlate to absorptions at ~ 2927 , 2853 , 1470 , and 1380 cm^{-1} , and a band from ~ 1050 – 950 cm^{-1} (Villanueva-Garcia et al. 2005; Abduriyim et al. 2009). Double bonds between carbons correlate with absorptions at ~ 3070 , 1640 , and 887 cm^{-1} . One could speculate that the former

(single bond) is sensitive to polymer orientation and the latter (double bond) is sensitive to the degree of polymerization.

Finally, the ester group contribution from the Baltic amber appeared in the variable of importance at 1170 cm^{-1} (figure 10). Interestingly KujiO had a similar contribution of the ester group.

Taken together, the frequencies of the variables of importance, their associated chemical groups and the Kruskal-Wallis test, suggested that the FTIR-ATR classification was predominantly sensitive to the age of the ambers and to some extent to location (ester group) and contaminations (eg. 660 cm^{-1} band, commonly associated with aromatic substitutions) or oxidative processes.

Conclusion

Cluster analysis of FTIR spectra of amber from 17, globally distributed localities spanning the Lower Cretaceous to Oligo-Miocene was carried out as a novel way to investigate amber chemistry and relationships. When all information is included in the analysis the resulting dendrogram is complex, and lacks clear patterns in clustering. However comparisons can be made between ambers when examining specific wavenumbers of importance using the Kruskal-Wallis statistical test. Multivariate analysis indicated the exocyclic methylene peak was the dominant factor in the clustering of the samples; this molecule has been previously linked to the age of ambers. Modelling of the variable of importance, helped resolve the contribution of the exocyclic methylene in total content and conformation; as well as confirming the role of ester and alkyl contents.

Further studies, particularly looking at sites where amber of multiple ages are available, may be useful in further distinguishing the changes taking place in amber chemistry through the fossilisation processes.

Acknowledgements

Christopher Nicholas, Didier Nerandean are thanked for providing specimens of Kuji and Isle of Wight (CN) and French (DN) ambers.

Supplementary information

Supplementary Table 1. Detailed sample list

Supplementary Figure S1. Combined spectra of all samples

References

- Abduriyim, A., Kimura, H., Yokohama, Y., Nakazono, H., Wakatsuki, M., Shimizu, T., Tansho, M. & Ohki, S. 2009. Characterization of "green amber" with infrared and nuclear magnetic resonance spectroscopy. *Gems and Gemology*, **45**, 158-177.
- Alonso, J., Arillo, A., Barron, E., Carmelo Corral, J., Grimalt, J., Lopez, J.F., Lopez, R., Martinez-Delclos, X., *et al.* 2000. A new fossil resin with biological inclusions in Lower Cretaceous deposits from Alva (Northern Spain, Basque-Canabrian Basin). *Journal of Paleontology*, **74**, 158-178.
- Anderson, K.B. 2006. The nature and fate of natural resins in the geosphere. XII. Investigation of C-ring aromatic diterpenoids in Raritan amber by pyrolysis-GC-matrix isolation FTIR-MS. *Geochemical Transactions*, **7:2**, <http://doi.org/doi:10.1186/1467-4866-7-2>.
- Angelini, I. & Bellintani, P. 2005. Archaeological ambers from northern Italy: an FTIR-DRIFT study of provenance by comparison with geological amber database. *Archaeometry*, **47**, 441-454.
- Bandel, K. & Vavra, N. 1981. Ein fossils Harz aus der Unterkeide Jordaniens. *Neues Jarbuch fur Geologie und Palaeontologie Monatscheft*, **1**, 19-33.
- Beck, C.W. 1986. Spectroscopic investigations of Amber. *Applied Spectroscopy Reviews*, **22**, 57-110.
- Breton, G. 2007. La bioaccumulation de microorganismes dans l'ambre: analyse comparée d'un ambre sparnacien et de leurs tapis algaires et bacteriens. *Comptes Rendus Palevol*, **6**, 125-133.
- Brody, R.H., Edwards, H.G.M. & Pollard, A.M. 2001. A study of amber and copal samples using FT-Raman spectroscopy. *Spectrochimica Acta Part A: Molecular and biomolecular spectroscopy*, **57**, 1325-1338.
- Cano, R., Poinar, H., Pieniazek, N., Acra, A. & Poinar, G. 1993. Amplification and sequencing of DNA from a 120-135 million year old weevil. *Nature*, **363**, 536-538.

