Wood Science and Technology
https://doi.org/10.1007/s00226-019-01111-1

ORIGINAL

A protocol for automated timber species identification using metabolome profiling

V. Deklerck, et al. ***[full author details at the end of the article]***

Received: 11 January 2019

Abstract

Using chemical fingerprints for timber species identification is a relatively new, but promising technique. However, little is known about the effect of pre-processing spectral data parameter settings on the timber species classification accuracy. Therefore, this study presents an extensive and automated analysis method using the random forest machine learning algorithm on a set of highly valuable timber species from the Meliaceae family. Metabolome profiles were collected using direct analysis in real-time (DART™) ionisation coupled with time-of-flight mass spectrometry (TOFMS) analysis of heartwood specimens for 175 individuals (representing 10 species). In order to analyse variability in classification accuracy, 110 sets of data pre-processing parameter combinations consisting of mass tolerance for binning and relative abundance cut-off thresholds were tested. Furthermore, for each set of parameters (designated "binning/threshold setting"), a random search for one hyperparameter of interest was performed, i.e. the number of variables (in this case ions) drawn randomly for each random forest analysis. The best classification accuracy (82.2%) was achieved with 47 variables and a binning and threshold combination of 40 mDa and 4%, respectively. *Entandrophragma angolense* is mostly confused with *Entandrophragma candollei* and *Khaya anthotheca*, and several *Swietenia* species are confused with each other due to the high similarity of their chemical fingerprints. *Entandrophragma cylindricum*, *Entandrophragma utile*, *Khaya ivorensis*, *Lovoa trichilioides* and *Swietenia macrophylla* are easy to discriminate and show less misclassifications. The choice of parameter settings, whether it is in the data pre-processing (binning and threshold) or classification algorithm (hyperparameters), results in variability in classification accuracy. Therefore, a preliminary parameter screening is proposed before constructing the final model when using the random forest algorithm for classification. Overall, DART-TOFMS in combination with random forest is a powerful tool for species identification.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00226-019-01111-1) contains supplementary material, which is available to authorized users.

Published online: 20 June 2019

Introduction

The trade in illegal timber species is still problematic and results in economic and ecological problems (Jolivet and Degen 2012). Vlam et al. (2018) state that trade from illegal harvesting leads to detrimental effects on the rich and biodiverse tropical forests and hurts the local economy through tax evasion practices. A first indication of illegal timber trade is often inconsistent paperwork; however, this might not be sufficient to prove illegal trade, and timber species identification might be required. Dormontt et al. (2015) report that for screening suspect material and identification of illegally sourced timber, there is currently a lack of forensic identification tools. Although several techniques exist to identify or determine the origin of the timber, these techniques lack the harmonious collaboration to support legal timber trade (Dormontt et al. 2015).

Traditional wood anatomy, for example, focuses on the anatomical features, such as parenchyma, vessels, rays and fibres to identify a species. Although wood identification using conventional optical light microscopy is usually sufficient to identify a wood sample to genus level, the technique sometimes fails to determine the species (Dormontt et al. 2015; Gasson 2011) and separating closely related taxa can be problematic as well (Deklerck et al. 2017). Automated classification techniques based on imagery transverse cross-sections are increasingly investigated and show promising results (Hermanson and Wiedenhoeft 2011; Ravindran et al. 2018; Rosa da Silva et al. 2017). However, such techniques depend on the availability of wood transverse cross-sections and these are not always easy to obtain. The application of near infrared (NIR) spectroscopy is less common but has shown success for discriminating between different species or even determining the provenance of a limited number of species (Brunner et al. 1996; Tsuchikawa et al. 2003; Pastore et al. 2011). DNA analysis is successful in identifying the species, however, at the expense of increased time (several days) and cost (Dormontt et al. 2015). In addition, analysing DNA sequences from wood samples for genetic differentiation is challenging due to the difficulties in isolating DNA from dried and processed wood (Höltken et al. 2012). Yet, DNA microsatellites have been used successfully for tropical timber tracing (Degen et al. 2013; Jolivet and Degen 2012; Tnah et al. 2010). Alternatively, stable isotope analysis can be used, but it is only capable, to a certain extent, to determine the origin of the traded timber, not the species (Dormontt et al. 2015; Kagawa and Leavitt 2010).

Recently, another technique has been developed using direct analysis in real-time (DART) (Cody and Laramée 2005) time-of-flight mass spectrometry (TOFMS). Wood slivers are placed in a heated helium gas stream for an average of eight seconds, which leads to thermal desorption and ionisation of the molecules. This results in a unique chemical pattern based on secondary metabolites (metabolic or chemical fingerprint), which is used to identify the species in question. DART-TOFMS has proven to be successful to discern between several timber species (Deklerck et al. 2017; Espinoza et al. 2014; Lancaster and Espinoza 2012; McClure et al. 2015; Musah et al. 2015) or even keratin types (Price et al.

2018). However, little is known about the effect of parameter settings in the data processing on species classification accuracy.

