Development and application of genetic timber tracking tools to help control illegal logging

Duncan Iain Jardine
(Grad Dip Research Methods, B Env Sci)

A thesis submitted to the

School of Biological Sciences
The University of Adelaide
AUSTRALIA

In fulfilment of the requirements for the degree of Doctor of Philosophy

August, 2019

Intentional blank page

THESIS DEDICATION

This thesis is dedicated to my daughters *Celia "Bug" Ivy Jardine* and *Eva "Bear" Rose Jardine*. Upon seeing this image of you, Celia, holding your toy planet and crying, I knew there is no question of whom this thesis research is for. Because you two, and the other children of your generation, are the ones that we most need to save the planet. I hope that my work will be able to be used to help fight the scourge of illegal logging and to help facilitate a more sustainable trade in timber species. I know that you will grow up to have a good solid understanding of what is fair and true, and I hope that this work will be able to help you.

Intentional blank page

TABLE OF CONTENTS

THESIS ABSTRACT

Illegal logging of timber is a significant problem worldwide. Apart from being a major driver of global deforestation, there are also severe economic implications. To abet the illegal trade, timber is often laundered into otherwise legitimate systems, tainting the supply chain, and sold on to unwitting consumers. Attempts to control and monitor the trade of illegal goods have been made by some of the major timber importing nations, with laws enacted that require importers to provide evidence of the species of timber and where it comes from. There are several scientific methods suitable for ascertaining this information from the timber itself; furthermore, these approaches can be used retrospectively by government agencies to assess the veracity of the declared claims.

This thesis focuses on the application of genetic methods for the monitoring and control of traded timber. Genetic methods have been used for identification purposes in many other species, most notably humans. Furthermore, they were recently recognised as being suitable for timber species in UNODC guidelines on forensic timber identification. Yet difficulties associated with obtaining DNA from timber material are holding back the uptake of genetic methods in timber identification. Compared to fresh leaf tissue, extracting DNA of sufficient quality and quantity from timber is not routine. This is a critical limitation, as without the capacity to extract DNA from timber, genetic techniques cannot be used effectively for supporting legislation. Additionally, for most timber species, there are currently no species-specific genetic markers that can be used for origin verification or individualisation tests available. Even for species where markers do exist, there is generally a lack of research concerning how well these approaches perform and the best way they can be utilised within a legal context.

My thesis presents work that was conducted as part of a patent application for extracting DNA from timber and other degraded plant material. It then focuses on the development of species-specific genetic markers for two timber species; bigleaf maple (*Acer macrophyllum*) and ayous (*Triplochiton scleroxylon*). Ayous

markers were used to conduct a population genetic analysis to identify the genetic structuring of the species across its native range. The results from that analysis were then incorporated into an assignment testing study, which was used to assess the claimed origins of blind test samples. The blind testing results were used to determine the effectiveness of the genetic markers and analysis approach to verify traded timber of ayous.

The DNA extraction project identified optimum conditions for successful extraction using our developed method, and the outcome was then compared to commercial kits. The patented protocol performed as well as the commercial kits and was often successful where the kits failed. The marker development studies were successful, and variable markers were developed for both species.

For ayous, the population genetic analysis identified three genetic populations across its natural range in tropical Africa. In addition to supporting previously identified patterns of genetic structure in the region, this work has furthered our understanding of the region's biogeography by identifying a novel genetic boundary occurring between the northwest and central samples from the Democratic Republic of the Congo (DRC). Samples from central DRC formed an isolated genetic group, while samples from the northwest of the country were more closely related genetically to samples from Cameroon and the Republic of the Congo. The study identified that the genetic patterns of ayous have likely been driven by geneflow and impacted by the contraction and expansion of forest boundaries in recent geological history.

The subsequent assignment testing found that assigning to the three genetic groups was the most successful approach, with lower success when assignment was based on country of origin. However as long as appropriate assessments were made to determine the significance of assignments, then acceptable results for country level claims were also achieved. Unfortunately, in ayous there is insufficient support in the data to perform origin assessments at the geographical population level.

In conclusion, this thesis demonstrates the capacity for using genetic methods to identify timber material. Furthermore, these methods are highly suitable for supporting law enforcement efforts to control and monitor the timber trade. Techniques that are underpinned by a scientific framework, such as genetics, can provide the certainty needed to meet legal requirements. Appropriate application of these techniques can assist measures to reduce the incentive for illegal logging practices; and promote the trade of legitimate products. Measures such as these can contribute to diminishing the negative environmental and socioeconomic impacts of illegal logging.

THESIS DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Duncan Iain Jardine
Canberra, AUSTRALIA
August 2019

ACKNOWLEDGMENTS

Jillian: my wife, friend, confidant, and listener of my rantings. We got you through your PhD and now it's my turn. I can't wait for the next set of adventures that will come our way; I know it is going to be amazing. When I first put this section together we just had **Bug** in our lives, now we have **Bear** as well, and that makes even more of a reason to finish this thesis. I look forward to reading this page again in our retirement on our writer's retreat in Cairns and contemplating where we have been since this day.

To my supervisors: **Andy** for having the confidence to give me a go, showing me great support and leadership and providing insight into how to get ahead in life. **Eleanor,** supervisor on the ground and guide throughout this candidature (oh, and translator when the science gets too intense). I admire your persistence and focus for life and hope that one day I can be as driven as you. Your direct knowledge of raising a young family whilst completing a PhD has been incredibly insightful for me. **Jeremy**, your energy towards my research has been immeasurable. Your assistance with my work is greatly appreciated. I am proud that I was given the opportunity to work with you.

Lasse: my good mate from the Thünen Institute. I really enjoyed the time we spent together at the lab in Hamburg, and I especially enjoyed the afternoon boßlen tournaments around the institute. Good luck with finishing your own PhD.

To the Lowe Lab Group members past and present: **Matt**, I really enjoyed the Friday office beers, it was been an essential shit talking exercise of my PhD. **The timber tracking team** (past and present): **Joey, Bianca, Rainbo, Kor** not only to thank you for assisting me with my research, but it is great to see a group of people that all want to use science to make to make a difference for a better world. **Hugh**, thanks for the job all those years ago. If you had of told me then that coming to Adelaide would change my life, I would probably have thought you were mad. And to the **Lab group people in general**: All of you have been a great group of individuals, who can bounce around ideas and share a laugh, whilst also being able to get some science done.