- Delclos, X., Briggs, D.E.G. & Penalver, E. 2004. Taphonomy of insects in carbonates and amber. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **203**, 19-64.
- Delclos, X., Arillo, A., Penalver, E., Barron, E., Soriano, C., Lopez Del Valle, R., Bernardez, E., Corral, C., *et al.* 2007. Fossiliferous amber deposits from the Cretaceous (Albian) of Spain. *Comptes Rendus Palevol*, **6**, 135-149.
- Desalle, R., Gatesy, J., Wheeler, W. & Grimaldi, D. 1993. DNA sequenced from a fossil termite in oligo-Miocene amber and phylogenetic implications. *Science*, **257**, 1933-1936.
- Elliott, G.N., Worgan, H., Broadhurst, D., Draper, J. & Scullion, J. 2007. Soil differentiation using fingerprint Fourier transform infrared spectroscopy, chemometrics and genetic algorithm-based feature selection. *Soil Biology and Biochemistry*, **39**, 2888-2896.
- Grimaldi, D.A. 1995. The age of Dominican amber. *In*: Anderson, K.B. & Crelling, J.C. (eds) *Amber, Resinite, and Fossils Resins*. American Chemical Society.
- Grimaldi, D.A. 1996. *Amber: window into the past*. Harry N. Abrams Inc., New York.
- Grimaldi, D.A., Engel, M.S. & Nascimbene, P.C. 2002. Fossiliferous Cretaceous amber from Myanmar (Burma): its rediscovery, biotic diversity, and paleontological significance. *American Museum Novitates*, **March**, 1-71.
- Guiliano, M., Mille, G., Onoratini, G. & Simon, P. 2006. Presence of amber in the Upper Cretaceous (Santonian) of La 'Mède' (Martigues, southeastern France). IRTF characterization. *Comptes Rendus Palevol*, **5**, 851-858.
- Guiliano, M., Asia, L., Onoratini, G. & Mille, G. 2007. Applications of diamond ATR FTIR spectroscopy to the characterization of ambers. *Spectrochimica Acta Part A: Molecular and biomolecular spectroscopy*, **67**, 1407-1411.
- Jarvis, R.M. & Goodacre, R. 2005. Genetic algorithm optimization for pre-processing and variable selection of spectroscopic data. *Bioinformatics*, **21**, 860-868.
- Jarvis, R.M., Broadhurst, D., Johnson, H.E., O'Boyle, N. & Goodacre, R. 2006. PyChem - a multivariate analysis package for python. *Bioinformatics*, **22**, 2565-2566.
- Jarzembowski, E.A. 1999. British amber: a little-known resource. *Estudios del museo de ciencias naturales de alva Numero especial*, **2**, 133-140.
- Katagiri, T., Mukai, M. & Yamaguchi, T. 2013. A new fossil moss *Muscites kujiensis* (Bryopsida) preserved in the late Cretaceous amber from Japan. *The Bryologist*, **116**, 296-301.
- Kimura, H., Tsukada, Y., Mita, H., Yamamoto, Y., Chujo, R. & Yukawa, T. 2006. A spectroscopic index for estimating the age of amber. *Bull. Chem. Soc. Jpn.*, **79**, 451-453.

- Lambert, J.B., Frye, J.S. & Poinar, G.O. 1985. Amber from the dominican republic: analysis by nuclear magnetic resonance spectroscopy. *Archaeometry*, **27**, 43-51.
- Lambert, J.B., Santiago-Blay, J.A. & Anderson, K.B. 2008. Chemical signatures for fossilized resins and recent plant exudates. *Angewandte Chemie International Edition*, **47**, 9608-9616.
- Litescu, S.-C., Teodor, E.D., Truica, G.-I., Tache, A. & Radu, G.-L. 2012. Fourier transform infrared spectroscopy - Useful analytical tool for non-destructive analysis. In: Theophanides, T. (ed) *Infrared Spectroscopy - Materials Science, Engineering and Technology*. InTech.
- Martill, D.M., Loveridge, R.F., Gomes de Andrade, J.A.F. & Cardoso, A.H. 2005. An unusual occurrence of amber in laminated limetstones: the Crato formation lagerstatte (early Cretaceous) of Brazil. *Paleontology*, **48**, 1399-1408.
- Neraudeau, D., Perrichot, V., Dejax, J., Masure, E., Nel, A., Philippe, M., Moreau, P., Guillocheau, F., *et al.* 2002. A new fossil locality with insects in amber and plants (likely Uppermost Albian): Archingeay (Charente-Maritime, France). *Geobios*, **35**, 233-240.
- Nicholas, C.J., Henwood, A.A. & Simpson, M. 1993. A new discovery of early cretaceous (Wealden) amber from the Isle of Wight. *Geology magazine*, **130**, 847-850.
- Nissenbaum, A. 1975. Lower Cretaceous amber from Israel. *Naturewissenschaften*, **62**, 341-342.
- Penalver, E., Delclos, X. & Soriano, C. 2007. A new rich amber outcrop with palaeobiological inclusions in the lower cretaceous of Spain. *Cretaceous Research*, **28**, 791-802.
- Penney, D. 2002. Spiders in upper cretaceous amber from New Jersey (Arthropoda: Araneae). *Palaeontology*, **45**, 709-724.
- Perrichot, V. 2004. Early Cretaceous amber from South Western France: insight into the Mesozoic litter fauna. *Geologica acta*, **2**, 9-22.
- Poinar, G. & Poinar, R. 2008. *What bugged the dinosaurs?* princeton University Press, Princeton/Oxford.
- Poinar, H., Poinar, G. & Cano, R. 1993. Oldest DNA from plants. *Nature*, **363**, 677.
- Russell, J.D. & Fraser, A.R. 1994. Infrared methods. In: Wilson, M.J. (ed) *Clay mineralogy: spectroscopy and chemical determinative methods*. Chapman and Hall, London, 16-67.
- Schlee, D. & Dietrich, H.G. 1970. Insektenfurender Bernstein aus der Uterkreide des Lebanon. *Neus Jarbuch fur Geologie und Palaeontologie Monatscheft*, **1**, 40-50.

- Schmidt, A.R., Ragazzi, E., Coppellotti, O. & Roghi, G. 2006. A microworld in Triassic amber. *Nature*, **444**, 835.
- Teodor, E., Litescu, S.-C., Neacsu, A., Truica, G. & Albu, C. 2009. Analytical methods to differentiate Romanian amber and Baltic amber for archaeological applications. *Open Chemistry*, **7**, 560-568.
- Trevisan, J., Angelov, P.P., Carmichael, P.L., Scott, A.D. & Martin, F.L. 2012. Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst*, **137**, 3202-3215.
- Van Bergen, P.F., Collinson, M.E., Scott, A.C. & de Leeuw, J.W. 1995. Unusual resin chemistry from the Upper Carboniferous pteridosperm resin rootlets. *In*: Anderson, K.G. & Crelling, J.C. (eds) *Amber, resinite and fossil resins*. American Chemical Society Symposium Series 617, Washington, 149-169.
- Vavra, N. 2009. The chemistry of Amber - Facts, findings and opinions. *Ann. Naturhist. Mus. Wien*, **111 A**, 445-474.
- Villanueva-Garcia, M., Martinez-Richa, A. & Robles, J. 2005. Assignment of vibrational spectra of labdatiene derivatives and ambers: a combined experimental and density functional theoretical study. *Arkivoc*, **4**, 449-458.
- Waggoner, B.M. 1994. Fossil microorganisms from Upper Cretaceous amber of Mississippi. *Review of Palaeobotany and Palynology*, **80**, 75-84.