In this study, the focus was on assessing the variability in classification accuracy using DART-TOFMS and different parameter settings employing the random forest algorithm on species from the Meliaceae family: *Entandrophragma angolense* (Welw.) C. DC., *E. candollei* Harms, *E. cylindricum* (Sprague) Sprague, *E. utile* (Dawe & Sprague) Sprague, *Khaya anthotheca* (Welw.) C. DC., *K. ivorensis* A. Chec., *Swietenia macrophylla* King, *S. mahagonie* (L.) Jacq. and *S. humilis* Zucc. species and *Lovoa trichilioides* Pierre ex Sprague. The *Entandrophragma* species, together with the *Swietenia*'s and *Khaya* species, are generally referred to as the mahoganies or the *acajous* from Africa (Beeckman 2003; Kasongo et al. 2019). The mahogany group is still one of the most highly valued and traded timber families (Gasson 2011), and it is vital to identify the species within the Meliaceae. Monthe et al. (2017) also state that the *Entandrophragma* genus is one of the most economically important genera in Africa. However, the four *Entandrophragma* species discussed in this paper are currently listed as vulnerable by the IUCN red list (www.iucnredlist.org). *L. trichilioides*, *K. anthotheca*, *K. ivorensis*, *S. macrophylla and S. humilis* are also listed as vulnerable. *S. mahagoni* is currently endangered and the entire *S.* genus is protected by the Convention on International Trade in Endangered Species (CITES, UNEP-WCMC). The popularity and threat of selective logging for *Swietenia macrophylla* justify its inclusion in CITES Appendix II (Braga et al. 2011; Gillies et al. 1999; Lemes et al. 2003, 2010; Novick et al. 2003). As Gasson (2011) indicated, custom officers could have trouble identifying a shipment of reddish brown wood and mark it as *Swietenia* sp., *Khaya* or *Entandrophragma* from Africa or a *dipterocarp* from SE Asia. Other than through wood anatomical descriptions and conventional optical light microscopy (for a full description see Electronic Supplementary Materials), several techniques have been used to identify timbers from the Meliaceae family. Höltken et al. (2012) developed DNA markers for the identification of several *Cedrela*, *Entandrophragma*, *Khaya* and *Swietenia* species in the Meliaceae family. Monthe et al. (2018) showed that *E. congoense* and *E. angolense* are distinct species based on their morphological traits and genetics. Rosa da Silva et al. (2017) identified and classified wood species, including several species from the Meliaceae family, using pattern recognition and anatomical characteristics of transverse cross-sections. Ravindran et al. (2018) identified 10 species in the Meliaceae family using transverse cross-sections and deep convolutional neural networks. Braga et al. (2011), Pastore et al. (2011) and Bergo et al. (2016) applied near infrared spectroscopy (NIR) for the identification of *S. macrophylla*.

There are two main objectives: (1) to provide an automated protocol to optimise timber identification using DART-TOFMS by determining the optimal parameter settings combined with random forest analysis and (2) to determine which species in the Meliaceae family are most often misclassified and as such pinpointing confounding species in timber trade.

Materials and methods

Sample collection and DART-TOFMS

Heartwood specimens of all species, *E. angolense*, *E. candollei*, *E. cylindricum*, *E. utile*, *K. anthotheca*, *K. ivorensis*, *L. trichilioides*, *S. macrophylla*, *S. humilis* and *S. mahagoni* were collected from different institutions and xylaria (see Supplementary Materials). Slivers (±2 to 3 cm long, max 4 mm wide) were sampled from these specimens and analysed using a DART-SVP ion source (IonSense, Saugus, MA, USA) coupled to a AccuTOF 4G time-of-flight mass spectrometer (JEOL USA, Peabody, MA, USA). The slivers were placed in the heated gas stream containing electronically excited helium atoms produced by the DART ion source. Spectra were obtained in positive ion mode, with the DART ion source parameters and mass spectrometer settings as defined in Evans et al. (2017), Lancaster and Espinoza (2012) and McClure et al. (2015) over a mass range of *m/z* 60 – 1100. A mass calibration standard (poly(ethylene glycol) 600 (Ultra, Kingstown, RI, USA)) was measured between every 5th sample. TSS Unity (Shrader Software Solutions, Inc., Grosse Pointe Park, MI, USA) data reduction software was used to export the txt-files of the mass-calibrated, centroided mass spectra.

Species classification analysis and settings optimisation

A heatmap of the spectra was constructed using Mass Mountaineer Mass Spectral Interpretation Tools software (massmountaineer.com). It shows the variation in intensity of ions within the samples and serves as a first visual check for differences between species. Random forest is a well-known machine learning technique and has been used in previous studies concerning timber identification using DART-TOFMS data (Deklerck et al. 2017; Finch et al. 2017; Paredes-Villanueva et al. 2018). It is important to obtain the optimal parameter settings for the model before concluding on species classification accuracy. Most algorithms using DART-TOFMS data require the use of a data frame, or matrix, where the columns and rows correspond to the ions (predictors, variables) and samples, respectively. Moreover, this requires binning ions with the same integer *m/z* values ("isobaric ions") based on the size of a mass tolerance expressed as milliDaltons (mDa) or millimass units (mmu). The minimum mass tolerance for discriminating between isobaric ions is related to the mass spectrometer resolving power (which is 10,000 using the full width at half maximum definition for this mass spectrometer) and mass accuracy (2–3 mDa). The larger the mDa tolerance level, the more isobaric ions are binned together and the fewer columns will be included in the data frame. Smaller mDa tolerance levels lead to less binning and more columns. This binning is required to have the same basis for comparison between samples. The abundance cut-off threshold setting determines which ion intensities will be included in the analysis. If, for a given sample, a certain intensity of an ion falls below the threshold that ion will not be included for that sample. Even though previous DART-TOFMS studies achieved