To my other Co-authors: **Celine, Bernd & Birgit**, Thank you for your swift revisions and comments and your contributions to this thesis work.

To my family: **Mum and Dad**, you brought me up right, I think, your values and ideologies have shown me the way to strive for the best. I'm sorry I didn't apply myself enough at school, but regardless, your messages have led me here. Who knew that was going to happen? **CJ**, brother, buddy, boofhead. You are as integral to this story as anyone else, and deservingly so need a mention. I really enjoy seeing you and Issa when we can, and I know you do too. I miss our random late night arguments, and the reciprocal rectifying that inevitably occurs.

To anyone else I may have missed, **thank you**. Even if it is not written down you deserve a cheer.

Chapter 1: General Introduction

Part 1: Thesis context

The importance of forests

Forests are an important biome – they extend over approximately 31% of the earth's land area[1], produce most of the oxygen required for life on earth (Raven et al. 2005), and contain ~80% of all aboveground and ~40% of belowground carbon stores (Dixon et al. 1994). Forests also support the majority of life on earth, with tropical forests in particular containing 60% of global biodiversity (Bradshaw et al. 2009). However, it is not just environmental capacity that forests provide, as they also have important cultural and community roles (Lowe et al. 2016; Nellemann and INTERPOL 2012; Sasaki et al. 2016). For example, ~18% of the world's population use forest products as the primary means of resources for building shelter (FAO 2014). Fuel derived from wood/timber (known as woodfuel) accounts for 6% of global primary energy supplies (FAO 2014), and approximately 840 million people (12% of global population) collect woodfuel and charcoal for their own use (FAO 2014). Considering that woodfuel is being used to cook and sterilize water, it also plays an important role in health and food safety for many people (FAO 2014). Yet by far the biggest contribution of forests for most people around the world is the economic value of timber. Estimates of the global wood trade put the value somewhere between US$320-350 billion annually (ITTO 2015; Nellemann and INTERPOL 2012). Furthermore, when incorporating the economic value of the entire forestry sector, it is estimated to be worth US$640 billion (~0.9% GDP) in income (FAO 2014). But many people have much more real connection to forests and forest products in their daily life than the mere numbers that show up in governments GDP and other statistics.

Economy of illegal logging

Alarmingly the economic foundation of the global timber trade is being undermined by the impact of illegal logging activities (Lowe et al. 2016). It is estimated that illegal logging deprives developing nations of ~US$15 billion annually in lost revenue and

[1]https://data.worldbank.org/indicator/AG.LND.FRST.ZS

taxes (EIA 2012). Globally, an estimated 30-50% of all timber traded comes from illegal sources (Degen 2012; Degen et al. 2013; Nellemann and INTERPOL 2012). In fact, illegal logging activities account for the largest proportion of illicit wildlife crimes. Its value (US$30-100 billion annually) is only surpassed by the trade in illicit drugs (~US$200 billion annually) (Nellemann and INTERPOL 2012). Some of the worst affected regions for illegal logging activities are from major forested regions, including the Amazon, Central Africa, Russia and South East Asia, where 50-90% of exports are thought to come from illegal sources (Degen 2012; Degen et al. 2013; Tnah et al. 2009). The prevalence of illegal logging is also exacerbated by socioeconomic factors such as government corruption, lack of economic prosperity, political instability, and conflicts (Lowe et al. 2016; Nellemann and INTERPOL 2012).

To further complicate the issue, the reasons for, and rates of, illegal logging are largely species dependant (Nellemann and INTERPOL 2012). The majority of illegally harvested timber comes from species that are of high economic importance, such as mahogany (*Swietenia macrophyla*), and rosewood (*Dalbergia* sp.). These high value species have been overharvested, reducing populations remaining in the wild to the extent that many are now listed as endangered or vulnerable on the International Union for Conservation of Nature and Natural Resources (IUCN) red list of threatened species[2] (Degen et al. 2013; Nellemann et al. 2014). These species are being deliberately targeted, and measures to mask the true origin (and even species) and launder them into legitimate logging supply chains are being employed (Nellemann and INTERPOL 2012; Vlam et al. 2018). Over 20 different methods for falsifying timber products have been recorded: from simply altering the species name or concession (origin) on export documents, moving an entire shipment to a new location, and generating illegitimate paperwork for already harvested logs (Nellemann and INTERPOL 2012). This diverse range of laundering techniques affects the trade in legitimate timber products, the impacts of which extend all the way to the end consumer.

From an import perspective (i.e. the consumer), it is estimated that ~5-10% of all timber imports into Europe (the EU), and 10-66% into China, are from illegal sources

[2]www.iucnredlist.org

(EIA 2012; Degen 2012). To further complicate the issue, the majority of timber (both legal and illegal) imported into China, is in the form of raw logs or sawn timber, which is then processed into furniture, paper pulp and other goods, and re-exported to consumer countries such as the EU (EIA 2012; Nellemann and INTERPOL 2012). This creates additional avenues for illegal timber to be laundered and on-sold as legitimised products.

Illegal logging is a vast and insidious problem that affects the whole planet, and mechanisms to control and monitor the situation need to be implemented urgently before more species are impacted. While consumer awareness can go some way to controlling the issues (Nellemann and INTERPOL 2012), realistically the best control mechanisms come via the creation and strengthening of domestic laws and international agreements. These measures would have the most impact on illegal logging activities. Furthermore, and the processes (e.g. forensics science) in place to provide enforcement of these laws.

Legislative avenues for controlling illegal logging

There is a broad legal spectrum for protecting species and pursuing those responsible for illegal logging activities. For species that have been heavily targeted and whose abundance has diminished significantly, the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) is the most applicable avenue to protect them (Dormontt et al. 2015; Johnson et al. 2014). CITES protects over 400 tree species (UNODC 2016). Under CITES, a species can be nominated by a signatory nation (if naturally occurring in that country) with all other signatory nations expected to control the movement of the species through their own legal framework (Johnson et al. 2014). There are three categories in which a species can be listed, each with a different level of associated restrictions. For an explanation of the three CITES appendices, the control mechanisms implemented, and the types of species listed in each category see Table 1.