494 **Figure captions**

495 Table 1. Sample list

496 Table 2. Vibrational frequencies, modes and chemical groups often found in amber

497 Figure 1. Map showing global distribution of samples used: 1 – New Jersey; 2 –
498 Dominican Republic; 3, 4, 5 – Bexhill and Isle of Wight, UK; 6, 7, 8, 9 - Tonnay-
499 Charente, Archingeay, Fouras, La Buzinie, South-West France; 10 – Lebanon; 11 –
500 Baltic region; 12 – Kuji, Japan.

501 Figure 2. Stratigraphic column showing the age distribution of amber samples.

502 Figure 3. Example of a typical amber FTIR-ATR spectrum showing regions of
503 interest (see text for details).

504 Figure 4. Dendrogram based on FTIR-ATR spectra PCA-DFA analysis.

505 Figure 5. Amalgamation plot comparison between PCA-DFA and GA-DFA. A steep
506 increase in the plot is indicative of the number of cluster. The PCA-DFA shows 2
507 clusters, the GA-DFA indicates 10 clusters.

508 Figure 6. Model error for the genetic algorithm discriminant function analysis (GA-
509 DFA). The fitness function maximizes the between/within group variance using the
510 sum of the eigenvalues (training and validation dataset) as a function of number of
511 variables. The best model is chosen to be the one with maximum fitness.

512 Figure 7. Dendrogram based on GA-DFA analysis

513 Figure 8. Frequency distribution of the variables of importance underlying the FTIR
514 based dendrogram

515 Figure 9. Median FTIR second derivative signal intensities for the different ambers at
516 3048, 920, 1054 and 2979 cm^{-1} . The FTIR intensities were multiplied by 10^5 for
517 clarity. A side-by-side comparison of two notched plots can be considered as a

graphical equivalent of a t-test. The red plus indicates outliers. See text for further interpretation.

Figure 10. Median FTIR second derivative signal intensities for the different ambers at 2910, 662, 2787 and 1170 cm^{-1} . The FTIR intensities were multiplied by 10^5 for clarity. A side-by-side comparison of two notched plots can be considered as a graphical equivalent of a t-test. The red plus indicates outliers. See text for further interpretation.

Figure 11. Median FTIR second derivative signal intensities for the different ambers at 987, 1830 and 1430 cm^{-1} . The FTIR intensities were multiplied by 10^5 for clarity. A side-by-side comparison of two notched plots can be considered as a graphical equivalent of a t-test. The red plus indicates outliers. See text for further interpretation.

Table 1. Sample list

Amber locality	Identifier	No. of specimens	No. of samples	Age (ref)
Dominican Republic	DOM	6	6	Late Oligocene-Miocene (Poinar, 1995)
Baltic	BAL	11	11	Late Oligocene-Early Eocene (Poinar, 1995)
Kawanuki Quarry, Kuji Japan.	KA-3	1	2	Early Oligocene (C. Nicholas, per.comm.)
Maita Beach, Kuji, Japan.	OK-2	1	4	Santonian (C. Nicholas, per.comm.)
Uge Beach, Kuji Japan.	UG-4, UG-1	2 (different beds)	6	Santonian (C. Nicholas, per.comm.)
Horinai Quarry, Kuji, Japan.	HO-5	1	3	Coniacian (C. Nicholas, per.comm.)
Tamagawa Beach (North), Kuji, Japan.	TAMI-1	1	3	Coniacian (C. Nicholas, per.comm.)
Tamagawa Beach (South), Kuji, Japan.	TAMIII-9	1	3	Coniacian (C. Nicholas, per.comm.)
New Jersey	NJ	16	16	Grimaldi (1996)
La Buzinie, France	BUZ_FR3	1	3	Cenomanian (Perrichot <i>et al.</i> , 2007)
Fouras, France	Fouras_FR1	1	3	Cenomanian (Neraudeau <i>et al.</i> , 2003)
Archingeay, France	ARCH_FR4	1	3	Late Albian (Neraudeau <i>et al.</i> , 2002)
Tonnay-Charente, France	T-C_FR2	2	7	Late Albian (Neraudeau <i>et al.</i> , 2005)
Chiltern Chine, Isle of Wight, UK	IoW	4 (2 different beds)	17	Barremian (Nicholas <i>et al.</i> , 1993)
Bexhill, Cooden Bay	CB	1	3	Hautervarian (unpublished data)
Lebanon	LEB	4	4	Valanginian
Bexhill, Sussex, UK	BH	14	14	Berriasian (Brasier <i>et al.</i> , 2009)

533

534

Table 2. Vibrational frequencies, modes and chemical groups often found in amber

Regions	Wavenumber (cm ⁻¹)	Vibration	Chemical group
1	887	γ_{C-H}	Exocyclic methylene (C=CH ₂)
	1015-1021	ν_{C-O}	
	1157	ν_{C-O}	Ester
	1204-1245	ν_{C-O}	Carboxylic acid
2	1268	ν_{C-O}	Carboxylic acid
	1375-1384	δ_{CH_3}	
	1443-1450	δ_{CH_2, CH_3}	
	1642	$\nu_{C=C}$	Exocyclic double bond
3	1695	$\nu_{C=O}$	Carboxylic acid
	1700-1705	$\nu_{C=O}$	Carboxylic acid
	1710-1715	$\nu_{C=O}$	Ketone
	1735	$\nu_{C=O}$	Ester