high classification accuracy, the majority of these reports relied upon chemometric methods that require selection of specific feature masses, either manually by inspection or by using mathematical methods such as Fisher ratio analysis (see Deklerck et al. 2017; Price et al. 2018). Little is known about the effect of choosing different binning and abundance parameters for data pre-processing with methods such as hierarchical clustering or random forest that do not rely upon feature selection. There are two steps in which parameters need to be optimised: first, the combination of an mDa binning and thresholding; second, the hyperparameters corresponding to the random forest model. In the latter case, the focus was on the number of variables drawn randomly for each random forest model, denoted by *mtry*.

The aforementioned txt-files were exported to Excel-csv files based on different binning and abundance cut-off threshold settings (mDa: 1–10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, threshold: 1–5, 110 combinations) using an in-house written R-code (RStudio Team 2015) written for automatic binning and thresholding (Supplementary Materials). When the abundance of an ion falls below the abundance threshold cut-off, it is removed from the txt-file before the data frame creation. In total, 100 random forest models were built (100 *mtry* settings, each with 500 decision trees) per mDa-bin/abundance cut-off threshold setting. The number of variables was randomly sampled between one and the maximum amount of variables in that dataset (depending on the binning, see above). It could be shown that the latter random grid search achieves better results, compared to standard grid search (Bergstra and Bengio 2012). Calculations were done using the *caret* (Kuhn 2018) and *mlbench* (Leisch and Dimitriadou 2010) R packages. Fivefold cross-validation was used in order to obtain an out-of-sample error estimate of the classification accuracy. Finally, a confusion matrix was constructed to determine which species were misclassified. To obtain a reliable out-of-sample estimate for the confusion matrix, fivefold cross-validation was performed on the complete dataset based on the best mDa binning, threshold and *mtry* settings. Finally, the five confusion matrices obtained were aggregated via summation.

Results and discussion

Figure 1 shows the chemical fingerprint heatmap for the different samples grouped by species. The *Swietenia* species have a very different chemotype compared to the other species. The ion around 871 *m*/*z* is an identifier for this group. Clear differences between the *Swietenia* species are not that obvious. *L. trichilioides* has a clear presence of ions around 395, 409 and 427 *m*/z, similar to the *Swietenia* species, but it is lacking the ion around 871 *m*/*z*. *K. ivorensis* differs by the high intensity of the ion around 338 *m*/*z*. It is more difficult to differentiate between the remaining species based on visual inspection of the heatmap only. *E. cylindricum* has a more or less clear presence of ions around 205, 713 and 829 *m*/*z* compared to the other species, while *E. utile* only has an ion signal around 829 *m*/*z*. There is a range of ions (429–574 *m*/*z*) that is present in *E. angolense*, but this is not consistent across all the samples of the species. *K. anthotheca* and *E. candollei* have a very similar chemotype and no immediate differences are visible.

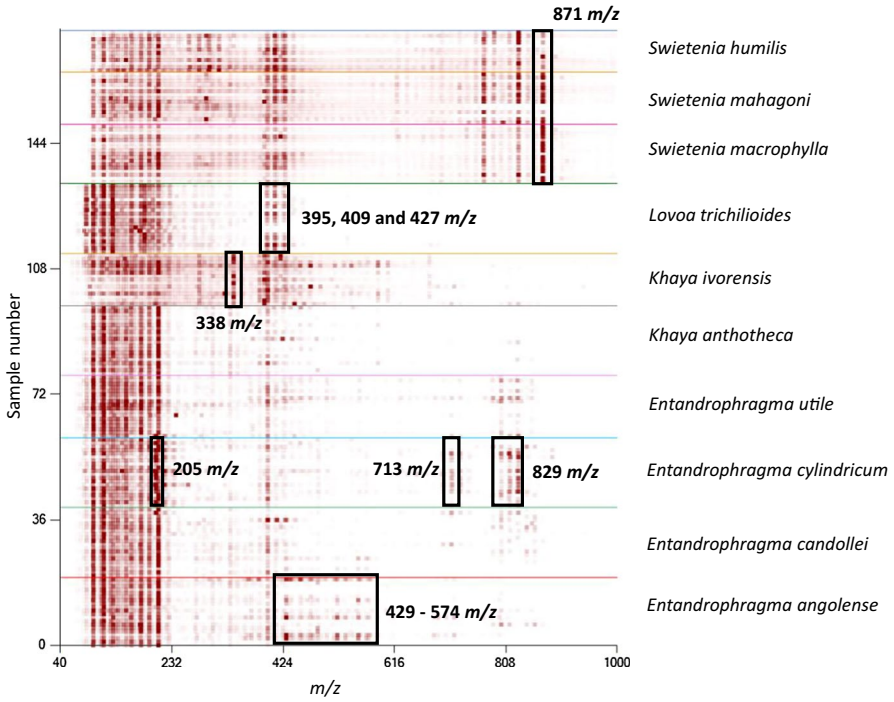

Fig. 1 Heatmap showing the presence of the ions for the different specimens per species. *Y*-axis: specimen number with the chemotype grouped per species, *x*-axis: mass to charge ratio (*m*/*z*) for the detected molecules. The colour intensity of the squares is an indication of the abundance of the ions in the specimen. Characteristic ions for each group of species are shown in black squares