Table 1: Explanation of CITES categories

Appendix	Control mechanisms	Types of species listed
CITES-I	Total trade ban (unless exceptional circumstances)	Threatened with global extinction
CITES-II	Restriction to both quantity and origin	Vulnerable to overexploitation (future endangerment)
CITES-III	Trade ban from particular countries	Country specific impacts

NB: table adapted from Dormontt (2015) & Johnson (2014)

The trade control of CITES-I species is relatively straightforward, with border control authorities able to halt the movement of products from these species with positive identification. However, for CITES-II & III species, knowledge of its origin becomes more important to ascertain legality. This is also the legal focus that is used for more easily traded/non-CITES listed species. Multiple domestic legislative mechanisms, such as the Australian *Illegal Logging Prohibition Act* (2012), United States of America (USA) *Lacy Act* (amended 2008), the EU *Forest Law Enforcement, Governance and Trade* (FLEGT) Action Plan (2003) and Timber Regulations (EUTR) (2010), and the Canadian *Wild Animal and Plant Protection and Regulation of International and Interprovincial Trade Act* (1992), have been implemented and highlight that countries around the world are becoming serious about reducing the problem of illegal logging activities (Degen et al. 2013; Dormontt et al. 2015; Jolivet and Degen 2012; Lowe et al. 2016). Whilst the trade restrictions associated with CITES listed species are for the benefit of the nominating country, the suite of timber trading laws approach illegal logging from an import (consumer) focus. Considering that it is the international demand for timber that incentivises the majority of illegal logging activities, this approach seems appropriate (Lowe et al. 2016). These laws place the responsibility for demonstrating legality onto the import trader rather than forcing exporting nation states to guarantee the legality of products. That is not to say that exporting countries cannot take responsibility for facilitating the requirements of importer-based timber-trading legislation. For example Voluntary Partnership Agreements (VPAs), which are bilateral trade agreements between the EU and certain timber producing countries (Ghana, Cameroon, Republic of Congo, Liberia, Central African Republic, Indonesia and Viet Nam, with several more being negotiated (Lao PDR, Malaysia, Thailand, Côte d'Ivoire, Democratic Republic of Congo, Gabon, Guyana and Honduras)) (ITTO 2015; Degen 2012; Jolivet and Degen 2012).

Requirements of the law

The general requirements for any consumer timber trading laws are the same as for trade in CITES-II & III listed species. They all require a statement or certificate that declares the name (scientific and sometimes common) and the origin of the material (where it was felled) (Lowe et al. 2016). They also emphasise that the importer must undertake due diligence to minimise the risk of the timber being harvested illegally

(i.e. the correct declaration of species and source forestry location) (Jolivet and Degen 2012). It is the responsibility of all actors in the supply chain (except the end consumer) to ensure this legality (Lowe et al. 2016). Recently several EU timber trading companies were fined for having a lack of due diligence for their timber imports[3] [4].

Verifying claims

The simplest way of verifying that claims are legitimate is to mark the logs in a shipment with a unique identifier that matches to the importation certificate. Traditionally this has been done with a brand or code, but being a fairly rudimentary method, it is easily forged. To further increase the reliability of these identification tools, more advanced Radio Frequency Identification (RFID) tags and electronic barcodes have been incorporated (Dormontt et al. 2015). The unique identifier system can also be used directly by a trader to verify that a supply chain is intact and not being affected by laundering. This approach also forms the basis of several supply chain certification schemes (e.g. Forest Stewardship Council (FSC)[5] or Programme for the Endorsement of Forest Certification (PEFC)[6]), that aim to improve the trading system and act as guarantors for more sustainable products (Sasaki et al. 2016; Tnah et al. 2009). These schemes have been shown to have a positive influence on consumer purchasing decisions, with suggestion there is a willingness to pay a premium for certified products, not just in timber (Yamamoto et al. 2014), but also for other products (e.g. Rainforest Alliance Certified coffee or tea) (Finkeldey et al. 2007; Lowe et al. 2010). Nevertheless, these schemes are being underwritten by a system that has the potential to be forged and mismanaged, thus increasing the chances that consumers are unwittingly supporting the trade of illegal product even if they are purchasing certified stock (Jolivet and Degen 2012; Lowe and Cross 2011; Lowe et al. 2010). Furthermore, most timber certified under these schemes are from North America and the EU (Nellemann and INTERPOL 2012). Considering that the vast majority of illegal logging activities are occurring external to these locations, the impact of certified products in reducing illegal logging rates is limited (Migone and Howlett 2012).

[3]www.edie.net/news/7/Furniture-retailer-Lombok-fined-by-Government-for-illegal-logging/

[4] www.flegt.org/news/content/viewItem/swedish-court-rules-myanmar-timber-documentation-inadequate-for-eu-importers/15-11-2016/57

[5]https://au.fsc.org/en-au/buy-fsc-certified/timber

[6]www.pefc.org/certification-services/supply-chain

Assessing the validity of evidence

However, the lack of suitability of these certificates and chain of custody techniques within a legal framework is even more of a concern to the limited use of certification schemes. Because the legitimacy of these claims cannot be independently verified, there is little evidentiary value in these approaches (Migone and Howlett 2012). Rather than attempting to invent mechanisms that would allow for independent assessment, a suite of alternative identification techniques that utilise the inherent features of timber itself have been suggested (Dormontt et al. 2015). These scientific methods have been suggested as applicable for working within timber trading and protection legislation and have been recently incorporated in UNODC forensic timber identification guidelines (UNODC 2016). These new methods can be grouped into three broad categories (visual, chemical, genetic) and have been used successfully in timber samples (Dormontt et al. 2015; Nellemann and INTERPOL 2012). Each method can be applied to identifying/verifying one of four different timber attributes (taxonomic name (genus/species), origin, date (of felling), and match), which will be discussed further in the following sections. The majority of these techniques are diagnostic tools that provide forensic identification. These diagnostic tests require specialist analysis, and results take time to generate. As an alternative to these advanced approaches, some techniques can be used as screening tests with results being quick to obtain. Additionally, these tests can provide swift and reliable field identification. Yet, screening tests still require forensic identification with a diagnostic technique to ensure a reliable outcome. All of these scientifically backed identification methods can be used either proactively (to substantiate claims made on certificates) or retroactively (to verify the legitimacy of claims). For a summary of the techniques available for timber identification, their category and application see Table 2.