	2800-3000	$\nu_{\text{CH,CH}_2,\text{CH}_3}$	Alkyl stretch (methyl / methylene)
4	3048	$\nu_{\text{C-H}}$	Exocyclic methylene
	3078	$\nu_{\text{H-C=}}$	
	3450	ν_{OH}	

a (Beck 1986; Angelini & Bellintani 2005; Villanueva-Garcia et al. 2005; Guiliano et al. 2006; Guiliano et al. 2007; Teodor et al. 2009; Litescu et al. 2012)

535

536 Supplementary information

Sample label	Group label
BH1	BH
BH2	BH
BH3	BH
BH4	BH
BH5	BH
BH6	BH
BH7	BH
BH8	BH
BH9	BH
BH10	BH
BH11	BH
BH12	BH
BH13	BH
BH14	BH
NJ1	NJ
NJ2	NJ
NJ3	NJ
NJ4	NJ
NJ5	NJ
NJ6	NJ
NJ7	NJ
NJ8	NJ
NJ9	NJ
NJ10	NJ
NJ11	NJ
NJ12	NJ
NJ13	NJ
NJ14	NJ
NJ15	NJ
NJ16	NJ
TCFR2c	TCFR
TCFR2b	TCFR
TCFR2a	TCFR
KujiUG4c	KujiU
KujiUG4b	KujiU
KujiUG4a	KujiU
KujiUG1c	KujiU
KujiUG1b	KujiU
KujiUG1a	KujiU

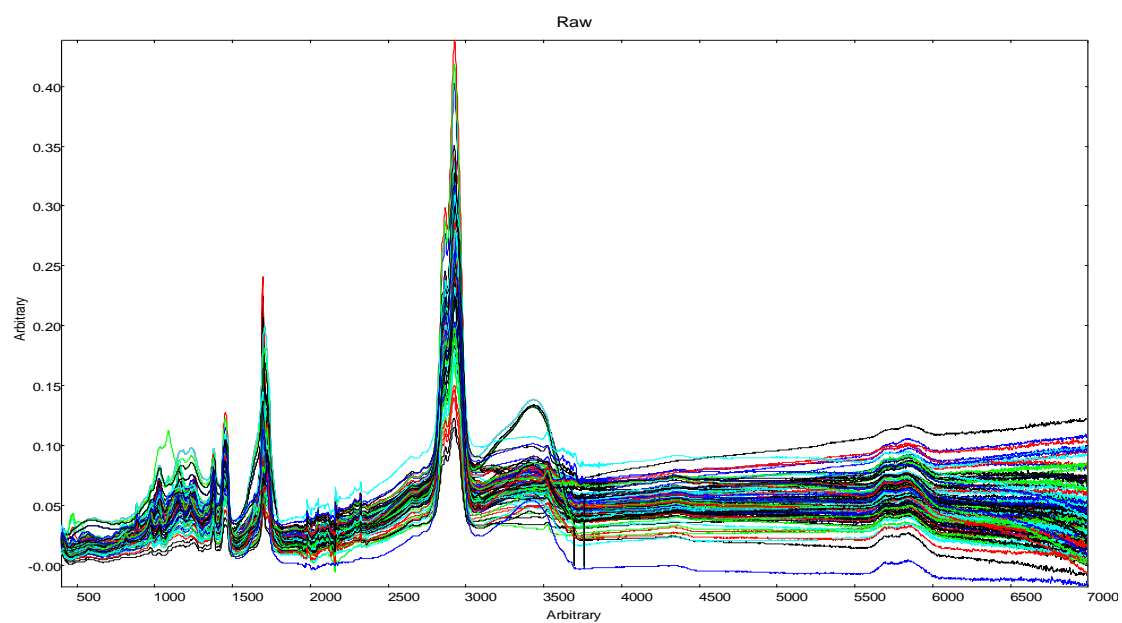
KujiTAMIII9c	KujiT
KujiTAMIII9b	KujiT
KujiTAMIII9a	KujiT
KujiTAMI1c	KujiT
KujiTAMI1b	KujiT
KujiTAMI1a	KujiT
KujiOK2d	KujiO
KujiOK2c	KujiO
KujiOK2b	KujiO
KujiOK2a	KujiO
KujiKA3b	KujiO
KujiKA3a	KujiO
KujiHO5c	KujiO
KujiHO5b	KujiO
KujiHO5a	KujiO
IoWMS3d	IoWMS
IoWMS3c	IoWMS
IoWMS3b	IoWMS
IoWMS3a	IoWMS
IoWMS2d	IoWMS
IoWMS2c	IoWMS
IoWMS2b	IoWMS
IoWMS2a	IoWMS
IoWMS1e	IoWMS
IoWMS1d	IoWMS
IoWMS1c	IoWMS
IoWMS1b	IoWMS
IoWMS1a	IoWMS
IoWHAd	IoWHA
IoWHAc	IoWHA
IoWHAb	IoWHA
IoWHAa	IoWHA
FourasFR1c	FR
FourasFR1b	FR
FourasFR1a	FR
Dom6	Dom
Dom5	Dom
Dom4b	Dom
Dom4a	Dom
Dom3	Dom
Dom2	Dom
Dom1	Dom
BuzFR3c	Buz
BuzFR3b	Buz
BuzFR3a	Buz
Bal11	Bal
Bal10	Bal
Bal9	Bal
Bal8	Bal
Bal7	Bal
Bal6	Bal
Bal5	Bal