Previous studies using DART-TOFMS and the random forest machine learning algorithm (Deklerck et al. 2017; Finch et al. 2017; Paredes-Villanueva et al. 2018) combined a 250 mDa and 1% threshold for their classification. However, little is known about the effect of these parameters on the random forest classification accuracy. Recent research by Beyramysoltan et al. (2018) mentions the effect of different threshold settings (1%, 2% or 3%) in combination with neural networks for identification of larva, pupa and adult life stages of carrion insects. A 2% threshold gave the highest prediction accuracy in their study. However, the screening space was limited, and the bin-size remained fixed at the resolution of the instrument (5 mDa). In the present study, the bin-size was varied as well, as optimising species identification was most important compared to potential compound identification. The heatmap showing the accuracy for the optimal model in terms of hyperparameters for the random forest, obtained after a random grid search as explained above, can be seen in Fig. 2. The overall best accuracy was 82.2%, using the 40 mDa binning and the 4% threshold setting combination with 500 trees and 47 variables for the final random forest model. Less optimal results were achieved when using higher thresholds and lower mDa bin-sizes. The 1 mDa binning and 5% threshold combination yields particularly inaccurate results. This

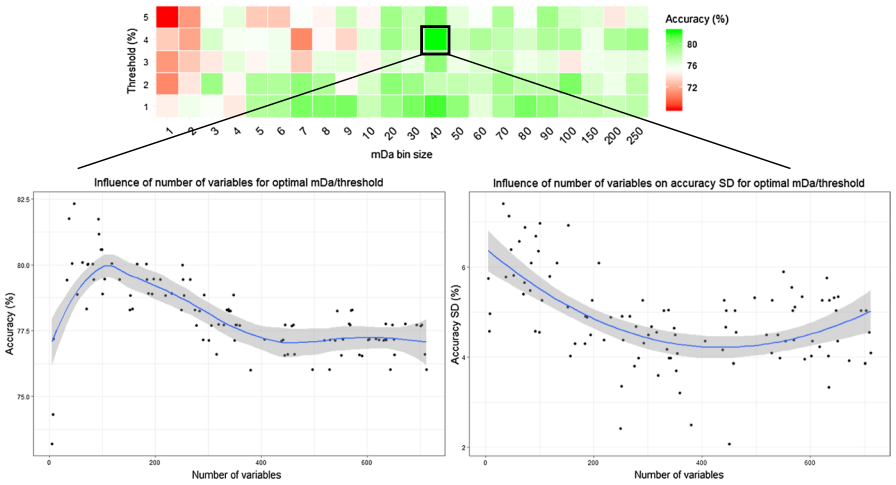

Fig. 2 Above: Heatmap showing the random forest classification accuracy for each mDa binning size/threshold combination. Below: For the best mDa bin/threshold (40–4%) combination, both the effect of the number of variables on the accuracy and the standard deviation are shown for 100 random forest runs with a random choice of variable number (mtry). The line represents the LOESS fitted regression curve with 95% confidence intervals (grey band)

might be explained by the choice of a narrow (low mDa) binning tolerance that is smaller than the statistical mass accuracy of the mass spectrometer resulting in a large number of ions. Further, the high abundance cut-off threshold leads to a limited selection of classifying ions. Additionally, including the so-called background or noise ions leads to a higher number of unnecessary variables. High classification accuracy was also achieved when taking a 40 mDa binning tolerance and a 1% threshold. As mentioned, previous studies used a 250 mDa binsize, mainly to reduce the number of variables and as such computation time, combined with a 1% threshold. However, this might lead to non-optimal classification accuracies as can be seen in Fig. 2. A preliminary screening is proposed to determine the best settings before moving towards the final model construction and validation. Choosing a large mass tolerance for binning increases the possibility that isobaric interferences and/or contaminants may be confused with species-specific compounds. For example, a peak at *m*/*z* 285.0763 is present in the DART mass spectra of *Dalbergia* species corresponding to a protonated compound with the formula $C_{16}H_{13}O_5$ (see also Lancaster and Espinoza 2012). A common interference may be observed at *m*/*z* 285.2794, corresponding to the formula $C_{18}H_{17}O_2$, for example, protonated octadecanoic acid or protonated methyl heptadecanoic acid. These species have masses that differ by 203 mDa and will not be separated with a binning tolerance of 250 mDa. There is a trade-off between compound selectivity and the number of features (columns) presented to the classification algorithm. Although the degree to which this influences the classification performance is beyond the scope of this paper, it is worthy of further investigation in future studies.