Table 2 Timber identification technologies: Listings of the various timber identification methodologies suitable for use in monitoring illegal logging, with information as to the application, and determination capacity for each technology.

		Application		Capacity to Determine				
Category	**Technique**	**Screening**	**Diagnostic**	**Genus**	**Species**	**Origin**	**Match**	**Date (of felling)**
Visual	*Wood anatomy*	Yes	Yes	Yes	Yes*	Yes*	No	No
	Machine vision	Yes	Yes	Yes	Yes*	No ^	No	No
	Dendrochronology	No	Yes	No	No	Yes*	Yes	Yes#
Chemical	*Mass spectrometry*	No	Yes	Yes	Yes	Yes	No	No
	Near Infrared spectrometry	Yes	Yes	Yes	Yes	Yes	No	No
	Stable isotopes	No	Yes	No	No	Yes	No	No
	Chemical detection tools (e.g. Sniffer dogs, chemical noses)	Yes	No	No	No	No	No	No
	Radiocarbon dating	No	Yes	No	No	No	No	Yes
Genetic	*DNA barcoding*	No	Yes	Yes	Yes	Yes*	No	No
	Population genetics & Phylogeography	No	Yes	No	Yes*	Yes	No	No
	DNA profiling	No	Yes	No	No	No	Yes	No

*NB: Chemical detection tools includes detector dogs and chemical noses, * = occasionally (depends on species/group), # = only where growth rings present, ^ = unknown. Table annotated from Dormontt (2015) & Lowe (2011)*

Name testing

For most laws, the "name" required is the common (and/or trade) and scientific (genus and species) (Lowe et al. 2016). Common names are not a useful identification measure. For example, a quick search of a wood identification database[7] revealed that there are 14 traded species with a common name mahogany in Latin and Central America (only two species are true mahogany (*Sweitenia* sp.), both of which are CITES-II listed), and 18 species are commonly known as teak (genuine teak, (*Tectona grandis*) is native to SE Asia, but is grown as a forestry species around the world, including Africa, where the CITES-II listed African teak (*Pericopsis elata*) is found). Scientific names are a far more reliable system to define a species, and better suited for testing to determine legitimacy. Additionally, they are the only accepted naming system within both the ICUN and CITES frameworks. Considering that an evolutionary framework underpins scientific names, tools that can utilise evolutionary or inherited characters are best suited for this type of assessment.

Techniques that have been used to answer questions regarding the name of a sample include wood anatomy, phytochemical analysis (using either mass spectrometry or Near Infrared Spectroscopy) and DNA barcodes. Wood anatomy examines the physical characteristics of the wood using standard anatomical characters and is underpinned by

[7]The Wood Explorer 2.0 (www.thewoodexplorerfe.com)

the terminology of the International Association of Wood Anatomists (Richter et al. 2004; Ruffinatto et al. 2015; Wheeler et al. 1989). There are significant skills required to be highly competent in using wood anatomy identification. To alleviate this issue, a computer-based approach, known as machine vision, is being developed (Hermanson and Wiedenhoeft 2011). Using the same physical characters used in manual wood anatomy techniques, machine vision aims to simplify and speed up the identification process and reduce the specialisation requirements that currently inhibit wood anatomy. Both mass spectrometry and Near Infrared Spectroscopy utilise natural chemical patterns that are unique to a taxonomic group (Dormontt et al. 2015). These patterns incorporate both the type of compounds and phytochemicals present for a species as well as the concentration (Finch et al. 2017). The third technique, DNA barcoding, is a genetic method that has been devised through the international barcode of life initiative[8]. This approach uses a set of standard genetic markers (i.e. the DNA barcodes) that can be used to identify taxonomic groups (Hollingsworth et al. 2011; Migone and Howlett 2012).

Due to the technical requirements of these techniques, they are best applied as diagnostic tools. Nevertheless, results can be obtained quickly using wood anatomy (both manual and machine vision) which means they are well suited for use as screening tools by less specialised personnel for rapid field identification (Dormontt et al. 2015). Similarly, due to the straightforwardness of analysing samples using Near Infrared Spectroscopy, and a specific form of mass spectrometry, known as time of flight (TOF-MS), they could also be used as screening tools (Dormontt et al. 2015; Finch et al. 2017; Lancaster and Espinoza 2012). Yet, for the most part the cost of the analysis equipment may be too expensive to see it being used routinely by frontline staff (Dormontt et al. 2015). A more cost-effective phytochemical approach is to use chemical detection tools, such as detector dogs or chemical noses, which can be trained to detect specific volatile organic compounds (VOC) unique to a particular taxonomic group (Ueland et al. 2016). While examples of their application in timber identification are limited (Braun 2013), they have been used successfully in animal wildlife species (Browne et al. 2006; Ueland et al. 2016). However, given the lack of certainty

[8]www.barcodeoflife.org/content/about/barcoding-landscape

surrounding chemical detectors, they may only be suitable as a screening tool (Browne et al. 2006; Schoon 1996)

Origin testing

The origin declaration for an importation refers to the location that the tree was felled or harvested from. The spatial resolution requirements for origin declarations differ between the various timber trading and protection laws. Typically, they are either to country or concession (the precise location within a country where the tree was felled) of harvest (see Lowe AJ *et al* (2016) & Dormontt EE *et al* (2015) for more information). The origin of a shipment is important because the legality of the material may be entirely location specific. For example, the trade of CITES-III listed *Quercus mongolica* (Mongolian oak) is banned from Russia, but not from China or Mongolia where it also naturally occurs. Timber is being illegally felled in Russia, then transported into China where forged Chinese origin declarations are then provided (EIA 2013; Smirnov et al. 2013). Recently a major US flooring retailer was found guilty of (knowingly) trading Russian stock with false origin claims[9]. Therefore, tests that can verify these claims of origin are important.

