ENVIRONMENTAL RESEARCH
LETTERS

RECEIVED
18 January 2023

REVISED
16 March 2023

ACCEPTED FOR PUBLICATION
28 March 2023

PUBLISHED
19 April 2023

LETTER

A new method for the timber tracing toolbox: applying multi-element analysis to determine wood origin

Laura E Boeschoten[1,*], Mart Vlam[1,2], Ute Sass-Klaassen[1,2], Barbara Rocha Venâncio Meyer-Sand[1], Ulfa Adzkia[3], Gaël D U Bouka[4], Jannici C U Ciliane-Madikou[4], Nestor L Engone Obiang[5], Mesly Guieshon-Engongoro[4], Joël J Loumeto[4], Dieu-merci M F Mbika[4], Cynel G Moundounga[6], Rita M D Ndangani[4], Dyana Ndiade Bourobou[6], Mohamad M Rahman[3], Iskandar Z Siregar[3], Steve N Tassiamba[7], Martin T Tchamba[7], Bijoux B L Toumba-Paka[4], Herman T Zanguim[7], Pascaline T Zemtsa[7] and Pieter A Zuidema[1]

[1] Forest Ecology and Forest Management, Wageningen University and Research, Wageningen, The Netherlands
[2] Forest and Nature Management, Van Hall Larenstein University of Applied Sciences, Velp, The Netherlands
[3] Faculty of Forestry and Environment, IPB University, Bogor, Indonesia
[4] Laboratory of Biodiversity and Ecosystems and Environmental Management, Faculty of Science and Technology, Marien Ngouabi University, Brazzaville, Congo
[5] Herbier National du Gabon, Institut de Pharmacopée et de Médecine Traditionelle, Libreville, Gabon
[6] Institute for Research in Tropical Ecology (IRET/CENAREST), Libreville, Gabon
[7] Laboratory of Environmental Geomatics, Department of Forestry, Faculty of Agronomy and Agricultural Sciences, University of Dschang, Dschang, Cameroon
[*] Author to whom any correspondence should be addressed.

E-mail: laura.boeschoten@wur.nl

Keywords: illegal logging, wood chemistry, random forest, blind sample analysis, tropical timber, ICP-MS, timber forensics

Supplementary material for this article is available online

Abstract

To effectively reduce illegal timber trade, law enforcers need forensic methods to independently verify claims of wood origin. Multi-element analysis of traded plant material has the potential to be used to trace the origin of commodities, but for timber it has not been tested at relevant large scales. Here we put this method to the test, by evaluating its tracing accuracy for three economically important tropical timbers: Azobé and Tali in Central Africa (22 sites) and Red Meranti on Borneo (9 sites). Wood samples from 991 trees were measured using Inductively Coupled Plasma Mass Spectrometry and element concentrations were analysed to chemically group similar sites (clustering) and assess accuracy of tracing samples to their origin (Random Forest models). For all three timbers, we found distinct spatial differences in chemical composition. In Central Africa, tracing accuracy was 86%–98% for regional clusters of chemically similar sites, with accuracy depending on the tracing question. These clusters were 50–800 km apart and tracing accuracy was highest when combining the two timbers. Tracing accuracy of Red Meranti on Borneo was 88% at the site level. This high accuracy at a small scale may be related to the short distances at which differences in soil type occur on Borneo. A blind sample analysis of 46 African timber samples correctly identified the origin of 70%–72% of the samples, but failed to exclude 70% of the samples obtained from different species or outside the study area. Overall, these results illustrate a high potential for multi-element analysis to be developed into a timber tracing tool which can identify origin for multiple species and can do so at a within-country scale. To reach this potential, reference databases need to cover wider geographic areas and represent more timbers.

1. Introduction

Illegal timber trade ranks among the most profitable wildlife crimes and the false declaration of origin is one of the main types of timber fraud (Dormontt et al 2015, Hoare 2015). This harms people, ecosystems and local economies and impairs initiatives to increase sustainable forest management and timber trade (Hoare and Uehara, 2022). The lack of methods to independently verify timber origin limits effective law enforcement and full supply chain transparency, as timber origin is currently verified by external documents and tags. Therefore, novel methods to trace timber origin based on intrinsic wood properties, rather than external documents, can play a crucial role in stopping illegal timber trade (Lowe *et al* 2016).

The multi-element analysis of wood is a promising new method for origin identification (Boeschoten *et al* 2022). A large number of elements (such as Mg, Ca, La) are measured simultaneously using mass spectrometry and based on this elemental composition, an origin-specific fingerprint is defined. By comparing the chemical fingerprint to an established geo-referenced dataset, the most likely origin of a sample can be determined. This can help to answer two relevant tracing questions: origin verification and assignment (Deklerck 2022). The verification question addresses whether a wood sample came from a certain (claimed) origin. On the other hand, the assignment question addresses which location a wood sample most likely originated from in case sample origin is entirely unknown. Both questions are relevant in tracing but require a different statistical approach. Multi-element analysis was already found to be useful for assignment and verification of a variety of commodities, such as green asparagus, bananas and tea (Gonzalvez, *et al* 2009, Ma *et al* 2016, Richter *et al* 2019), and has been successful at both regional and continental scales (Joebstl *et al* 2010, Baroni *et al* 2015). For timber, further research across different countries and species is essential to understand how the method operates (Boeschoten *et al* 2022).

Multi-element analysis may provide advantages in origin determination compared to other timber tracing methods under investigation, such as stable isotopes, genetic or direct analysis in real time - time of flight mass spectrometry (DART-TOFMS) analyses (Lee *et al* 2015, Vlam *et al* 2018, Deklerck *et al* 2020). As wood chemical composition is thought to be influenced by soil chemistry (Boeschoten *et al* 2022), the chemical profile of multiple species might overlap. Even if species differ in elemental composition (Amais *et al* 2021), the variation between sites could still be reflected in multiple species. In that case, their reference data could be combined, which would reduce the reference data collection effort drastically (Gasson *et al* 2021). This is contrary to other methods that need a unique fingerprint for each species.

Challenges are expected in the application of elemental analysis for timber tracing as the chemical composition of timber is more complex than most other commodities. First, wood formation is a process of many years and the traded part of the tree, the heartwood, is thus influenced by environmental and ontogenetic changes over the tree's lifetime (Hietz *et al* 2015, Scharnweber *et al* 2016, Hevia *et al* 2017). Many other commodities are only affected by the environmental conditions of one year, which may lead to a more unique fingerprint per region. Second, the method is only useful for tracing timber when variation is present with defined differences between origins (Hevia *et al* 2017). The applicability of the method thus depends on natural variation in elemental composition within a species distribution range, but limited research is available on this for trace elements in wood. Wood elemental composition has been associated with soil physical and chemical properties (Boeschoten *et al* 2022), so the method may be especially successful in regions where soil chemistry exhibits strong spatial variation.