Bal4	Bal
Bal3	Bal
Bal2	Bal
Bal1	Bal
ArchFR4c	Arch
ArchFR4b	Arch
ArchFR4a	Arch
TCFR2d	TCFR
TCFR2c	TCFR
TCFR2b	TCFR
TCFR2a	TCFR
Leb1	Leb
Leb2	Leb
Leb3	Leb
Leb4	Leb
CB1a	CB
CB1b	CB
CB1c	CB

537

538

539 Figure S1. Combined FTIR-ATR raw spectra (110 spectra) in the 400 – 7000 cm^{-1}
540 range.



541

542

543

544

545 Figure 1.



546

547

548

549

550

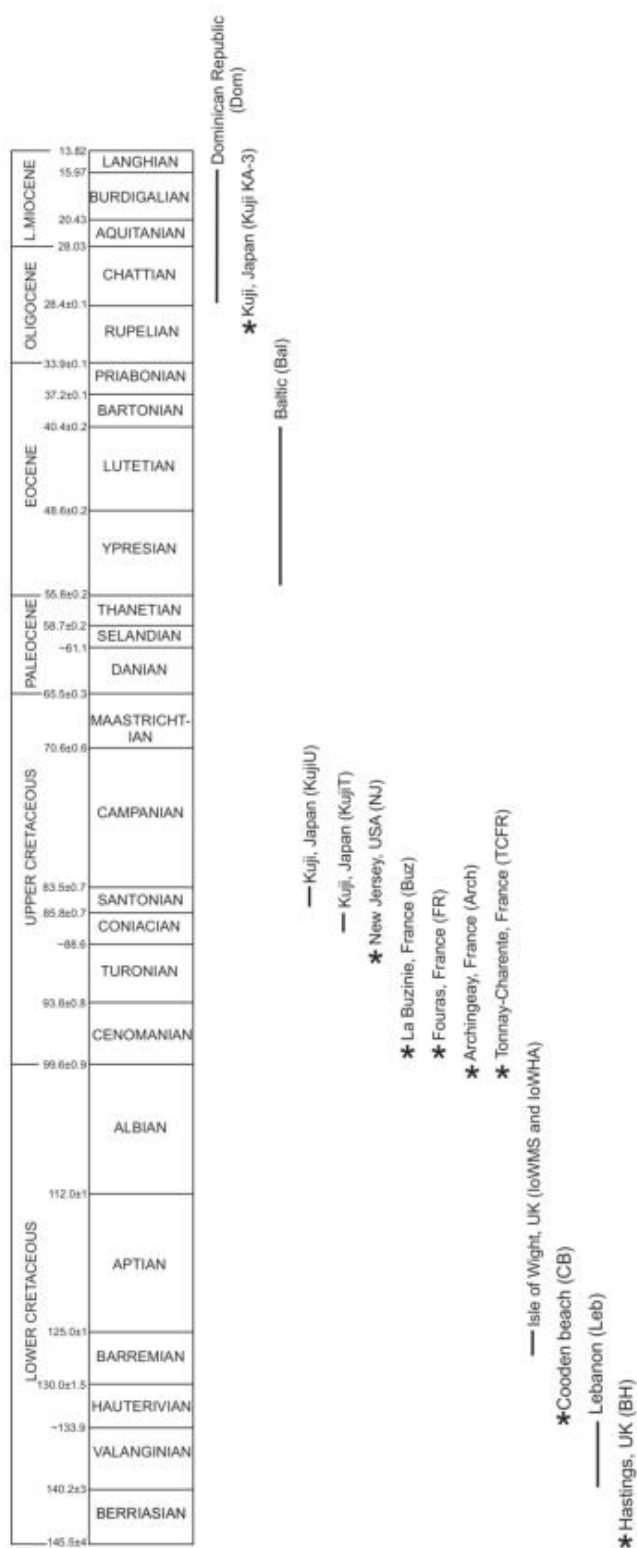


Figure 3

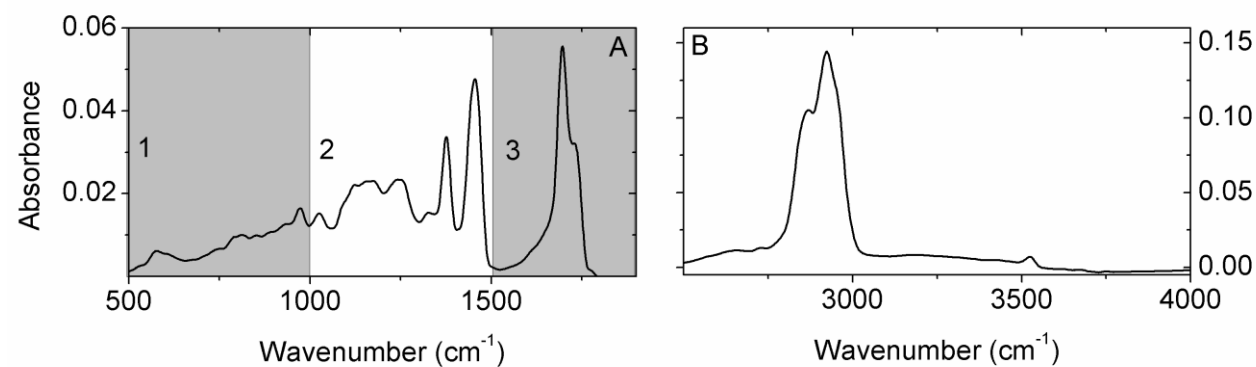
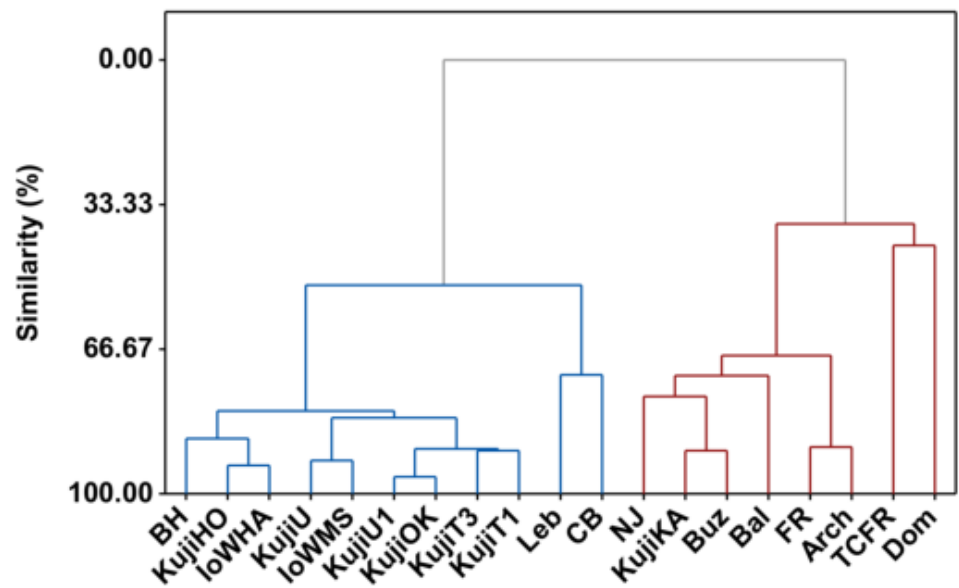
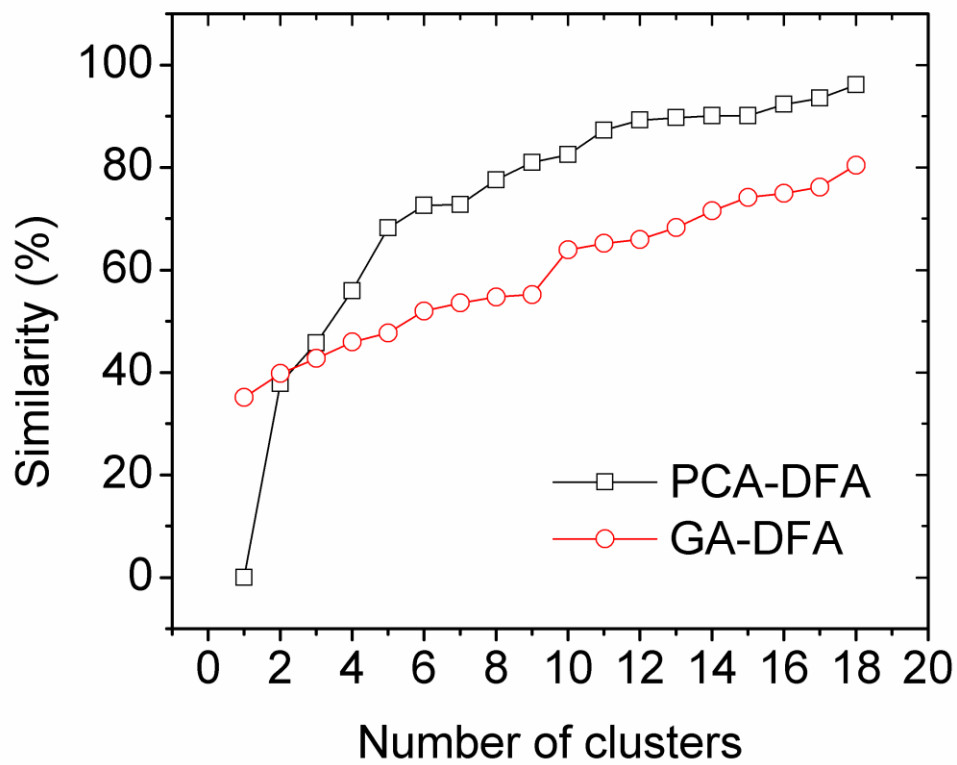


Figure 4.



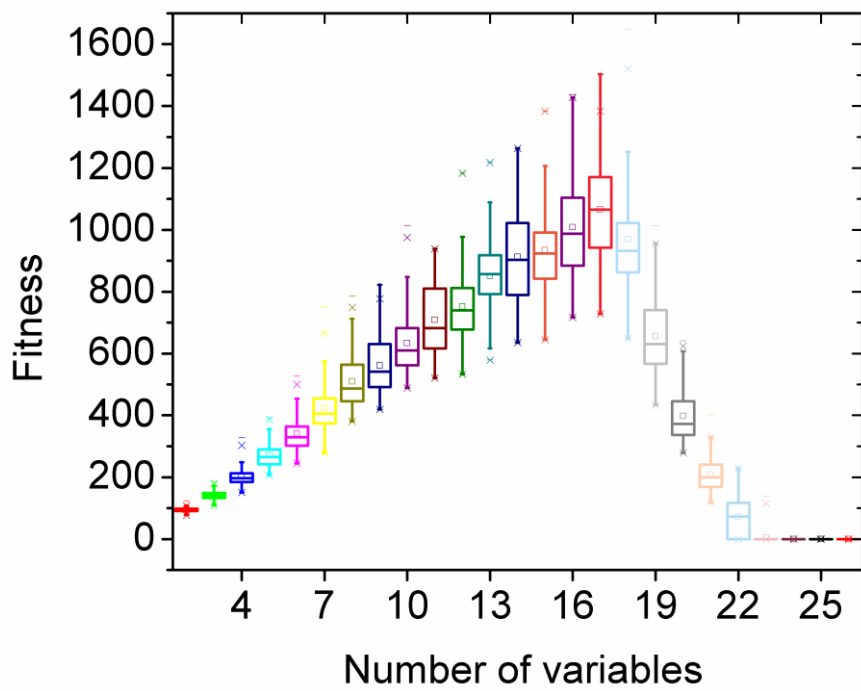
562 Figure 5.



