**Appendix A: Literature search methods**

Databases

The following databases were used to survey the scientific literature: Web of Science (WoS) All Databases (Web of Science Core Collection, BIOSIS Citation Index, CABI: CAB Abstracts, Current Contents Connect; KCI-Korean Journal Database; MEDLINE).

Search string

The search string has three sections, relating to biota, analyses and methods. Each section comprises a combination of ‘Topic’ and ‘Title’ fields separated by the search operators AND, OR and NOT. In WoS when the Topic field is used the following fields within a record are searched for the search terms: Title, Abstract, Author Keywords, Keywords Plus®. When the Title field is used only the title of the journal article, proceedings paper, book or book chapter are searched for the search terms.

*Section 1*

The first section addresses the target biota and is relevant to both genetic and chemical identification methods. The terms timber, tree, wood or forest appeared to be the most widely used terms in the targeted literature*.* The Title field was adopted in the first query as initial searches indicated that using the Topic field included many irrelevant records in the results. Initial searches for genetics indicated that there were papers in the WoS database that did not have the terms timber, tree, wood or forest in the Title field but were relevant. To capture these papers in the WoS results a query was included in section 1 to include papers with the following key genetic terms in the Title field (e.g., SNP, SSR, microsatellite, marker, genetic diversity, population structure, genetic structure, DNA or assignment).

We searched using the genera of the target species. To exclude papers investigating other biota from the WoS results, a query was included to exclude any papers containing the following search terms in the Topic field: bacteria, fungi, microbial, soil, bird, insect, frog, rabbit, mouse or bear.

*Section 2*

The second section included search terms relating to the analyses we were targeting. In the case of genetics, a query was included to find papers containing the following search terms in the Topic field: timber tracking, provenance, geographical origin, forensic, source identification, population detection, origin detection, species identification, species differentiation and illegal logging. A query was included in this section to exclude any papers from the WoS results containing the following search terms in the title field: biology, invasive, cultivate, sex, growth, pollen, breed and extract.

*Section 3*

The third section focused on search terms relevant to the identification method (e.g., genetic, or chemical). For genetics a query was included to find papers containing the search terms SNP, SSR and microsatellites in the Topic field as these can highlight genetic differences between geographical regions (e.g., populations of a species, logging concessions, individuals within a population) (Degen & Fladung, 2008; Lowe & Cross, 2011). A query was included to find and exclude any papers with the following search terms in the Topic field referring to other types of markers used in population genetic and phylogeographic studies based on them not working as reliably as SSRs and SNPs in timber: RAPD, allozyme, AFLP, isozyme, matK, ITS, trn and rbcl.