The effect of the number of variables for the best binning and threshold combination on the accuracy and standard deviation can also be seen in Fig. 2. Here, 100 random forest models were run (with random number of variables) for the 40 mDa binning and 4% threshold combination. The line represents the LOESS fitted regression curve with 95% confidence intervals (grey band). When using less than 30 variables, the accuracy is at a minimum, due to insufficient information for the classification algorithm to discriminate between species. The highest accuracy is achieved for 47 variables, after which the accuracy decreases with increasing number of variables. The lowest standard deviation, however, is achieved with 400 variables, although variation in the results is still observed.

The confusion matrix after fivefold cross-validation using the optimal pre-processing parameters and *mtry* settings can be seen in Fig. 3. This allows us to evaluate the results of species identification and understand the correlation with the heatmap shown in Fig. 1.

Samples from *E. angolense* are most confused with *E. candollei* and *K. anthotheca* and vice versa. This is in line with what can be noted in the heatmap, the three species have very similar chemical fingerprints. It has been shown before that *E. angolense* and *K. anthotheca* are difficult to adequately identify using texture analysis of anatomical transverse cross-sections (Rosa da Silva et al. 2017); however, there are minor anatomical differences (Supplementary Materials). However, the classification success in the study by Rosa da Silva et al. (2017) is still higher, even with more species included, compared to using chemical fingerprints (74.8% to 57.1%). Although there is a group of ions for *E. angolense* (429 – 574 *m/z*) that are

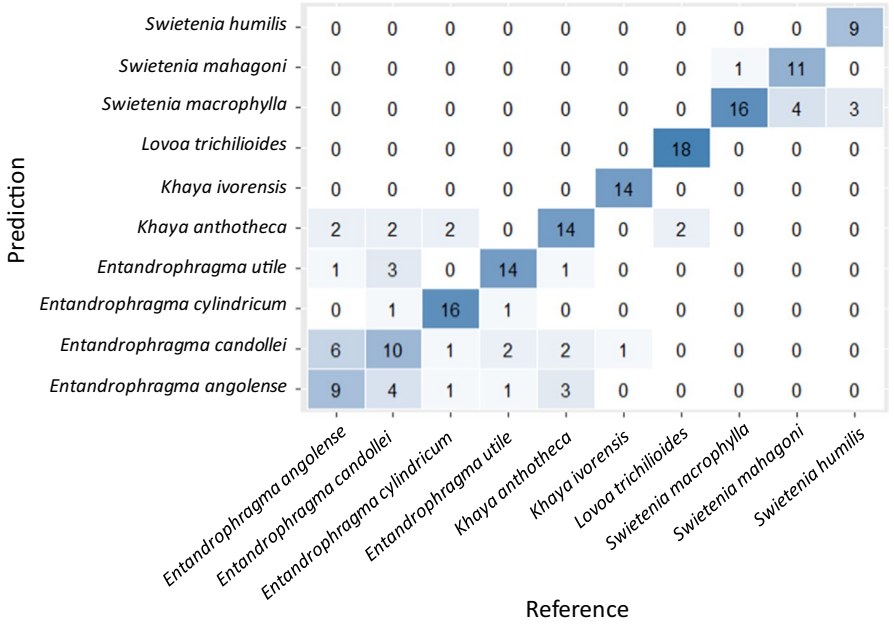

Fig. 3 Out-of-sample confusion matrix for the optimal mDa-threshold setting and random forest model

clearly unique, this is not consistent in all the samples. In the technique described by Rosa da Silva et al. (2017), *E. candollei* (97.3%) was very distinguishable from other *Entandrophragma* species. However, *E. angolense* and *E. utile* were harder to distinguish between each other, which is not a problem with chemical fingerprinting.

Entandrophragma cylindricum is classified with high accuracy, as expected, since it has a unique chemical fingerprint showing three identifying ion groups (Fig. 1). This is different from using conventional light microscopy, where it is hard to distinguish between *E. cylindricum* and *E. utile*, especially when sampling is suboptimal (Supplementary Materials).

There are some equivocal assignments between the different *Swietenia*—species, as their chemical fingerprint is similar. 25% of the *S. mahagoni* samples and a third of the *S. humilis* samples are being classified incorrectly as *S. macrophylla* (Fig. 3). The difficulty of separating the *Swietenia* species was already indicated by Höltken et al. (2012) using DNA analysis. Although the results are similar, as a technique, DART-TOFMS shows fewer constraints for obtaining the wood samples, and the only requirement is the need for heartwood. From the confusion matrix (Fig. 3), it is clear that the classification of *Swietenia macrophylla* separately was more successful and similar to the results obtained by Bergo et al. (2016), although in their paper only four species were considered and species-specific PLS-DA models based on NIR were used. In addition, the random forest classification accuracy would greatly increase if only this *Swietenia* species was considered. The combination of wood images and convolutional networks achieved high classification success for both *Swietenia macrophylla* and *mahagoni* (100% and 91.4%, respectively), yet *Swietenia humilis* was not included in that paper (Ravindran et al. 2018). Identification of species within the *Swietenia* genus using conventional light microscopy is less straightforward (Supplementary Materials). It should also be noted here that distinguishing between the genus *Swietenia* and *Khaya* using conventional light microscopy is not straightforward, although this is not a problem with DART-TOFMS spectra.