This is the first study to put multi-element analysis for timber tracing to practice. We did so by determining its applicability for three economically important tropical timber species: two from Central Africa, traded as Azobé (*Lophira alata*) and Tali (*Erythrophleum ivorense* and *E. suaveolens*), and Red Meranti from Borneo (*Shorea* spp.). We first analysed whether the two timbers from Central Africa shared their variation in elemental composition across the study region. Then we developed classification models for all timber species based on the elemental composition, addressing both origin verification and assignment accuracy. We also investigated at what scale we found tracing potential for these three timber species. In order to understand what determines elemental variation across a landscape, we visualised how key elements varied across the study area and assessed whether wood elemental compositions correlated with soil variables that affect element uptake. Lastly, we simulated a real-life tracing case by applying the classification models to a set of blind samples, to determine their most likely origin.

2. Materials and methods

2.1. Study design

The study was conducted on three valuable internationally traded tropical timber species: two from Central Africa (Azobé; *Lophira alata*, Ochnaceae and Tali; *Erythrophleum ivorense* and *E. suaveolens*, Fabaceae), and one from South-East Asia (Red Meranti; *Shorea* spp., Dipterocarpaceae). They represent a single-species timber, Azobé, and two multi-species timbers. Tali timber is sourced from two botanical species and Red Meranti is generally accepted to be sourced from ten species (Klaassen 2018). The botanical species that make up Tali and Red Meranti are hard to distinguish in the field, therefore they were only identified to species level if leaf, flower and/or fruit material was available. Otherwise samples were identified at genus level. As many individual trees did not have branches below three meters and we had no equipment available for sampling higher branches, fresh leaves could not be collected at every tree. Furthermore, as the aim of this study was to test chemical differences at the timber level, we did not pursue further species-level analysis. All three timbers grow in evergreen and moist deciduous forests.

Samples in Central Africa were taken from 22 study sites across the main timber exporting countries: Cameroon (seven sites), Gabon (seven sites) and the Republic of the Congo (eight sites), between September 2019 and April 2022 (figure 2(A)). Samples were taken from both species at all sites except for three locations: at two sites only Azobé was sampled and at one site only Tali was sampled due to low occurrence of the other species. Samples on Borneo were taken from nine study sites across the provinces West, East and Central Kalimantan, Indonesia, between January 2020 and February 2022 (figure 2(B)). All sites were natural forest concessions, accessed in collaboration with the operating forestry companies.

2.2. Sample collection

At each site, we sampled heartwood from 20–30 trees per timber species. Sampled trees within one site were located between 100 m and 5 km apart. Trees were either standing or recently felled and were of at least 30 cm diameter at breast height (DBH). A heartwood sample was collected from each tree as an increment core (Haglöf Increment borer 350 mm × 5,15 mm; $n = 238$), with a FAMAG plug cutter of 15 mm diameter ($n = 562$), as a wood chunk ($n = 23$) or as a wood powder sample obtained with an electrical drill ($n = 170$). All samples were taken 10–20 cm into the tree. Additionally, global positioning system (GPS) coordinates and DBH were recorded. This resulted in a geolocated database of 179 Red Meranti, 420 Azobé and 394 Tali samples. For every tree, we also obtained soil variables at their growth locations from www.soilgrids.org (Poggio *et al* 2021): pH, cation exchange capacity (CEC), clay content and soil organic matter content.

2.3. Chemical analysis

Chemical analysis was performed following Boeschoten *et al* (2022). In short, a 1.0 g subsample of heartwood was cut from 3–5 cm of wood and dissolved in 70% HNO_3 by heating in a microwave (CEM Mars 6). In the resulting solutions, element concentrations were determined by inductively coupled plasma mass spectrometry (NexION 350D, PerkinElmer). Lichen and/or Rye grass were used as certified reference material to guarantee measurement accuracy. The lowest detection limit per element was calculated as three times the intensity of that element in a blank standard. If elements were found in quantities below the detection limit in more than 200 samples (the equivalent of a quarter of the trees), they were excluded. This resulted in a multi-elemental composition of 41 elements at concentrations varying between 0.001 $\mu g\ kg^{-1}$ (terbium) and 6 $g\ kg^{-1}$ (potassium) in the wood samples.

2.4. Statistical analysis

2.4.1. Species differences across sites

All statistical analyses were performed in R version 4.1.0 (R Core Team 2021). Differences in the multielemental composition of Azobé and Tali were tested by distance based redundancy analysis (db-RDA) using the vegan package (Oksanen *et al* 2020), based on Chord distances. In addition to this multivariate analysis, we investigated whether individual elemental concentrations differed between the two species by constructing mixed effect models for all elements using site as random factor. These were performed using the lme4 package (Bates *et al* 2015).

2.4.2. Random forest (RF) classification of sites and clusters

We performed multiple RF analyses for site and cluster classification using the randomForest package (Liaw and Wiener 2002), see table 1 for an overview. The two potential tracing questions were addressed: assignment (*where does this sample most likely originate from?*) and verification (*does this sample come from location X?*). First we tested origin assignment for the three timber species separately (three site-level RF models). Additional to the absolute element concentrations we included elemental ratios, based on the elements that were most important in the RF assignment models. All 1:1 ratios for the top ten elements of each RF model were tested as additional variables in the RF assignment models, as well as ratios that were previously described to vary geographically (Hevia *et al* 2017). Only ratios that improved assignment success were included. Assignment success per model was then calculated as: 100% minus the RF out-of-bag error rate. We also fitted a fourth RF model to assign site of origin combining both Azobé and Tali data.

We continued with a cluster analysis to test similarities between sites. Sites were aggregated into clusters based on divisive clustering using the cluster package (Maechler *et al* 2021). Site distances were calculated using Mahanalobis distances, based on standardised mean elements and elemental ratios, using the HDMD package (McFerrin 2013).