**Table A1.** Genetic and chemical search strings

|  |  |  |  |
| --- | --- | --- | --- |
|  | Section | Field tag | Search string components (separated by AND) |
| All identification methods | Biota | Title (TI) | TI=(timber OR tree\* OR wood OR forest) |
| Topic (TS) | NOT TS=(\*bacteri\* OR fung\* OR microb\* OR soil\* OR bird OR insect OR frog\* OR rabbit\* OR mouse OR bear\* |
| Topic (TS) | TS=(Abies OR Acacia OR Acer OR Afzelia OR Agathis OR Amburana OR Aniba OR Aningeria OR Anisoptera OR Anthocephalus OR Antiaris OR Apuleia OR Aquilaria OR Araucaria OR Aspidosperma OR Aucoumea OR Bagassa OR Baillonella OR Bertholletia OR Betula OR Bowdichia OR Brosimum OR Bulnesia OR Caesalpinia OR Calophyllum OR Calycophyllum OR Canarium OR Capirona OR Carapa OR Cariniana OR Caryocar OR Castanea OR Cedrela OR Cedrelinga OR Ceiba OR Celtis OR Chorisia OR Chukrana OR Copaifera OR Couratari OR Cupressus OR Cylicodiscus OR Dalbergia OR Dicorynia OR Dinizia OR Diospyros OR Dipterocarpus OR Dipteryx OR Dryobalanops OR Endospermum OR Entandrophragma OR Eperua OR Erythrophleum OR Eucalyptus OR Eusideroxylon OR Fagus OR Ficus OR Fitzroya OR Fraxinus OR Gonystylus OR Goupia OR Guaiacum OR Guarea OR Guibourtia OR Gyrinops OR Handroanthus OR Hopea OR Hura OR Hymenaea OR Hymenolobium OR Intsia OR Jacaranda OR Juglans OR Khaya OR Koompassia OR Lagerstroemia OR Larix OR Lophira OR Machaerium OR Manilkara OR Microberlinia OR Micropholis OR Milicia OR Millettia OR Minquartia OR Myroxylon OR Nauclea OR Ocotea OR Ormosia OR Pachira OR Palaquium OR Paraberlinia OR Parashorea OR Peltogyne OR Pericopsis OR Picea OR Pilgerodendron OR Pinus OR Piptadeniastrum OR Platymiscium OR Podocarpus OR Podophyllum OR Pometia OR Populus OR Pterocarpus OR Pterygota OR Qualea OR Quercus OR Roupala OR Ruizterania OR Santalum OR Sextonia OR Shorea OR Simarouba OR Sorbus OR Swartzia OR Swietenia OR Swintonia OR Tabebuia OR Taxus OR Tectona OR Terminalia OR Tetraberlinia OR Tetracentron OR Tetramerista OR Tipuana OR Toona OR Triplochiton OR Virola OR Vitex OR Vouacapoua OR Widdringtonia OR Xylia) |
| Analysis | Topic (TS) | TS=("timber trac\*" OR provenanc\* OR "geographic\* origin" OR forensic OR "source identification" OR population detect\* OR origin detect\* OR species identification OR species differentiat\* OR illegal logging) |
| Exclusions | Title (TI) | NOT TI=(biolog\* OR invasi\* OR cultiva\* OR \*sex\* OR growth OR pollen OR breed\* OR extract\*) |
| Identification method specific | Genetics | Topic (TS) | TS=(polymorph\* OR genetic diversity OR “population structure” OR “genetic structure” OR SNP\* OR SSR\* or microsatellite OR marker\* OR genetic assignment OR phylogeograph\* OR DNA) |
| Title (TI) | TI = SNP\* OR SSR\* OR microsatellite\* OR marker\* OR “genetic diversity” OR “population structure” OR “genetic structure” OR DNA OR assignment) |
| Topic (TS) | NOT TS=(RAPD\* OR allozyme\* OR AFLP OR isozyme\* OR matk OR ITS\* OR trn\* OR rbcl) |
| DART TOFMS | Topic (TS) | TS= species differentiation OR species identification OR species classification OR source identification OR timber identification |
| Title (TI) | TI= species differentiation OR species identification OR source identification OR species classification OR chemical fingerprinting OR metabolome profiling OR DART TOFMS OR mass spectrometry OR chemical differentiation |
| Title (TI) | NOT TS= metabolomics |
| Stable isotopes | Topic (TS) | TS=Isotop\* OR stable isotop\* OR isotop\* ratio OR isotop\* variation OR isoscap\* OR isotop\* analysis OR isotop\* signature OR ICP-MS OR IRMS |

Time period

In the iterative process of developing the search string, initial search results indicated that the genetic markers we are focusing on such as SSRs only appear in the literature from 1994 onwards so a custom date range of 1994 to 2021 was applied to the search.

Exclusion criteria

Literature searches were carried out for genetics, DART TOFMS and stable isotopes in between June and September 2021. We read titles and abstracts of all studies and excluded those that were not relevant to the review, excluding studies that did not contain target species, used methods outside the scope of the review, assessed species but no geographical provenance identification, assessed microbiology, plant physiology, silviculture, climate change, parentage and mating systems, clonal species and forest restoration or were not peer reviewed. For genetics we excluded studies that did not report population differentiation statistics.

Data collection

For each paper included in the review, the following is an example of what data was collected: reference information (e.g., journal and authorship), species information (e.g., taxa, common names, geographic range, and GTTN geographic region), methodology (e.g., taxa or origin resolution, sample source, no. trees sampled, no. of populations or concessions sampled), clustering results, self-assignment and blind sample assignment percentages. For genetics the data collected included marker information (no. and type of markers used), genetic diversity statistics (e.g., no. of haplotypes, heterozygosity, inbreeding coefficient), genetic differentiation statistics (e.g., Fst. Gst and Dst statistics). For isotopes the isotopes measured, and isotopes used for origin determination were recorded. For mass spectrometry data included type of mass spectrometer, spectra included in the dataset and tissue type.

**References**

Degen B, Fladung M. 2008. Use of DNA-markers for tracing illegal logging, In: B. Degen, (Ed.), *Proceedings of the International Workshop Fingerprinting Methods for the Identification of Timber Origins*: 6−14.

Lowe AJ, Cross HB. 2011. The application of DNA methods to timber tracking and origin verification. *IAWA J*. 32: 251−262.