An origin test is undertaken by identifying the population (or group) of reference individuals with the most similar profile (Ogden and Linacre 2015). The techniques that can be used in origin assessments can be broken into two groups; those that utilise evolutionary based approaches (wood anatomy, mass spectrometry and population genetics & phylogeography (hereafter population genetics)) and those that utilise environmental factors (stable isotopes and dendrochronology). Population genetics is the only evolutionary approach that is targeted for origin testing, as the genetic markers used for these analyses have been specifically designed for the species in question, with limited capability of working outside that species. Whereas for wood anatomy and mass spectrometry, origin assessments are undertaken using the same approach as for name testing. The inclusion of these techniques is ultimately due to the fine scale level of taxonomic determination that can be obtained for some species (Finch et al. 2017; Gasson 2011).

[9]www.justice.gov/opa/pr/lumber-liquidators-inc-sentenced-illegal-importation-hardwood-and-related-environmental

From a purely scientific perspective, population genetics is the study of the genetic patterns that occur across the range of a species. Known as spatial genetic structure, by analysing individuals from across the distribution of the species an understanding of the pattern can be obtained. The genetic patterns of a given species are influenced by a range of factors including life history traits, biogeographical barriers and historical distribution patterns (Lowe and Cross 2011; Vlam et al. 2018). As a rule, individuals that are close in distance are more genetically similar than those further away, and this pattern can be exploited and used in origin tests to identify the most likely origin or to disprove the claimed origin of a test sample (Ogden and Linacre 2015). As mentioned previously, the type of genetic marker used for origin verification tests (and population genetic studies) is different to that needed for species identification. Typically, there is insufficient variation to use a DNA barcode-based approach (but taxa specific markers may be suitable (see Ogden (2015)). Instead, population-based studies make use of more species-specific amplicons that have small hyper variable regions (either Short Tandem Repeats (STR) (also known as microsatellites) or Single Nucleotide Polymorphisms (SNP)) nested within conserved non-variable sequences are better suited (Ogden and Linacre 2015).

Instead of using an evolutionary based framework, the environmental focused techniques - stable isotopes and dendrochronology, use variations and conditions specific to a certain location as the mechanism for origin verification (Dormontt et al. 2015). Stable isotope analysis focuses on the ratios of certain elements (such as hydrogen, carbon, nitrogen, oxygen, sulphur, strontium) (Dormontt et al. 2015; Vlam et al. 2018). For each element, the ratio is strongly influenced by the geographical location, creating a pattern that can be used to assign test samples (Dormontt et al. 2015). Dendrochronology is a form of wood anatomy testing which analyses the growth rings (number, size and pattern) present in wood. The growth rings are influenced by climatic fluctuations and annual growing patterns for an individual tree. Dendrochronology can be used in origin tests by aligning these growth rings with the climatic patterns of a particular location (Speer 2010). However, dendrochronology is only usable in species where distinct growth rings can be identified. Growth rings most readily form in species that occur in regions with distinct seasonal patterns (e.g. temperate forests) and/or in

species that present clear growing and resting phases over a time period (e.g. teak, *Tectonia grandis*) (Dormontt et al. 2015).

Regardless of the technique used for origin verification testing, all are limited by the degree of variation present for a species and/or geographical location, which will affect the spatial resolution to which the origin of a samples can be assessed (Vlam et al. 2018). In some instances, the ability to assign to concession or even country may be limited, making it hard to apply them to the requirements of timber trading and protection legislation. Additionally, for all techniques, rarely do the natural biological boundaries that define evolutionary groups or locations mimic the geopolitical borders that are in place today (Ogden and Linacre 2015). Unless a clear geo-genetic/geopolitical alignment is present then it may be difficult to distinguish between countries with a shared border (Ogden and Linacre 2015).

Date (of felling)

Knowing the date that a sample of timber was felled (or harvested) can be important, as the legality of the felling may have changed over time. Generally, this change is in the form of a legal felling becoming illegal, but in some rare instances the reverse can also occur (for example in 1975 *Platymiscium pleiostachyum* was a CITES-I listed species but was delisted to CITES-II in 1990 ([CITES] 2018; Groves and Rutherford 2015). Regardless of the alteration, the time period when this change occurs determines the legal status of the importation. A good example of this is the trade in Rhinoceros horns, where in some countries it is still legal to trade "historical" or "pre-CITES" material, as long as sufficient evidence as to the age can be provided (see footnote[10] for the Australian government requirements). Therefore, verification of the age of the material is critical. CITES is the most likely legislation that would necessitate any date analysis, as it involves the trade of material from rare and endangered species. Technically, dates are also important in consumer timber trading legislation, where changes to the legality of logging concessions would influence an importation, but it is difficult to see governments making this much effort for species that are not endangered, vulnerable or otherwise at-risk.

[10] www.environment.gov.au/biodiversity/wildlife-trade/permits/pre-cites-certificates

There are two technologies that have been proposed for date determination in timber material, dendrochronology and radiocarbon dating (UNODC 2016; Dormontt et al. 2015). However, given the error rates that occur with dendrochronology (Jones and Daniels 2012), only radiocarbon dating appears to be a reliable technique. Simplified, radiocarbon dating is a technique that measures the ratio between radiocarbons (^{14}C to ^{12}C), which can inform the age of the sample (Ramsey 2008).

To verify the harvesting date of a timber sample, the outermost layer is tested; this is the youngest growth ring (UNODC 2016). If the analysis is to be considered reliable, then this outermost layer needs to be the cambium of the tree, as without this, the harvesting date cannot be ascertained and only a snapshot of the trees life history can be concluded (Dormontt et al. 2015). Considering that the majority of timber material traded has been processed to some degree (e.g. sawnwood lengths) whereby the cambium has been removed, the practicality of using date determination tests is probably reduced (Dormontt et al. 2015).