A second set of RF assignment analyses was then performed to assign samples to clusters (cluster-level models). Similar to the site assignment, we developed three species specific RF assignment models (one for each species) and a fourth where Azobé and Tali were combined. As clustering resulted in an imbalanced sample design, we applied Synthetic Minority Over-Sampling Technique (SMOTE) resampling using the UBL package (Branco, *et al* 2016) and we set aside 30% of the samples in each iteration before resampling to ensure full independence of the test samples. Assignment success of the final models was evaluated using 50-fold cross-validation, calculated as: 100% minus the percentage of incorrect assigned trees.

Lastly, a third set of RF classification analyses was performed to develop verification models rather than assignment models. Eight RF verification models were developed, one model per cluster for both Azobé and Tali. These models each classified two groups: one group included all reference samples from that cluster and the other group included all other reference samples. Before fitting the models, SMOTE resampling was applied to each reference dataset and 30% was set aside per iteration to ensure full independence of the test samples.

2.4.3. Soil drivers of chemical composition

To test which soil variables were linked to wood chemical variation, we focused on the top five elements or ratios of the cluster level RF models. Soil data from Soilgrids was centred and scaled (Poggio *et al* 2021). For each of the five top variables, a full model was dredged to generate a list of candidate models ranked by Akaike information criterion (AICc) using the MuMIn package (Barton 2020). Full model averaging was performed using candidate models within the top two ΔAICc units (Burnham and Anderson 2002).

2.5. Blind sample collection and classification

Parallel to the main sampling campaign, a second set of samples was arranged by a third party. They collected additional samples of which 46 were anonymised and labelled as Azobé and Tali. No blind samples were taken of Meranti. Each sample fitted one of three categories: (1) sampled within the same reference sites during the main sampling campaign, (2) sampled within 20 km of the reference sites but not from the same campaign, (3) sampled outside the reference sites and/or from species not covered in the reference samples. Multi-element composition was measured in the same way as the reference samples.

Using the RF classification models of the reference dataset, both tracing questions were addressed for all blind samples (origin verification and assignment, see section 2.4.2). Origin verification was tested based on the eight RF cluster verification models (V1–V8, table 1). Then, before origin assignment, it was tested whether the blind samples should be assigned at all or whether they should be excluded based on their chemical composition using multiple outlier statistics (univariate, principle component analysis, Chord distances and isotree distances). Subsequently, all blind samples were assigned to their most likely cluster (model A4 for both Azobé and Tali) and site (A5 for Azobé and A8 for Tali) of origin (table 1). After verification and assignment, we performed two robustness tests, to take into account uncertainty in the classifications because of (1) a small measurement error or (2) uncertainty in the RF models themselves. If the result of one of the robustness tests was that a blind sample was not classified to a single origin in more than 80% of the cases, the sample was recorded as ‘not assigned’. A detailed description of the blind samples analysis is attached in supplementary note S1.

3. Results

3.1. Species and site differences in Central Africa

The two timbers from Central Africa, Azobé and Tali, had a significantly different chemical fingerprint (figure 1). This was also reflected in individual mixed effect models: 35 of the 41 element concentrations were significantly different (table S1). Across our study area, 28 elements were more abundant in Azobé, whereas 7 were more abundant in Tali. However, the second db-RDA axis also reflected a significant common origin pattern. Both timbers shared high or low concentrations of certain elements (such as As, Co or Y) at the same sites. This reflects a common spatial pattern across the study area.

3.2. Origin verification and assignment models

In Central Africa, the cluster analysis identified eight chemical clusters, including 1–7 sites each (figure 2(A)). The distances between sites from the same cluster were 20–340 km, whereas some sites that were assigned to different chemical clusters were also as close as 50 km. Furthermore, all chemical clusters of sites were geographically well defined, i.e. there was no spatial mixing of clusters. On Borneo there were five chemical clusters of 1–4 sites each (figure 2(B)). However, in contrast to Central Africa, on Borneo the sites that were chemically similar were not always geographically close: one cluster consisted of one site in East, one in Central and two sites in West Kalimantan. Additionally, from three sites in West Kalimantan that were located within 40 km, one was chemically distinct from the other two.

The eight RF verification models for clusters in Central Africa were successful in verifying origin (table 1). They yielded a verification accuracy between 91.6% (green cluster, Republic of the Congo, model V7) and 97.6% (darkblue cluster, west Gabon, model V4, table 1), with 96.0% accuracy on average. Because of the lack of geographical correspondence in the clusters on Borneo, no cluster-level verification was performed. Accuracy was slightly lower in the RF assignment analysis (table 1). The individual model for Red Meranti performed best at assigning the site (model A7, 87.6% correct assignment) and cluster of origin (model A3, 92.6%). As expected, the cluster assignments were better than the site assignments, because the grouping into clusters with high similarity increases the chance of correct classification. Successful site pairwise assignment in the RF

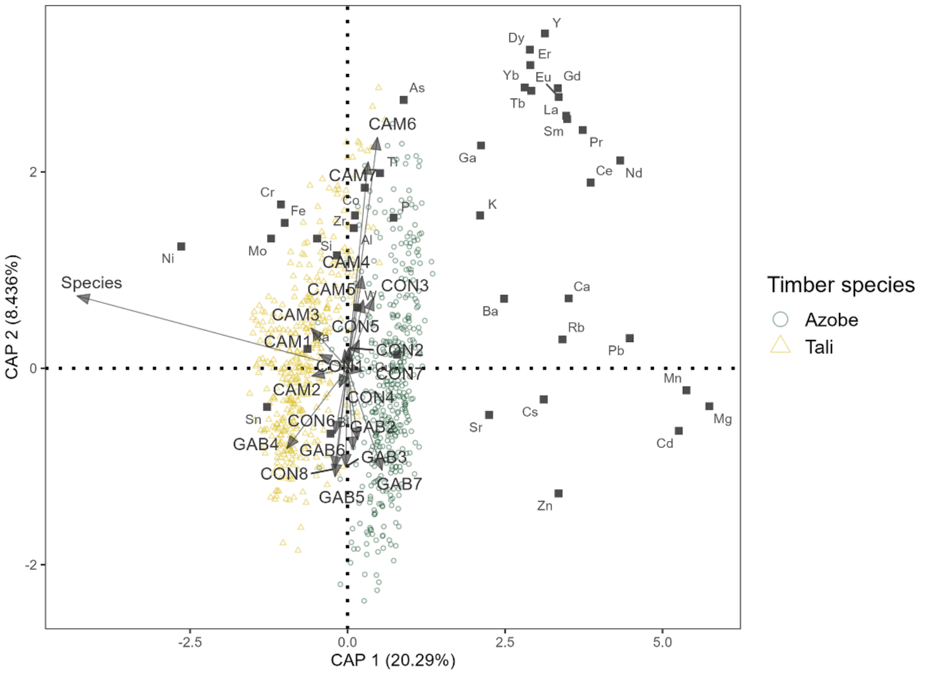

Figure 1. Distance based redundancy analysis (db-RDA) of heartwood elemental composition of the African timber species Azobé and Tali. Green and yellow dots indicate relative position of the wood samples, based on Chord distances of square root transformed element concentrations, coloured by timber species. Grey squares indicate weight and direction of the element concentrations in the wood. Vectors indicate the weight and the direction of the significant explanatory variables: timber species and sites of origin. Percentages in axis labels denote the percentage of total variance explained by this component. The Indonesian Red Meranti was not included as it does not co-occur with the other two species.

models went up with increasing distance in Central Africa but not on Borneo (figure S1).