563

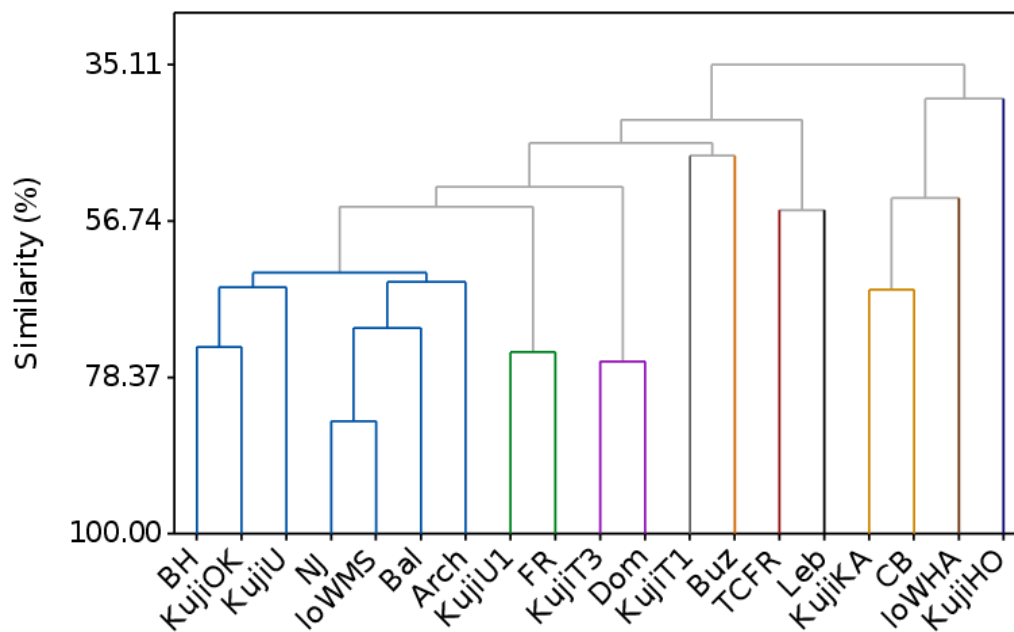
564

565 Figure 6.



566

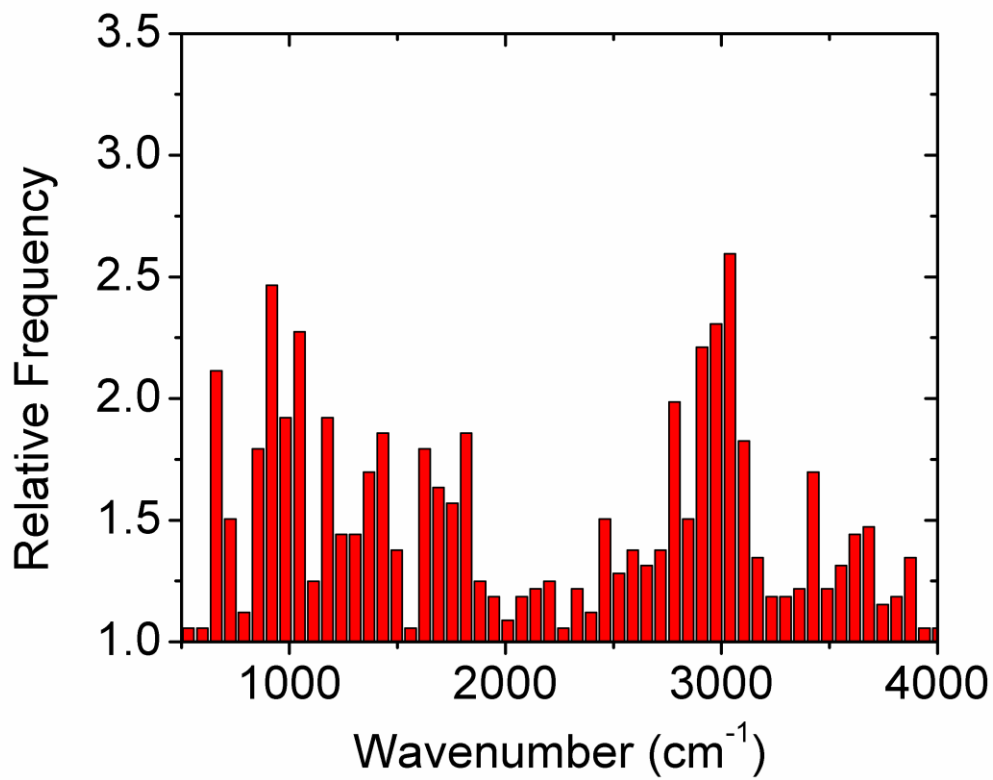
567 Figure 7.