Match (individualisation) testing

The final focal area for timber identification is match (or individualisation) testing. This type of analysis identifies the significance of a comparison between two or more samples (supposedly from the same individual) against other individuals (UNODC 2016). If the samples were from the same individual, there would be a statistically significant match. Conversely, the absence of any significant result would indicate that the samples are not from the same individual. There are two key applications of match testing within the timber legislation framework: trading (supply chain integrity) and protection (proof of illegal activities) (Lowe and Cross 2011; Nowakowska et al. 2015). From a trading perspective, individualisation testing would be suitable to verify the integrity of a supply chain (Lowe et al. 2010; Migone and Howlett 2012). Integrity testing is done by comparing logs with the same unique identifier (i.e. RFID tag or barcode) at various points along the supply chain (Dormontt et al. 2015). When match testing is used for protection purposes, it is typically employed retroactively to prove if a tree was illegally harvested. For this approach, a comparison is undertaken between the stump of an illegally felled tree and lumber suspected to originate from that tree (UNODC 2016).

There are no technologies that are specifically designed for match testing, but two have been proposed, dendrochronology and DNA profiling (UNODC 2016; Dormontt et al. 2015). Dendrochronology can be applied by comparing the growth ring patterns between samples. However, this technique has been included as a hypothetical method as there are no examples of applying it in this manner (Dormontt et al. 2015). Further research may alter the application of this method to individualisation testing. The second technique, DNA profiling, uses genetic markers for matching samples (Lowe and Cross 2011). Markers that have been designed for population studies or origin determination, such as STRs and SNPs, are well suited, but any amplicon with sufficient variation can be used.

Limitations of the scientific timber identification techniques

There are a wide variety of scientific technologies suitable for timber identification in a forensic context, and each is best applied for answering a particular question regarding a sample's identity. Therefore, no single method could be considered superior to the others. Nevertheless, there are issues that affect these techniques collectively; some of which are inherent to the method. Firstly, the narrow application of each approach in timber identification means that if inquiries into multiple areas of identity are required then more than one technology will be needed to get all the answers (Dormontt et al. 2015). Similarly, the cost of the equipment and/or the level of expertise required to generate results can inhibit the uptake and preference for any one of the technologies (Dormontt et al. 2015). For all techniques there are large investments (money and/or expertise) required, so there will be costs associated with using any method for getting the answers. There is no clear solution to these problems. While the cost of equipment may be reduced in the future, identification capacity is unlikely to be altered because the level of expertise will always restrict the uptake of any method.

Nevertheless, there are also problems that could potentially be overcome. Firstly, there are only limited examples of using any of these techniques in a timber identification manner (Dormontt et al. 2015; Vlam et al. 2018). A lack of clear understanding of the incorporation and use in a legal context truly make it difficult to direct scientific timber identification. Finally, there is a clear lack of reference data available for many of the

technologies, which is inhibiting a more universal uptake (Dormontt et al. 2015; Linacre et al. 2011; Ogden and Linacre 2015). Yet reference content does not need to be generated by forensic facilities to be considered usable in court. Laboratories or research groups that use the aforementioned technologies for their own research can generate reference content for timber species in a forensic context (Johnson et al. 2014).

Part 2: framing the thesis

Introduction: Timber identification testing using genetic methods

This thesis focuses specifically on the use of genetic methods in timber identification testing (aka DNA timber tracking). The use of genetic methods in human criminal cases (DNA identification of human tissue) is fairly well known and has been referenced widely in popular culture. The vast research that has gone into human forensics should attest to the reliability of genetic techniques in legal proceedings (Linacre et al. 2011; Ogden and Linacre 2015). There are also examples of genetic methods being used for non-human wildlife (plant and animal) species in court cases as witnesses for human crimes (Menotti-Raymond et al. 1997; Yoon 1993), victims of human crimes (White et al. 2012), as well as wildlife species directly involved in the crime (e.g. attacks) (Clarke and Vandenberg 2010).

However, to date, there are no examples of using genetic timber identification results in court. But this should not diminish the use of genetic techniques, considering that there are limited examples of prosecutions using any of the timber identification techniques as evidence. In fact, despite the extent of timber trading and protection legislation, there have only been a handful of cases where illegal loggers have been successfully prosecuted at all[11] [12] [13].

[11] www.straitstimes.com/asia/se-asia/indonesian-pulp-company-guilty-of-illegal-logging
[12] www.justice.gov/opa/pr/lumber-liquidators-inc-sentenced-illegal-importation-hardwood-and-related-environmental
[13] www.justice.gov/opa/pr/gibson-guitar-corp-agrees-resolve-investigation-lacey-act-violations

Examples of application

Despite the lack of legal exposure, there are numerous scientific publications that have applied genetic methods to timber identification, including name testing (Asif and Cannon 2005; Höltken et al. 2012; Jiao et al. 2014; Muellner et al. 2011), origin verification (Degen et al. 2013; Finkeldey et al. 2010; Ng et al. 2017; Ng et al. 2016; Ogden 2008; Tang et al. 2011), as well as match testing for supply chain integrity testing (Jolivet and Degen 2012; Lowe et al. 2010; Tnah et al. 2009), and positive matching of individual logs (Nowakowska et al. 2015; Tereba et al. 2017; White et al. 2000). However, considering the scale of the legitimate timber industry (large number of species, global footprint of trading, value of the industry) as well as the known impact and size of illegal practices, the quantity of relevant publications should be many times larger than presented here and further work is still required. It is not only examples that are lacking, the volume of reference data from which timber identification testing can be undertaken is also thoroughly underwhelming. A 2013 internal review by our research group found that in 100 of the most traded species (including those critically endangered, endangered or vulnerable) fewer than 20 had any species-specific population genetic markers available, and only 12 of these had incorporated range wide sampling strategies. Furthermore, of the 80 species with no population studies, 42 had no species level DNA barcodes on either Genbank or BOLD (Barcode Of Life Database). The shortage of reference data not only reduces the capacity to conduct timber identification tests, but it may also be negatively impacting the uptake of genetic timber identification methods by new facilities, especially government forensic departments.