We tested two more assignment models for the samples from Central Africa, in which we combined Azobé and Tali in one model to check their combined accuracy. The model that assigned the eight clusters yielded a correct assignment rate of 88.0% (model A4), which was comparable to the individual models (model A1 and A2). Interestingly, the variable timber species was not included in the most important distinguishing variables of the combined models (model A4 and A8). This again indicates that for certain elements, which are apparently distinguishing, the differences between sites were larger than the species differences.

3.3. Soil chemistry and wood elemental concentrations across the study area

In the RF assignment models, both essential and non-essential elements as well as some ratios were most important for site classifications (figure 3). There was no overlap in the set of key elements in Central Africa and Borneo. All multiple regression models associated wood element concentrations with soil variables but model fit was low overall (av.adj. R^2 between 0.11 and 0.47, figure 4). The low model fit likely results from the variation of wood chemical composition within a site, which is not covered at the same resolution in the gridded soil data. As the soil data originates from interpolated maps, small-scale variation is not well represented in the explanatory variables. Therefore the models mostly describe the large-scale patterns across the study areas.

3.4. Blind sample verification and assignment

Regarding the first question of verification, 70% of the blind samples was included in the correct cluster. Only five samples were included in the incorrect cluster (false positives). From these five, four samples were from a different species (figure 5(A) and table S2) and the fifth was a Tali sample that was included in the correct cluster but also in a neighbouring cluster, both in Cameroon. However, the incorrect exclusion error of the verification analysis was higher (false negatives): 9 of the 36 blind samples originating within the reference clusters were excluded from their own cluster. The cluster in south-west Cameroon stood out as four out of the five samples (80%) were incorrectly excluded from that cluster, whereas in the other clusters a maximum of 30% (3 out of 10) of the blind samples was incorrectly excluded, with an average of 14% of incorrect exclusions per cluster.

As for the second question, blind sample assignment, the first step of exclusion went well for the blind samples that originated from within the sites

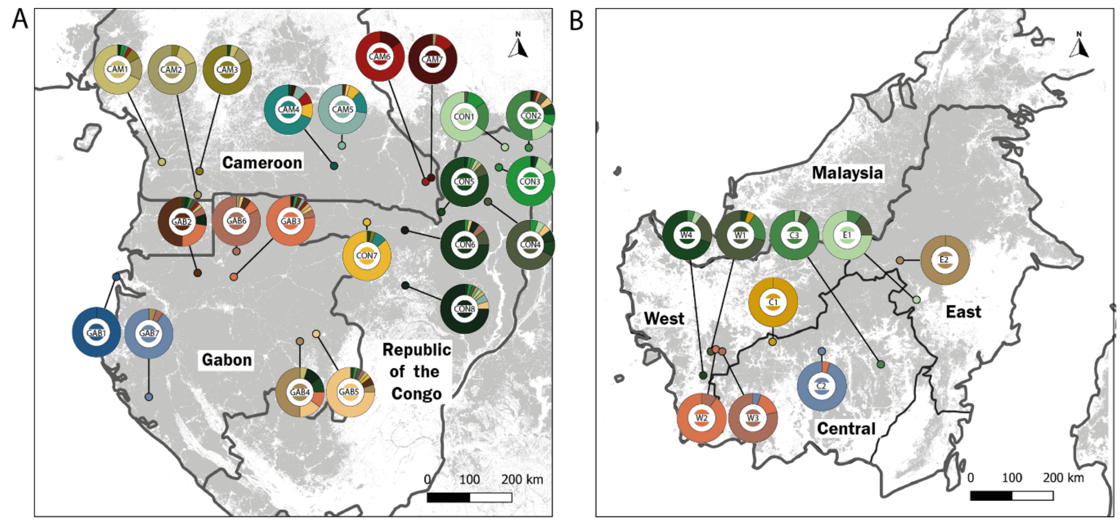

Figure 2. Confusion charts of the random forest assignment models for (A) the 22 study sites in Central Africa, model A8 in table 1, and (B) the 9 study sites on Borneo, model A7. Colours in the inner circle indicate the colour of the unique site. Colours in the outer circle indicate to which site the trees of that location were allocated, clusters from the cluster analysis are grouped and indicated in the different colour shades. No cluster-level verification was performed on Borneo because of the lack of geographical correspondence in the clusters. Primary tropical forest extent from Global Forest Watch in light grey (Turubanova *et al* 2018).