Both *K. ivorensis* and *L. trichilioides* perform well with only one sample for each species misidentified. As shown in Fig. 1, these species have a clear set of unique ions, which is consistent across all samples. In comparison, the method by Ravindran et al. (2018) to identify *K. ivorensis* performed poorly (76.1%), however, different species from the Meliaceae family were included in that study. DART-TOFMS could work complementary here, not only as a quick screening technique but also to help identify difficult species. It should be noted here that the classification success for the discussed techniques highly depends on the parameters used and which samples or species are included in the study.

Conventional light microscopy to determine the genus is feasible; however, differences on the species level are less obvious and might be challenging (Dormontt et al. 2015; Gasson 2011). Here, DART-TOFMS might play a role in discriminating species that are difficult to separate based on wood anatomy (Deklerck et al. 2017). Although the technique is relatively new, there have been studies on the effectiveness for difficult to identify species using wood anatomy, and satisfactory classification results were achieved (Deklerck et al. 2017; Lancaster and Espinoza 2012; McClure et al. 2015). The main advantages of DART-TOFMS for timber identification purposes are, as indicated, the versatility for forensic questions,

the quick screening time and the small sample size needed (heartwood sliver). These slivers are easy to obtain from, for example, imported musical instruments of which the species identity is uncertain. These types of products are harder to test using DNA analysis. Another difference is that DNA focusses on specific markers for species identification, where the identification using mass-spec data is done by classification or matching algorithms (Hartvig et al. 2015; Lancaster and Espinoza 2012). NIR is promising as well but there are too few studies to be able to assess the viability as a forensic technique. Deciding on species identification, however, should still be done by looking at inconsistencies in the paperwork together with a combination of different timber identification methods.

Conclusion

Little is known about the effect of pre-processing parameter settings for DART-TOFMS spectra when the random forest algorithm is used for classification. There is substantial variability in classification accuracy depending on the selection of the mass tolerance for binning and the abundance cut-off threshold. To tackle this problem, a framework is proposed allowing for an automated screening of the parameter space to retrieve the optimal settings for timber identification. The optimal combination for the Meliaceae species evaluated in this study is a mass tolerance bin-size of 40 mDa and a 4% abundance cut-off threshold, leading to an overall classification accuracy of 82.2%. The optimal number of variables included in the random forest analysis depends on these pre-processing parameters and species included in the analysis. *Entandrophragma angolense* is poorly discriminated from *Entandrophragma candollei* and *Khaya anthotheca*, and there is misclassification within the *Swietenia* genus due to very similar metabolic fingerprints. *Entandrophragma cylindricum*, *Entandrophragma utile*, *Khaya ivorensis*, *Lovoa trichilioides* and *Swietenia macrophylla* are easy to discriminate and show less misclassifications. Combining DART-TOFMS spectra with the preliminary parameter screening and the random forest algorithm allows for a consistent illegal timber identification pathway. However, a combination of different timber identification techniques is still advised for difficult and confounding species.

Acknowledgements The authors would like to thank Stijn Willem (UGent-Woodlab), Pam McClure and Erin Price (US Fish and Wildlife Forensic Laboratory) for their help with the sample preparation. This research was conducted under the HerbaXylaRedd BELSPO-project (Brain.be – code: BR/143/A3/HERBAXYLAREDD). The findings and conclusions in the article are those of the authors and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Beeckman H (2003) De microscopische schoonheid van mahonie (The microscopical beauty of mahogany). CR Interdisciplinair Tijdschrift Voor Conservering En Restauratie 4(2):18–27

Bergo MCJ, Pastore TCM, Coradin VTR, Wiedenhoeft AC, Braga JWB (2016) NIRS identification of Swietenia macrophylla is robust across specimens from 27 countries. IAWA J 37(3):420–430. https://doi.org/10.1163/22941932-20160144

Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. J Mach Learn Res 13:281–305. https://doi.org/10.1162/153244303322533223

Beyramysoltan S, Giffen JE, Rosati JY, Musah RA (2018) Direct analysis in real time-mass spectrometry and Kohonen artificial neural networks for species identification of larva, pupa and adult life stages of carrion insects. Anal Chem 90:9206–9217. https://doi.org/10.1021/acs.analchem.8b01704

Braga JWB, Pastore TCM, Coradin VTR, Camargos JAA, da Silva AR (2011) The use of Near InfraRed Spectroscopy to Identify solid wood specimens of *Swietenia Macrophylla* (CITES Appendix II). IAWA J 32(2):285–296

Brunner M, Eugster R, Trenka E, Bergamin-Strotz L (1996) FT-NIR spectroscopy and wood identification. Holzforschung 50(2):130–134. https://doi.org/10.1515/hfsg.1996.50.2.130

Cody RB, Laramée JA (2005) Versatile new ion source for the analysis of materials in open air under ambient conditions. Anal Chem 77(8):2297–2302

Degen B, Ward SE, Lemes MR, Navarro C, Cavers S, Sebbenn AM (2013) Verifying the geographic origin of mahogany (*Swietenia macrophylla* King) with DNA-fingerprints. Forensic Sci Int Gen 7(1):55–62. https://doi.org/10.1016/j.fsigen.2012.06.003