A significant reason that both reference data and published examples of timber identification are limited is most likely attributed to constraints associated with the DNA within timber material (timber DNA). The problem starts with the timber material itself. As a rule, compared to fresh leaf tissue timber DNA is of suboptimal quality and quantity. Furthermore, there are complexities in extracting and using timber DNA that require suitable DNA extraction protocols to facilitate successful extraction. These difficulties, in turn affect many parts of the downstream analysis. The following sections will discuss the extent of these problems in further detail.

Suboptimal quality and quantity of timber DNA

Undoubtedly the most significant limitation for the genetic identification of timber material is associated with the suboptimal quality and quantity of timber DNA (Lowe and Cross 2011). The quality of the DNA that can be obtained during an extraction procedure is negatively impacted by measures used by a plant to survive (e.g. protect themselves from predation) (Pirttilä et al. 2001). Specifically, macromolecules, secondary compounds and metabolites (such as polysaccharides, polyphenols and tannins) occur naturally and serve to protect tissue from attack by foreign organisms (Pirttilä et al. 2001). These compounds readily co-extract with DNA and can affect any part of a genetic analysis, from interference with the cell lysis during DNA extraction or in the downstream amplification/analysis (Alaeddini 2012; Fang et al. 1992; Verbylaite et al. 2010).

While chemical interferences are a universal problem for all plant species and tissue types, there is a particular limitation more prevalent in timber DNA. Because the vast majority of cells in timber material were not living at the time the tree was felled, the quantity of DNA present is reduced (Pirttilä et al. 2001). If you take a cross section of a tree trunk, it is only the outer most layer, the vascular cambium, which contains active cells (Raven et al. 2005). As a tree grows, this cambium layer extends outwards, with new cells being formed (Raven et al. 2005). The original cells are then superseded, remaining in place to either be repurposed as secondary xylem or phloem cells or to provide structural integrity for the tree as it grows (Raven et al. 2005). Furthermore, as cellular activity is not required for this new function, the redundant cells undergo some form of pre-mortem DNA decay and die (Alaeddini et al. 2010; Willerslev and Cooper 2005). Known as apoptosis, it is the active, planned cessation of a cell, with the cellular components being reabsorbed by the surrounding cells (Bortner et al. 1997; Campbell et al. 2006; Katsuhara and Kawasaki 1996). If apoptosis is effective, DNA should be fragmented to the point that it can be reabsorbed into the adjacent cells, reducing the total quantity of DNA present in the remaining tissue. Alternatively, the DNA strands that do remain are heavily fragmented, averaging only 300-400bp in length (Fondevila et al. 2008; Höltken et al. 2012).

Complexities in extracting and using timber DNA

The effect of suboptimal DNA is also exacerbated by the difficulty in the extraction of clean and useable products from timber material. This means that the protocol used to extract DNA needs to be effective and achieve certain objectives that may be unique to timber material. Firstly, any chemical inhibitors need to be removed or neutralised. However, there is no commonality in the types or concentrations of chemicals present between species across the plant kingdom, which reduces the universal application of any single DNA extraction protocol. The range and quantities of inhibiting compounds is so vast and complex that a review (Varma et al. 2007), posed the question: "*Plant Genomic DNA isolation: An art or a science?*" as the title of their paper. Yet despite this, the extraction of DNA from leaf and cambium tissue is relatively straightforward and is typically undertaken with kit-based DNA extraction protocols (Colpaert et al. 2005; Lowe and Cross 2011; Verbylaite et al. 2010). Nevertheless, extraction of timber and degraded material remains much more complex and unreliable (Lowe 2007; Lowe and Cross 2011).

Suitability of current markers for timber DNA

Because timber DNA has been fragmented and is usually in a limited concentration, there are also difficulties associated with amplifying timber DNA using many of the genetic markers available (Cooper and Poinar 2000; Gugerli et al. 2005; Willerslev and Cooper 2005). There may be problems in amplifying some genetic markers, especially DNA barcodes. Many of the universal plant DNA barcodes are unsuitable for working with degraded timber DNA template (Höltken et al. 2012). With median amplicon lengths ranging from 468bp (*psbK-I*) to 994bp (*trnL-F*) (Hollingsworth et al. 2011), they are too long to amplify successfully in timber DNA. Likewise, STR amplicon lengths are often too long for timber DNA, ranging in size from 80-300bp (Alaeddini et al. 2010; Fondevila et al. 2008; Schmerer et al. 1999). Additionally, STR markers can produce erroneous results that arise from problems with allelic dropout or miscoding (Fondevila et al. 2008). This issue occurs even in good quality samples, so it is expected more frequently in degraded samples, such as timber (Butler et al. 2003; Gugerli et al. 2005; Schmerer et al. 1999). In an attempt to overcome these problems when using timber samples, several papers have reported repeating the genotyping of samples (Degen et al. 2013; Jolivet and Degen 2012), but this is not an ideal solution. SNP markers too are

limited. While being relatively simple and available in short amplicons, they can be readily affected by base alterations in timber samples (Alaeddini et al. 2010). Additionally, the discriminatory power of SNPs is far less than in STRs. For example, in humans a four-fold increase in the number of screened loci was required to produce similar discriminatory results (Alaeddini et al. 2010; Sobrino et al. 2005). The lack of suitability of most existing makers to amplify timber DNA means that there is a further reduction in the already limited amount of reference data (DNA barcodes and species-specific population markers) that can be used for timber identification purposes.

Overcoming the limitations to genetic timber identification

While the constrictions associated with timber DNA presently inhibit the usability of current genetic methods for timber identification, they can be overcome. The challenges faced in extraction and amplification are not unique to timber DNA and have been discussed at length in similar research fields such as ancient DNA (aDNA) (Gugerli et al. 2005; Pääbo et al. 2004; Willerslev and Cooper 2005) and forensic DNA (Alaeddini 2012; Alaeddini et al. 2010; Schneider 1997). Furthermore, both areas have been highly successful in their ability to extract, amplify and analyse DNA from highly degraded and sensitive material. So utilising advances identified from these areas can help in applying them to timber DNA. The next sections will discuss some solutions to allow the use of genetic methods for timber identification more readily.