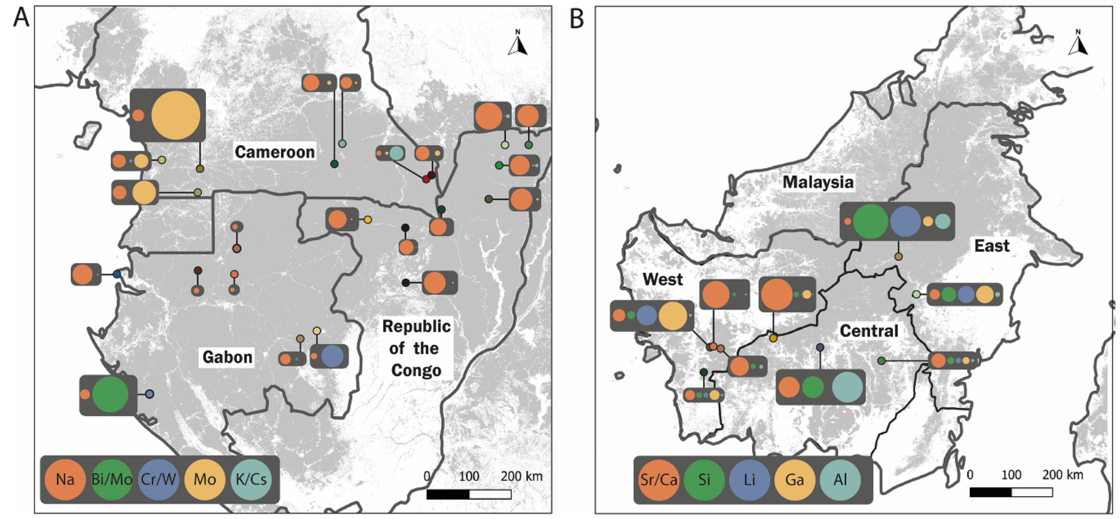

Figure 3. Site mean relative wood elemental concentrations of the five most important variables in the random forest models A4 (A, Central Africa) and A3 (B, Borneo) respectively. Site mean concentrations are scaled within each element, therefore the size of the circles indicates relative abundance per element at each site. Primary tropical forest extent from Global Forest Watch in light grey (Turubanova *et al* 2018).

with reference samples (category 1; correct exclusion of 90% for Azobé and 77% for Tali; Table S2). However, it did not exclude samples from outside the study area or from different species well, seven out of ten were not excluded (category 3). Furthermore, about half of the samples (6/13) that were collected close to the reference sites (<20 km; category 2) were excluded based on their chemical composition even though they should have been included. After exclusion, assignment to cluster level was of similar success as the RF self-assignment: 72% of the blind samples was correctly assigned to its cluster of origin (figure 5(B) and table 1). Similar to the verification analysis, most incorrect assignments were from the cluster in south-west Cameroon: none of the five samples was correctly assigned. Furthermore, assignment success at the site level was considerably lower compared to that at cluster level: only 30% of the blind samples was assigned to the correct site of origin.

Table 1. Success rates of the random forest cluster verification (model V1–V8), cluster assignment (model A1–A4) and site assignment models (model A5–A8). The verification models included both Azobé and Tali reference data and are presented per country (Rep o/t Congo as Republic of the Congo). The assignment models included reference data of each species individually and a combined model with both Azobé and Tali. Clusters are depicted in figure 2. The number of clusters, sites and trees indicate their respective totals as included in the models.

1. Verification models

Country		Cluster-level success		# sites	# trees
Cameroon	(V1)	Khaki	97.8%	3	120
	(V2)	Lightblue	96.4%	2	79
	(V3)	Red	98.1%	2	79
Gabon	(V4)	Darkblue	97.6%	2	40
	(V5)	Orange	94.4%	3	120
	(V6)	Lightbrown	94.8%	2	60
Rep o/t Congo	(V7)	Green	91.6%	7	280
	(V8)	Yellow	97.5%	1	36
Average			96.0%		

2. Assignment models

		Cluster-level success	# clusters		Site-level success	# sites	# trees
Azobé	(A1)	85.7%	8	(A5)	80.1%	21	418
Tali	(A2)	87.2%	7	(A6)	68.8%	20	394
Red Meranti	(A3)	92.6%	5	(A7)	87.6%	9	179
Azobé and Tali	(A4)	88.0%	8	(A8)	74.4%	22	812

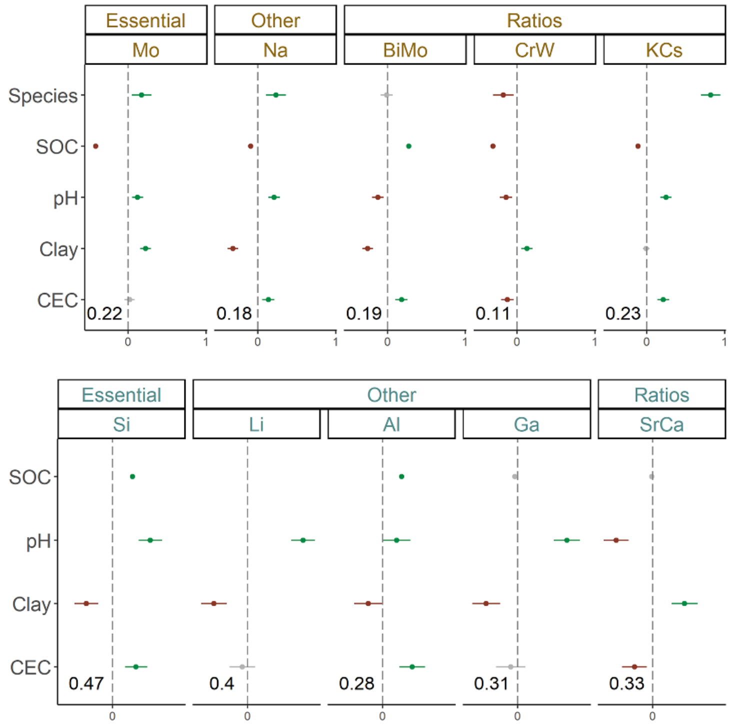

Figure 4. Multiple regression models of the five most important elements in the random forest assignment models in (A) Central Africa, model A8, and (B) Borneo, model A7. Element concentration was tested as a function of: soil organic carbon content (g kg^{-1}), pH, clay content (g kg^{-1}) and cation exchange capacity (mmol(c) kg^{-1}; all from Poggio *et al* 2021), plus timber species (in Central Africa only). Coefficient plots include slope estimates (mean and 95 CI), averaged across top-ranked models. The CI of the red (negative) and green (positive) circles does not overlap with zero and therefore denotes a strong effect. A positive species effect indicates a higher value in Tali compared to Azobé. The grey circles were included in the top models but their CI overlaps with zero, therefore they do not indicate a strong effect on the response variable. Numbers in the panels represent av. adj. R^2 of the best models.

