Wood from Angiosperms and Gymnosperms Shows Distinctive Lignin Peaks in Direct Analysis in Real Time (DART) Mass Spectra

Robert B. Cody*1, Edgard O. Espinoza2, Erin R. Price3, and Pamela J. McClure2

1 JEOL USA, Inc., 11 Dearborn Rd., Peabody MA 01960 USA
2 US Fish and Wildlife Service, Office of Law Enforcement, National Fish & Wildlife Forensic Laboratory, 1490 East Main Street, Ashland, OR 97520, USA
3 US Forest Service International Programs Wood Identification and Screening Center, 1490 East Main Street, Ashland, OR 97520, USA

*Corresponding author. E-mail: cody@jeol.com

Abstract
A data set was constructed consisting of 3021 mass spectra randomly selected from the ForeST© (Forensic Spectra of Trees) database of mass spectra for wood analyzed by Direct Analysis in Real Time ionization coupled with time-of-flight mass spectrometry (DART-TOFMS). Clear and reproducible differences were observed between the lignin peaks for angiosperms and gymnosperms, with DART-TOFMS spectra of angiosperms showing significantly higher abundances for peaks associated with syringyl subunits. These differences can be used to provide support for enforcing trade laws by accurately identifying the source of finished wood products when anatomical differences are not present.

Introduction
With an estimated 73,300 species of trees worldwide [1] taxonomic identification of timber is decidedly challenging. Logging practices in developing tropical countries often involves destructive clear-cutting of trees from the angiosperm taxa. When these logs are processed it is often done in ways that destroy the traditional anatomical features used for morphological identification. These wood-based products can then be imported into consumer countries with the claim that the wood originated from gymnosperm trees. The advantage to the importer is the low taxation rate afforded conifers, giving them an additional financial incentive.

Direct Analysis in Real Time ionization coupled with time-of-flight mass spectrometry (DART-TOFMS) has become an important tool in combatting the illegal trade of timber protected by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES)[2-16]. A key development in this application has been the creation of the ForeST© (Forensic Spectra of Trees) database of DART mass spectra of samples from commercial sources and curated xylaria[15]. The database contains more than 12,000 spectra comprised from 1267 species encompassing 241 genera and 59 families.
Lignin is a ubiquitous component of wood; therefore the lignin peaks in DART mass spectra are not useful for chemotaxonomic classification at the species level. In fact, the presence of lignin peaks in DART mass spectra can lead to erroneous identifications in ForeST© database searches[15]. It can be helpful to omit the lignin peaks for that reason. However, there are useful structural differences between the lignin composition that can differentiate at the clade level between angiosperms (hardwood) and gymnosperms (softwood). Lignin in gymnosperms contains guaiacyl (G) and p-hydroxyphenyl (H) subunits. Lignin in angiosperms also contains syringyl (S) subunits[17].

Mass spectrometry is a key analytical method for the characterization of lignin. A review of mass spectrometric methods for the analysis of lignin was recently published by Letourneau and Volmer[18]. The detection of lignin peaks by DART-TOFMS was first reported by Jones et al. in an article about the pyrolysis of polar biomass[19]. Adams proposed structures for the principal peaks arising from the DART-TOFMS analysis of lignin in an article about the analysis of printing and writing papers[20]. Guo et al. detected lignin peaks by DART mass spectrometry in an evaluation of the deterioration state of archaeological wooden artifacts[21]. Crawford et al. compared DART and spray ionization methods for lignin analysis mass spectrometry and concluded that spray methods produce more high-mass species than DART[22].

As referenced from Section IX of the “Harmonized Tariff Schedule of the United States Revision 11”[23]), identifying whether timber originated from an angiosperm or gymnosperm is a key requirement for enforcing United States trade laws which largely focus on conifers (a group of gymnosperms that includes Pinaeaceae, Podocarpaceae, Cupressaceae, Araucariaceae, Cephalotaxaceae, Phyllocladaceae, Sciadopityaceae, and Taxaceae)[24] and non-conifers. By claiming wood products originated from gymnosperms rather than angiosperms, or vice versa, importers can easily avoid paying large tariffs on their shipments.