Deklerck V, Finch K, Gasson P, Van den Bulcke J, Van Acker J, Beeckman H, Espinoza E (2017) Comparison of species classification models of mass spectrometry data: kernel discriminant analysis vs random forest; a case study of Afrormosia (*Pericopsis elata* (Harms) Meeuwen). Rapid Commun Mass Sp 31(May):1582–1588. https://doi.org/10.1002/rcm.7939

Dormontt EE, Boner M, Braun B, Breulmann G, Degen B, Espinoza E, Gardner S, Guillery P, Hermanson JC, Koch G, Lee SL, Kanashiro M, Rimbawanto A, Thomas D, Wiedenhoeft AC, Yin Y, Zahnen J, Lowe AJ (2015) Forensic timber identification: it's time to integrate disciplines to combat illegal logging. Biol Conserv 191:790–798. https://doi.org/10.1016/j.biocon.2015.06.038

Espinoza EO, Lancaster CA, Kreitals NM, Hata M, Cody RB, Blanchette RA (2014) Distinguishing wild from cultivated agarwood (*Aquilaria* spp.) using direct analysis in real time and time of-flight mass spectrometry. Rapid Commun Mass Sp 28(3):281–289. https://doi.org/10.1002/rcm.6779

Evans PD, Mundo IA, Wiemann MC, Chavarria GD, McClure PJ, Voin D, Espinoza E (2017) Identification of selected CITES-Protected Araucariaceae using DART TOFMS. IAWA J 38(2):266–281

Finch K, Espinoza E, Jones FA, Cronn R (2017) Source identification of western Oregon Douglas-Fir wood cores using mass spectrometry and random forest classification. Appl Plant Sci 5(5):1–49. https://doi.org/10.3732/apps.1600158

Gasson P (2011) How precise can wood identification be? Wood anatomy's role in support of the legal timber trade, especially CITES. IAWA J 32(2):137–154

Gillies ACM, Navarro C, Lowe AJ, Newton AC, Hernández M, Wilson J, Cornelius JP (1999) Genetic diversity in mesoamerican populations of mahogany (*Swietenia macrophylla*), assessed using RAPDs. Heredity 83(6):722–732. https://doi.org/10.1046/j.1365-2540.1999.00626.x

Hartvig I, Czako M, Kjær ED, Nielsen LR, Theilade I (2015) The use of DNA barcoding in identification and conservation of rosewood (*Dalbergia* spp.). PLoS ONE 10(9):e0138231. https://doi.org/10.1371/journal.pone.0138231

Hermanson JC, Wiedenhoeft A (2011) A brief review of machine vision in the context of automated wood identification systems. IAWA J 32(2):230–250

Höltken AM, Schröder H, Wischnewski N, Degen B, Magel E, Fladung M (2012) Development of DNA-based methods to identify CITES protected timber species: a case study in the Meliaceae family. Holzforschung 66(1):97–104. https://doi.org/10.1515/HF.2011.142

Jolivet C, Degen B (2012) Use of DNA fingerprints to control the origin of sapelli timber (*Entandrophragma cylindricum*) at the forest concession level in Cameroon. Forensic Sci Int Gen 6(4):487–493. https://doi.org/10.1016/j.fsigen.2011.11.002

Kagawa A, Leavitt SW (2010) Stable carbon isotopes of tree rings as a tool to pinpoint the geographic origin of timber. J Wood Sci 56(3):175–183. https://doi.org/10.1007/s10086-009-1085-6

Kasongo E, Louppe D, Monthe F, Hardy O, Mbele Lokanda FB, Hubau W, Van den Bulcke J, Van Acker J, Beeckman H, Bourland N (2019) Enjeux et amélioration de gestion de Entandrophragma: arbres africains potentiellement en danger (Management problems and improvements of Entandrophragma: African trees are potentially in danger). Bois et Forêts de Tropiques 339:75–94

Kuhn M (2018) Package classification and regression training ('caret'). Repository CRAN, R package

Lancaster C, Espinoza E (2012) Analysis of select *Dalbergia* and trade timber using direct analysis in real time and time-of-flight mass spectrometry for CITES enforcement. Rapid Commun Mass Sp 26(9):1147–1156. https://doi.org/10.1002/rcm.6215

Leisch F, Dimitriadou E (2010) Package machine learning benchmark problems ('mlbench'). Repository CRAN, R package

Lemes MR, Gribel R, Proctor J, Grattapaglia D (2003) Population genetic structure of mahogany (*Swietenia macrophylla* King, Meliaceae) across the Brazilian Amazon, based on variation at microsatellite loci: implications for conservation. Mol Ecol 12(11):2875–2883. https://doi.org/10.1046/j.1365-294X.2003.01950.x

Lemes MR, Dick CW, Navarro C, Lowe AJ, Cavers S, Gribel R (2010) Chloroplast DNA microsatellites reveal contrasting phylogeographic structure in mahogany (*Swietenia macrophylla* King, Meliaceae) from Amazonia and Central America. Trop Plant Biol 3(1):40–49. https://doi.org/10.1007/s12042-010-9042-5