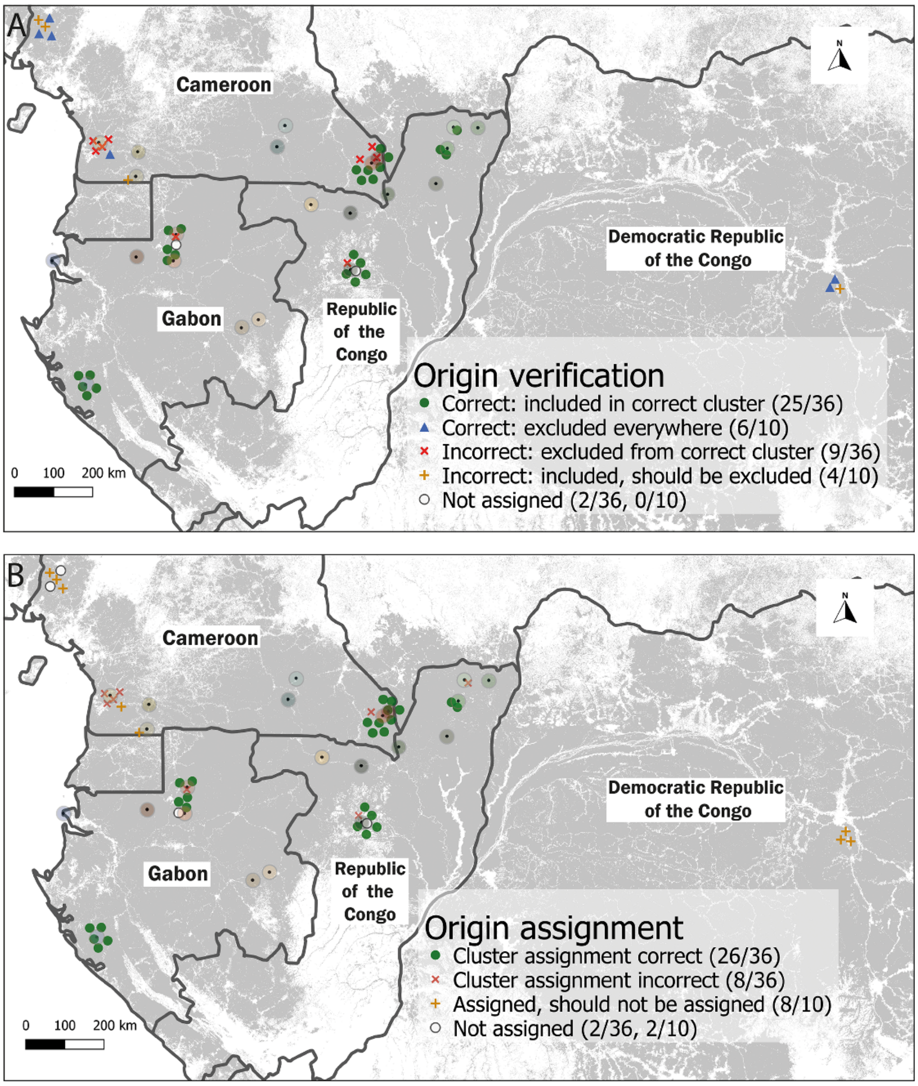

Figure 5. Blind sample origin and results of (A) the sample origin verification (*does this sample come from cluster X?*) and (B) origin assignment (*where does this sample come from?*), both at the cluster level. To avoid overlap, blind sample symbols were scattered around their site of origin and this does not reflect distance to that site. Origin verification was based on the eight RF models per cluster (models V1–V8, table 1). Origin assignment was based on RF model A4. The total number of samples that was correct or incorrect is indicated in the legend, compared to the total per group: 36 samples originated from <20 km of a site (category 1 + 2), 10 blind samples originated from >20 km of a site or were another species (category 3). For percentages per species and blind sample category see table S2. Primary tropical forest extent from Global Forest Watch in light grey (Turubanova *et al* 2018).

4. Discussion

4.1. What drives variation in timber multi-elemental composition?

This study provides insights into the drivers of wood multi-elemental composition for multiple species and continents. As expected, co-occurring timber species differed in their chemical composition (figures 1 and S1, Amais *et al* 2021). Nevertheless, they also shared a spatial pattern for many essential and non-essential elements (figure 1). This overlapping spatial variation was reflected in the combined RF assignment models: their assignment success was similar to the individual models (table 1) and timber species was not amongst the most important variables.

In tracing studies on other commodities, elemental composition was mostly associated with soil type and soil chemistry (Baroni *et al* 2015, Wang *et al* 2020), but studies investigating this are limited. In wood, the important variables for assignment in the RF models were associated with soil physical and chemical properties (figure 4). The direction of the effects mostly reflected the buffer capacity of the soils: a positive effect of pH and cation exchange capacity on the wood elemental concentrations reflects that chemical concentrations in wood are higher on soils with a high nutrient retention capacity (figure 4). These results underline that soil chemistry is an important driver of wood elemental composition (Boeschoten *et al* 2022).

4.2. What is the performance of multi-elemental analysis for timber tracing?

The RF classification models confirmed that there is potential for multi-elemental tracing of timber (figure 2). In Central Africa, distinct chemical patterns were found in two timber species (Azobé and Tali) within the three most important timber exporting countries. The scale at which tracing performed best was not the site level (assignment accuracy of 74.4%) but the cluster level (assignment accuracy of 88%), which were between 50 and 800 km apart (figures 2 and S1). On Borneo, site-level assignment accuracy was already high (87.6%), with distinct chemical patterns across provinces. Even though accuracy increased at the cluster level on Borneo as well (92.6%), these clusters overlapped geographically and are thus unlikely to improve tracing results. This difference between Central Africa and Borneo could arise from the spatial scales at which soil characteristics differ, which is especially clear when comparing the soil types in our study areas. Whereas our study area in Central Africa almost exclusively consists of ferralsol, the area on Borneo consists largely of nitisol which alternates with cambisol and ferralsol even at 10 km (Poggio *et al* 2021).

To put elemental tracing in perspective, we compare its performance to that of other methods, such as genetics and stable isotope ratios. Genetic studies at the population level reported self-assignment success rates between 55% (Ng *et al* 2017) and 88% (Chaves *et al* 2018), which is comparable to the site level accuracy of 68.0%–87.6% presented here (table 1). For a Red Meranti species, genetic site-level assignment accuracy in Western Malaysia was lower than our chemical site-level assignment accuracy of Red Meranti on Indonesian Borneo (60.60% vs 87.6%, Chin Hong *et al* 2017). Overall, genetic studies that were applied at the same scales reported similar success rates to this study (Low *et al*, 2022). Compared to the genetic literature, much less work has been done on other chemical methods for tracing purposes. Site-level assignment success based on stable isotope ratios varies between studies, mostly illustrating that the method has more potential at large geographical scales such as the state or country level (Vlam *et al* 2018, Watkinson *et al* 2020, Paredes-Villanueva *et al* 2022). Therefore our results suggest a higher potential for elemental tracing compared to stable isotope ratios at finer spatial scales.