568

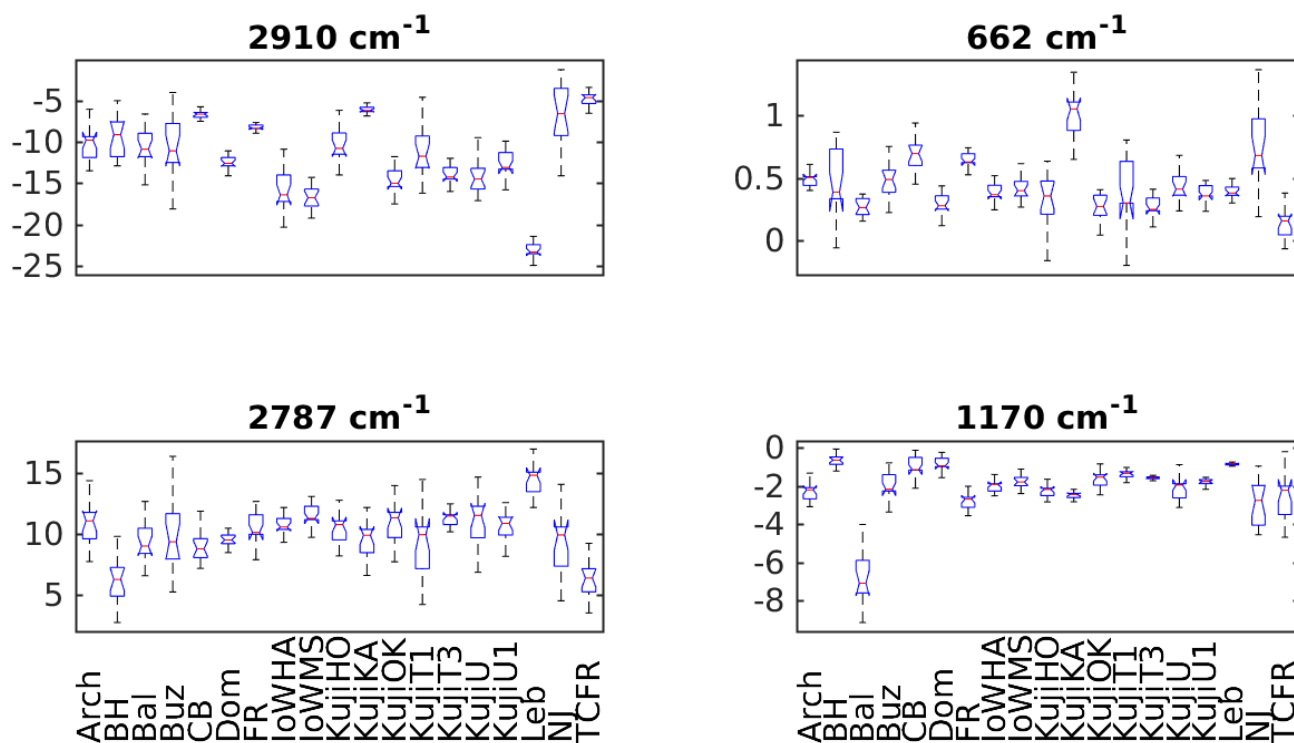
569

570 Figure 8.



571

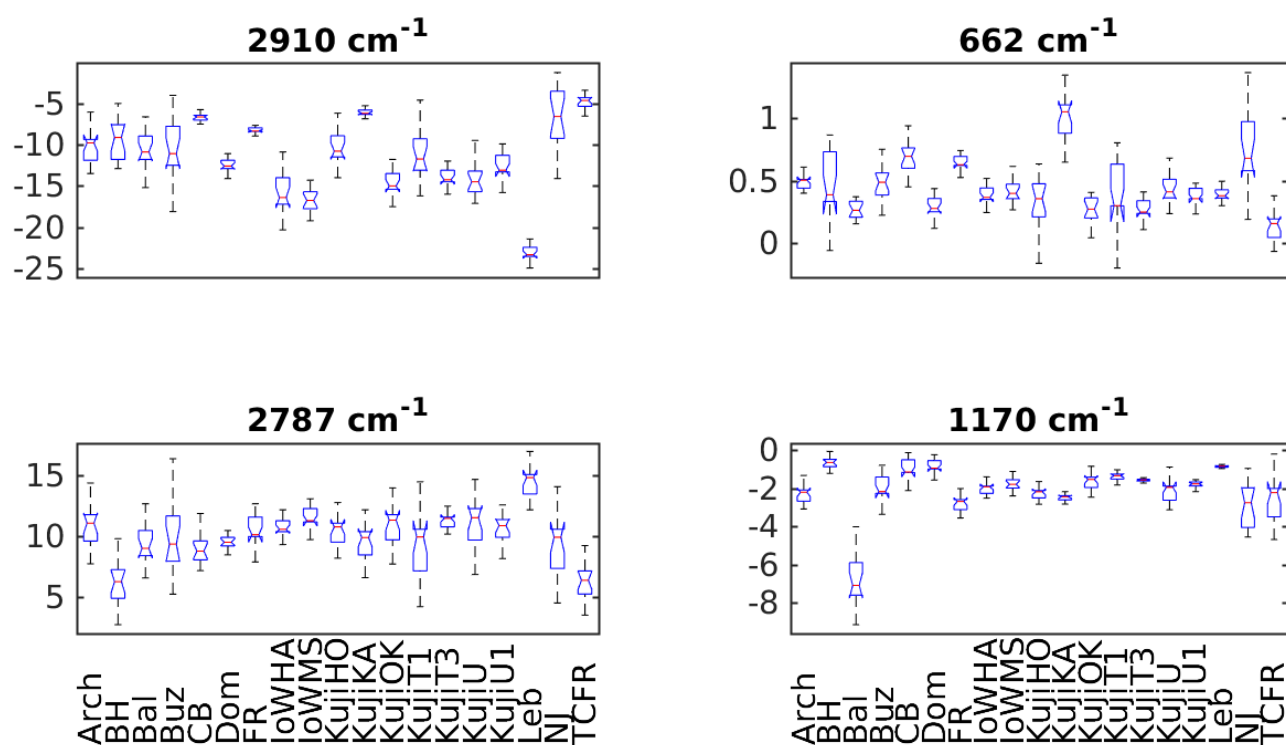
572 Figure 9.



573

574

575 Figure 10.



576