McClure PJ, Chavarria GD, Espinoza E (2015) Metabolic chemotypes of CITES protected *Dalbergia* timbers from Africa, Madagascar, and Asia. Rapid Commun Mass Sp 29(9):783–788. https://doi.org/10.1002/rcm.7163

Monthe FS, Duminil J, Tosso F, Migliore J, Hardy OJ (2017) Characterization of microsatellite markers in two exploited African trees, *Entandrophragma candollei* and *E. utile* (Meliaceae). Appl Plant Sci 5(2):1600130. https://doi.org/10.3732/apps.1600130

Monthe FK, Duminil J, Kasongo Yakusu E, Beeckman H, Bourland N, Doucet J-L, Sosef MSM, Hardy OJ (2018) The African timber tree *Entandrophragma congoense* (Pierre ex De Wild.) A. Chev is morphologically and genetically distinct from *Entandrophragma angolense* (Welw.) C.DC. Tree Genet Genomes 14(5):66. https://doi.org/10.1007/s11295-018-1277-6

Musah RA, Espinoza EO, Cody RB, Lesiak AD, Christensen ED, Moore HE, Maleknia S, Drijfhout FP (2015) A high throughput ambient mass spectrometric approach to species identification and classification from chemical fingerprint signatures. Sci Rep 5(February):11520. https://doi.org/10.1038/srep11520

Novick RR, Dick C, Lemes MR, Navarro C, Caccone A, Bermingham E (2003) Genetic structure of Mesoamerican populations of Big-leaf mahogany (*Swietenia macrophylla*) inferred from microsatellite analysis. Mol Ecol 12(11):2885–2893. https://doi.org/10.1046/j.1365-294X.2003.01951.x

Paredes-Villanueva K, Espinoza E, Ottenburghs J, Sterken MG, Bongers F, Zuidema PA (2018) Chemical differentiation of Bolivian *Cedrela* species as a tool to trace illegal timber trade. Forestry 00:1–11. https://doi.org/10.1093/forestry/cpy019

Pastore TCM, Braga JWB, Coradin VTR, Magalhães WLE, Okino EYA, Camargos JAA, Bonzon de Muñiz GI, Bressan OA, Davrieux F (2011) Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: discrimination of true mahogany, cedar, andiroba, and curupixá. Holzforschung 65(1):73–80. https://doi.org/10.1515/HF.2011.010

Price ER, McClure PJ, Jacobs RL, Espinoza EO (2018) Identification of rhinoceros keratin using direct analysis in real time time-of-flight mass spectrometry and multivariate statistical analysis. Rapid Commun Mass Spectrom 32:1–7. https://doi.org/10.1002/eqe.3063

Ravindran P, Costa A, Soares R, Wiedenhoeft AC (2018) Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks. Plant Methods 14(1):1–10. https://doi.org/10.1186/s13007-018-0292-9

Rosa da Silva N, De Ridder M, Baetens JM, Van den Bulcke J, Rousseau M, Martinez Bruno O, Beeckman H, Van Acker J, De Baets B (2017) Automated classification of wood transverse cross-section micro-imagery from 77 commercial Central-African timber species. Ann For Sci 74(2):30. https://doi.org/10.1007/s13595-017-0619-0

Tnah LH, Lee SL, Ng KKS, Faridah Q-Z, Faridah-Hanum I (2010) Forensic DNA profiling of tropical timber species in Peninsular Malaysia. Forest Ecol Manag 259:1436–1446. https://doi.org/10.1016/j.foreco.2010.01.017

Tsuchikawa S, Inoue K, Noma J, Hayashi K (2003) Application of near-infrared spectroscopy to wood discrimination. J Wood Sci 49(1):29–35. https://doi.org/10.1007/s100860300005

UNEP-WCMC (n.d.) Convention on international trade in endangered species of Wild Fauna and Flora. Appendices I, II and III. Retrieved from https://cites.org/sites/default/files/notif/E-Notif-2016-068-A.pdf. Accessed 4 Jan 2019

Vlam M, de Groot GA, Boom A, Copini P, Laros I, Veldhuijzen K, Zakamdi D, Zuidema PA (2018) Developing forensic tools for an African timber: regional origin is revealed by genetic characteristics, but not by isotopic signature. Biol Conserv 220(January):262–271. https://doi.org/10.1016/j.biocon.2018.01.031

Affiliations

V. Deklerck[1,2] · T. Mortier[3] · N. Goeders[1,3] · R. B. Cody[4] · W. Waegeman[3] · E. Espinoza[5] · J. Van Acker[1] · J. Van den Bulcke[1] · H. Beeckman[2]

✉ V. Deklerck
victor.deklerck@ugent.be

[1] UGent-Woodlab, Laboratory of Wood Technology, Department of Environment, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000 Ghent, Belgium

[2] Service of Wood Biology, Royal Museum for Central Africa (RMCA), Leuvensesteenweg 13, 3080 Tervuren, Belgium

[3] Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure Links 653, 9000 Ghent, Belgium

[4] JEOL USA, Inc., Peabody, MA 01960, USA

[5] U.S. National Fish and Wildlife Forensic Laboratory, 1490 East Main Street, Ashland, OR, USA