In addition to performance of the RF classification models, the blind sample tests in Central Africa provided a proof of principle for multi-elemental tracing. Verification accuracy was 72% (does this sample come from cluster X?) and assignment accuracy was 70% (where does this sample come from?; table S2). Yet, the blind sample exclusion showed that methods to identify samples from outside the area or from species not covered by the reference database had a low success rate (63%, table S2). Additionally, the 25% of false negatives in the blind sample verification analysis (exclusion of timber that should be included; table S2) are especially sensitive as they can lead to erroneous accusations of timber fraud. Better exclusion criteria are needed, which can be achieved by increased sampling effort but also by combining it with other methods such as wood anatomy for species identification.

Blind sample test results based on other methods are limited, but those available showed a comparable accuracy. In Central Africa, higher successes were reported for genetic tracing of Tali (92%, Vlam *et al* 2018) as well as when assigning batches of samples of *E. cylindricum* (86%, Jolivet and Degen 2012). Assignment of those same Tali blind samples based on stable isotope ratios was unsuccessful however (Vlam *et al* 2018). Other studies verifying blind sample origin based on stable isotope ratios had mixed results. The predicted area of origin of oak samples in the US encompassed the true origin for 78% of the blind samples (Watkinson *et al* 2020). Furthermore, two investigations on tropical timbers reported a country-level verification accuracy of 87% (Förstel *et al*, 2011) and 50%–70% (Degen *et al* 2015), and 20%–50% (Degen *et al* 2015) at the region-level. Based on those comparisons, multi-elemental tracing of timber is a novel method that seems to achieve sub-country level origin identification at an accuracy that ranges between isotopic and genetic analyses.

4.3. The next steps for timber tracing with multi-element analysis

Based on these results we identified key points to be addressed in order to make multi-element analysis operational. Origin verification analysis must be improved, specifically the exclusion of samples from a specific origin. This can be achieved by increased sampling within clusters, to better cover the spread in chemical composition per cluster, as well as by sampling additional trees from different species and origins, to improve cluster specificity. New opportunities could also emerge with the addition of more species to the reference dataset. With site differences larger than differences between species (figure 1), data collection efforts could be reduced by combining the information of multiple species into a single reference dataset. This is a great advantage over other tracing methods that require a species-specific reference set, such as genetic tracing methods, but should be tested with more species.

Lastly, the statistical analyses can be developed further to enable large-scale application of the method. Adding samples to the reference database in a more scattered sampling design allows for interpolating spatial analyses, similar to isoscapes in stable isotope tracing (Bowen 2010). That would open up opportunities to move forward from site- or cluster-oriented verification of a blind sample towards identifying a potential region of origin.

5. Conclusions

This study illustrates a high potential for multi-element analysis to be developed into a timber tracing tool, because (1) wood origin could be identified at a sub-country scale in Central Africa and on Borneo (from 50–800 km), (2) overall cluster verification success is high, even though it varied per cluster (96.0% in the RF self-assignment and 70% for the blind samples) and (3) the spatial variation in Central Africa included non-species specific patterns. This is a first indication that it may be possible to combine multiple species in one reference dataset from which to build classification models. Furthermore, wood chemical composition was associated with soil chemistry, which can help to identify regions where this method may be most promising. However, the methods to exclude samples originating from outside of the reference dataset must be improved to apply elemental tracing in a forensic context. These findings should be confirmed for other sites and species in order to implement multi-element analysis as a timber tracing tool.

Data availability statement

The data cannot be made publicly available upon publication due to legal restrictions preventing unrestricted public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This study was supported by the Dutch Research Council (NWO-TTW-OTP-16427). Additional fieldwork support was received from the Alberta Mennega Foundation and from WorldForestID. We thank all the collaborating timber companies and their field teams for facilitating the fieldwork, colleagues at our partner institutes IPB University, University of Dschang, Marien Ngouabi University, IRET/CENAREST and the National Herbarium of Gabon for help in arranging fieldwork, our colleagues of WFSR for assistance in the laboratory work and our colleagues from the Forest Ecology and Forest Management writing group for feedback on the manuscript. We also thank the Indonesian Ministry of Environment and Forestry for the permission to access the plant sample genetic resources for research purposes with the permit number SK Dirjen: SK.92/KSDAE/SET.3/KSA.2/4/2022.

Conflict of interest

Authors declare no conflict of interest.

ORCID iDs

Laura E Boeschoten https://orcid.org/0000-0002-6061-2194
Mart Vlam https://orcid.org/0000-0002-0038-6586
Ute Sass-Klaassen https://orcid.org/0000-0002-8479-3209
Barbara Rocha Venâncio Meyer-Sand https://orcid.org/0000-0003-1291-8145
Gaël D U Bouka https://orcid.org/0000-0001-8850-522X
Rita M D Ndangani https://orcid.org/0000-0002-3065-6655
Iskandar Z Siregar https://orcid.org/0000-0002-5419-482X
Steve N Tassiamba https://orcid.org/0000-0003-4899-4537
Pieter A Zuidema https://orcid.org/0000-0001-8100-1168

References

Amais R S, Moreau P S, Francischini D S, Magnusson R, Locosselli G M, Godoy-Veiga M, Ceccantini G, Ortega Rodriguez D R, Tomazello-Filho M and Arruda M A Z 2021 Trace elements distribution in tropical tree rings through high-resolution imaging using LA-ICP-MS analysis *J. Trace Elem. Med. Biol.* **68** 126872

Baroni M V, Podio N S, Badini R G, Inga M, Ostera H A, Cagnoni M, Gautier E A, García P P, Hoogewerff J and Wunderlin D A 2015 Linking soil, water, and honey composition to assess the geographical origin of Argentinean honey by multielemental and isotopic analyses *J. Agric. Food Chem.* **63** 4638–45

Barton K 2020 MuMIn: multi-model inference *R package version 1.43.17*

Bates D, Mächler M, Bolker B and Walker S 2015 Fitting linear mixed-effects models using lme4. R package version 1.1–27.1 *J. Stat. Softw.* **67**

Boeschoten L E *et al* 2022 Clay and soil organic matter drive wood multi-elemental composition of a tropical tree species: implications for timber tracing *Sci. Total Environ.* **849** 157877

Bowen G J 2010 Isoscapes: spatial pattern in isotopic biogeochemistry (https://doi.org/10.1146/annurev-earth-040809-152429)

Branco P, Ribeiro R P and Torgo L 2016 *UBL: An R Package for Utility-Based Learning*

Burnham K P and Anderson D R 2002 *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* 2nd edn (New York: Springer)