The traditional method for differentiating between angiosperms and gymnosperms is by anatomical analysis [25, 26]. As described by Wheeler [27], 90% of gymnosperm wood is composed of longitudinal tracheid cells with the remaining 10% consisting of rays, while angiosperms are comprised of vessels and fibers with between 10-30% of the wood consisting of rays. Due to the necessity of identifying finished products, wood anatomy is not always possible especially in cases of wood pellets, paper, cardboard, compressed wood, plywood, and veneers where the anatomical features are absent making alternative methods necessary. Because of the growing use of DART-TOFMS to identify wood, we investigated whether the method can be used to distinguish between wood from angiosperms and gymnosperms.

Methods

The ForeST© database contains an extensive collection of positive-ion DART-TOF mass spectra measured using an AccuTOF™-DART® mass spectrometer (JEOL USA, Inc., Peabody MA USA) by holding slivers of wood in the helium DART gas stream with the gas heater set to 350°C. Mass spectra containing lignin peaks were extracted for analysis from the raw text-file format data from which the ForeST© database was created. The methods used to acquire these mass spectra and develop the database are described in previous publications[2, 7]. Relative peak abundances in the ForeST© database mass spectra are normalized to the most abundant (base) peak as 100%.

A test set of data was constructed with mass spectra of wood samples from a variety of species of wood from both gymnosperm and angiosperm clades. Approximately 20 of the initially selected mass spectra
with low signal-to-noise ratios that did not exhibit lignin peaks were omitted. The final test set consisted of 1313 randomly selected mass spectra of angiosperms and 1708 mass spectra of gymnosperms. Masses and tentative assignments for lignin peaks in the DART mass spectra taken from Table 4 in reference [28] are shown in Table 1.

Table 1. Lignin peaks used for classification taken from Adams[28]

<table>
<thead>
<tr>
<th>Peak</th>
<th>Mass</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>C₆H₅O</td>
<td>131.0497</td>
</tr>
<tr>
<td>B</td>
<td>C₁₀H₉O₂</td>
<td>161.0603</td>
</tr>
<tr>
<td>C</td>
<td>C₁₀H₁₁O₂</td>
<td>163.0759</td>
</tr>
<tr>
<td>D</td>
<td>C₁₀H₁₇O₃</td>
<td>179.0708</td>
</tr>
<tr>
<td>E</td>
<td>C₁₁H₁₃O₃</td>
<td>193.0865</td>
</tr>
<tr>
<td>F</td>
<td>C₁₁H₁₃O₄</td>
<td>209.0814</td>
</tr>
<tr>
<td>G</td>
<td>C₁₀H₁₇O₄</td>
<td>273.1127</td>
</tr>
<tr>
<td>H</td>
<td>C₁₀H₁₇O₄</td>
<td>285.1127</td>
</tr>
<tr>
<td>I</td>
<td>C₁₀H₂₁O₄</td>
<td>301.1440</td>
</tr>
<tr>
<td>J</td>
<td>C₁₀H₁₉O₅</td>
<td>303.1233</td>
</tr>
<tr>
<td>K</td>
<td>C₁₀H₁₉O₅</td>
<td>327.1233</td>
</tr>
<tr>
<td>L</td>
<td>C₁₈H₂₁O₆</td>
<td>333.1338</td>
</tr>
<tr>
<td>M</td>
<td>C₁₀H₁₉O₅</td>
<td>339.1233</td>
</tr>
<tr>
<td>N</td>
<td>C₂₀H₂₁O₆</td>
<td>357.1338</td>
</tr>
<tr>
<td>O</td>
<td>C₂₀H₂₃O₇</td>
<td>387.1444</td>
</tr>
<tr>
<td>P</td>
<td>C₂₂H₂₅O₇</td>
<td>401.1600</td>
</tr>
<tr>
<td>Q</td>
<td>C₂₂H₂₅O₈</td>
<td>417.1549</td>
</tr>
</tbody>
</table>

Mass Mountaineer software[29] was used for all data processing. For Principal Component Analysis (PCA) and Discriminant Analysis of Principal Components (DAPC), lignin peak abundances in each mass spectrum were normalized to the sum of the measured abundances for the lignin peaks. The exact mass tolerance was set to 25 mmu (0.025u) to accommodate measurement variability in the ForeST® database mass spectra.

Results

The heat map in Figure 1 shows only the lignin peaks from Table 1 for all 1313 mass spectra of angiosperms and 1708 mass spectra of gymnosperms. Darker points representing peaks with higher relative abundance show that the higher-m/z peaks, especially the syringyl m/z 193.0865 and m/z 209.0814, are more abundant in the angiosperm mass spectra. This trend is also visible in Figure 2 showing the averaged mass spectra for 102 randomly selected angiosperm mass spectra and 88 randomly selected gymnosperm mass spectra. Peaks corresponding to trisubstituted phenols (2,6-Dimethoxy-4-[prop-1-enyl] phenol, sinapaldehyde, and syringyl dimers) are clearly more abundant in the averaged DART mass spectrum of angiosperms.
Figure 1. Heat map showing the relative abundances of lignin peaks in the test data set comprised of 3021 DART mass spectra. Letter symbols at the top represent the compounds that show the most visually distinguishable differences between the two clades. The overall relative abundance of each lignin peak in the angiosperm and gymnosperm mass spectra (Figure S1) shows that only peaks A, C, and D (3-methyl-benzofuran, 2-methoxy-4-[prop-1-enyl] phenol, and coniferaldehyde) are more abundant in wood from gymnosperms than in angiosperms.
Figure 2. Averaged mass spectra for 102 randomly selected angiosperm and 88 mass spectra randomly selected gymnosperm mass spectra with the lignin peaks labeled.

Principal Component Analysis (Figure S2) shows separation between the two clades that is readily visible (Figure 4). The first eleven principal components cover 90.4% of the variance. The biplot (Figure S3) confirms that the peaks corresponding to syringaldehyde and sinapyl dimers are dominant classifiers in the DART-TOFMS mass spectra of angiosperms.

To test the validity of using lignin peaks to distinguish between wood from angiosperms and gymnosperms, 30% (907) of the training set mass spectra were randomly selected and removed from the training set to be treated as “unknowns” to determine classification accuracy. DAPC using 11 principal components correctly classified all but 63 of the 907 test mass spectra, for a classification accuracy of 93%.

CONCLUSIONS

In this study we evaluated 1708 mass spectra of softwood and 1313 mass spectra of hardwoods selected from all families included in the ForeST© database. It is evident that diagnostic molecules (e.g., sinapaldehyde) are a reliable proxy indicators of clade origin. The number of taxa evaluated confirms that these markers are reliable and reproducible.

The inference that can be reached from a rapid DART MS analysis allows for the verification of legal timber trade, and provides an approach to identify wrongful claims made to avoid import or export tariffs. In conclusion, the method we presented herein is efficient and can be used to assist in detecting fraudulently labeled timber imports.

Associated Content

The ForeST© database is available to qualified forensic laboratories by application to the U.S. Fish and Wildlife Service-National Fish and Wildlife Forensic Laboratory.

Supporting data

Relative abundance of lignin peaks in angiosperms and gymnosperms; Principal Component Analysis using the lignin peaks to distinguish between angiosperms and gymnosperms in all 3021 mass spectra randomly selected from the ForeST© database; biplot showing how the compounds associated with the lignin peaks contribute to the PCA classification; confusion matrix showing the correct and incorrect assignments for the 803 DART mass spectra selected for method validation.

Author information

Notes:

The authors declare the following competing financial interest(s): Robert B. Cody is a coinventor of the DART ion source and is employed by JEOL USA, Inc., the supplier of the AccuTOF-DART mass spectrometer. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the U.S. Fish and Wildlife Service or the U.S. Forest Service.
References