Chaves C L, Degen B, Pakull B, Mader M, Honorio E, Ruas P, Tysklind N and Sebbenn A M 2018 Assessing the ability of chloroplast and nuclear DNA gene markers to verify the geographic origin of Jatoba (Hymenaea courbaril L.) timber *J. Hered.* **109** 543–52

Degen B, Bouda H-N and Blanc-Jolivet C 2015 *ITTO project PD 620/11 Rev.1 (M): development and implementation of a species identification and timber tracking system with DNA fingerprints and isotopes in Africa Tech. Rep*

Deklerck V, Lancaster C A, Van Acker J, Espinoza E O, Van den Bulcke J and Beeckman H 2020 Chemical fingerprinting of wood sampled along a pith-to-bark gradient for individual comparison and provenance identification *Forests* **11** 1–13

Deklerck 2022 Timber origin verification using mass spectrometry: challenges, opportunities, and way forward *Forensic Sci. Int. Animals Environ.* **3** 100057

Dormontt E E *et al* 2015 Forensic timber identification: it's time to integrate disciplines to combat illegal logging *Biol. Conserv.* **191** 790–8

Förstel H *et al* 2011 *Final project report WWF Germany—fighting illegal logging through the introduction of a combination of the isotope method for identifying the origins of timber and DNA analysis for differentiation of tree species project Tech. Rep* (Berlin: WWF Germany)

Gasson P E, Lancaster C A, Young R, Redstone S, Miles-Bunch I A, Rees G, Guillery R P, Parker-Forney M and Lebow E T 2021 WorldForestID: addressing the need for standardized wood reference collections to support authentication analysis technologies; a way forward for checking the origin and identity of traded timber *Plants People Planet* **3** 130–41

Gonzalvez A, Armenta S and de la Guardia M 2009 Trace-element composition and stable-isotope ratio for discrimination of foods with protected designation of origin *TRAC Trends Anal. Chem.* **28** 1295–311

Hevia A, Sánchez-Salguero R, Camarero J J, Buras A, Sangüesa-Barreda G, Galván J D and Gutiérrez E 2017 Towards a better understanding of long-term wood-chemistry variations in old-growth forests: a case study on ancient Pinus uncinata trees from the Pyrenees *Sci. Total Environ.* **625** 220–32

Hietz P, Horsky M, Prohaska T, Lang I and Grabner M 2015 High-resolution densitometry and elemental analysis of tropical wood *Trees* **29** 487–97

Hoare A 2015 *Tackling Illegal Logging and the Related Trade: What Progress and Where Next?*

Hoare A and Kanashiro Uehara T H 2022 *Establishing fair and sustainable forest economies: lessons learned from tackling illegal logging Tech. Rep.* (London: Royal Institute of International Affairs)

Joebstl D, Bandoniene D, Meisel T and Chatzistathis S 2010 Identification of the geographical origin of pumpkin seed oil by the use of rare earth elements and discriminant analysis *Food Chem.* **123** 1303–9

Jolivet C and 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. Genet.* **6** 487–93

Klaassen R K W M (ed) 2018 *Houtvademecum* 11th edn (Almere: Centrum Hout)

Lee S, Park B S, Lee D, Chung H and Lee K-S 2015 Spatial variability in hydrogen and oxygen isotopic composition of Korean Red Pine and its implication for tracing wood origin *Environ. Earth Sci.* **73** 8045–52

Liaw A and Wiener M 2002 randomForest: classification and regression by randomForest *R package version 4.6–14*

Low M C *et al* 2022 Tracing the world's timber: the status of scientific verification technologies for species and origin identification *IAWA J.* 1–22

Lowe A J *et al* 2016 Opportunities for improved transparency in the timber trade through scientific verification *BioScience* **66** 990–8

Ma G, Zhang Y, Zhang J, Wang G, Chen L, Zhang M, Liu T, Liu X and Lu C 2016 Determining the geographical origin of Chinese green tea by linear discriminant analysis of trace metals and rare earth elements: taking Dongting Biluochun as an example *Food Control* **59** 714–20

Maechler M *et al* 2021 *cluster: cluster analysis basics and extensions R package version 2.1.2*

McFerrin L 2013 HDMD: structural analysis tools for high dimensional molecular data *R Package version 1.2*

Ng Chin Hong, Lee S L, Tnah L H, Ng K K S, Lee C T, Diway B and Khoo E 2017 Geographic origin and individual assignment of Shorea platyclados (Dipterocarpaceae) for forensic identification *PLoS One* **12** e0176158

Oksanen J *et al* 2020 vegan: community ecology package *R Package Version 2.5–7*

Paredes-Villanueva K, Boom A, Ottenburghs J, Van Der Sleen P, Manzanedo R D, Bongers F and Zuidema P A 2022 Isotopic characterization of Cedrela to verify species and regional provenance of Bolivian timber *Tree-Ring Res.* **78** 73–89

Poggio L, de Sousa L M, Batjes N H, Heuvelink G B M, Kempen B, Ribeiro E and Rossiter D 2021 SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty *Soil* **7** 217–40

R Core Team 2021 R: A Language and Environment for Statistical Computing

Richter B, Gurk S, Wagner D, Bockmayr M and Fischer M 2019 Food authentication: multi-elemental analysis of white asparagus for provenance discrimination *Food Chem.* **286** 475–82

Scharnweber T, Hevia A, Buras A, van der Maaten E and Wilmking M 2016 Common trends in elements? Within- and between-tree variations of wood-chemistry measured by x-ray fluorescence—a dendrochemical study *Sci. Total Environ.* **566–567** 1245–53

Turubanova S, Potapov P V, Tyukavina A and Hansen M C 2018 Ongoing primary forest loss in Brazil, Democratic Republic of the Congo, and Indonesia *Environ. Res. Lett.* **13** 074028

Vlam M, de Groot G A, Boom A, Copini P, Laros I, Veldhuijzen K, Zakamdi D and Zuidema P A 2018 Developing forensic tools for an African timber: regional origin is revealed by genetic characteristics, but not by isotopic signature *Biol. Conserv.* **220** 262–71

Wang Z, Erasmus S W, Dekker P, Guo B, Stoorvogel J J and van Ruth S M 2020 Linking growing conditions to stable isotope ratios and elemental compositions of Costa Rican bananas (Musa spp.) *Food Res. Int.* **129** 108882

Watkinson C J, Gasson P, Rees G and Boner M 2020 The development and use of isoscapes to determine the geographical origin of Quercus spp. in the United States *Forests* **11** 862