Timber specific DNA extraction protocols

As previously mentioned, the extraction of DNA from timber of sufficient quantity and quality is not as routine when compared to fresh leaf and cambium tissue (Lowe 2007; Lowe and Cross 2011). The first publications that use timber DNA in their research have been around for nearly 20 years (De Filippis and Magel 1998; Dumolin-Lapègue et al. 1999), and there are multiple extraction protocols reported as suitable for timber material (Varma et al. 2007; Weising et al. 1991). Of the DNA extraction protocols that have been published, the most commonly used kit based protocol is a modified DNeasy Plant Mini kit (QIAGEN; Hilden, Germany) (Deguilloux M F et al. 2002; Jiao et al. 2014; Rachmayanti et al. 2009), with the innuPREP Plant DNA Kit (ANALYTIK JENA; Jena, Germany) also being reported (Telle and Thines 2008; Verbylaite et al. 2010). The CTAB (cetyltrimethylammonium bromide) based protocol (including variations to the original

Doyle and Doyle protocol) (Doyle and Dickson 1987; Doyle and Doyle 1987; Varma et al. 2007) is the most widely reported in-house extraction protocol. Other in-house protocols to be reported include PTB (N-phenacylthiazolium bromide) (Asif and Cannon 2005) and ATMAB (Alkyltrimethylammonium bromide) (Degen et al. 2013).

Choice of genetic markers

When selecting which genetic markers to use, the first thing that needs to be considered is its suitability for amplifying timber DNA. For a genetic marker to be ideal for timber DNA it needs to be short in length and have a simple variation. As a consequence, many existing genetic markers are unsuitable for working with timber DNA. In some instances, these markers can be modified to work in timber DNA by shortening the amplicon length and utilising additional priming sites internal to existing ones. There are a few DNA barcode examples of this (Höltken et al. 2012; Jiao et al. 2014; Taberlet et al. 2006). However, these alterations limit the capacity of the barcode. Because the modified amplicons can only be designed for a target species/genus/family, which reduces the universality of the marker. Amplicon modifications are most effective in population genetic markers and have been well reported in STR loci. Known as "miniSTR's", they have been shown to be an effective and reliable approach for amplifying DNA from degraded material such as timber (Alaeddini et al. 2010; Butler et al. 2003; Fondevila et al. 2008). The idea of a miniSTR is to locate new priming sites as close to the repeat motif as possible. But, there are only a finite number of appropriate sites, which limits the usability of miniSTR markers (Butler et al. 2003).

Alas, either the majority of existing genetic markers cannot be modified, or it is not cost effective to do so. This means that novel markers that are suitable for timber DNA need to be developed to increase the usability of genetic methods in timber identification. Fortunately, the selection and development of novel genetic markers comes at an exciting time in genetics (Cross et al. 2016; Ogden 2011). The advancements of new sequencing techniques have been incredibly advantageous for many organisms, especial for non-model species (i.e. timber species) (Cross et al. 2016). In a nutshell these new techniques, commonly known by several names such as High Throughput Sequencing (HTS) (the term used hereafter), Massively Parallel Sequencing (MPS) or Next Generation Sequencing (NGS), allow for an exponentially greater amount of information

to be derived at any given time (Cross et al. 2016). HTS is most suitable during the marker development and screening phases of laboratory procedures. It is through the large scale/high throughput techniques incorporated by these platforms, that the identification and provision of results with a high degree of support (coverage) are identified (Ebert and Peakall 2009).

Furthermore, HTS is suitable for developing novel DNA barcodes as well as species-specific STR and SNP variable amplicons. For DNA barcodes, the most suitable technique for developing new markers would be via whole genome comparisons. The development of novel DNA barcodes that are suitable for timber and other degraded material has been identified as the next phase of the Barcode of Life initiatives (Hollingsworth PM (June 2017) *pers. comm.*). The discovery of informative species-specific markers can be undertaken in a variety of ways (Cronn et al. 2012). Of particular interest is the use of restriction-enzyme enrichment techniques, which, when combined with HTS, can provide large datasets and identification of high confidence loci at a cost effective price (Cronn et al. 2012; Stolle and Moritz 2013; van Orsouw et al. 2007). In human forensics, miniSTRs along with SNPs are regarded as the most suitable amplicon for working with degraded DNA (Butler et al. 2003; Fondevila et al. 2008; Sobrino et al. 2005). But given the high degree of uncertainty associated with STRs in general (Ballantyne et al. 2007), especially in amplifying timber DNA, the application of them as a verified tool for monitoring and tracking timber may not be appropriate.

SNPs are the most common form of polymorphism in a genome (Cronn et al. 2012; Sobrino et al. 2005) and are well suited for use in HTS (Sobrino et al. 2005). SNPs are also found in shorter amplicons than STRs. They have also been shown to be more immune to the effects of inhibition, making them ideal for amplifying degraded DNA samples (Fondevila et al. 2008; Sobrino et al. 2005). Given that SNPs are being suggested as an ideal marker type for working with degraded DNA (Ogden and Linacre 2015), and that results generated by this method can be provided, reviewed and replicated, it would be wise to consider applying this marker type to amplify timber DNA.

Using a novel approach to genotyping

Once a set of suitable genetic markers has been identified, they will need to be amplified and analysed. Unlike the amplification of both DNA barcodes and STR markers, which use capillary separation technologies (i.e. Sanger sequencing; genotyping by capillary electrophoresis), there is no single methodology used to screen for SNP variation (Fondevila et al. 2008; Sobrino et al. 2005). Additionally, a review of the various SNP genotyping methodologies suitable for forensic identification applications found it difficult to identify the most appropriate approach (Fondevila et al. 2008; Sobrino et al. 2005). The review also identified that it was entirely dependent on the application of the genotyping (Fondevila et al. 2008; Sobrino et al. 2005). There are some examples of SNP assays for origin testing in wildlife species and these have been summarised in Table 3.

Table 3 SNP genotyping techniques: Various SNP genotyping techniques and examples of application in origin testing of wildlife species

Reaction	Detection	Name	Species/group	Reference
Oligonucleotide ligation	Electrophoresis	ϒSNPlex™	Grapevine cultivars	1
	Array	ΦGoldenGate™	Fish	2
			Eucalyptus	3
Allele Specific Hybridisation	⍰FRET	ϒTaqMan®	Ramin	4
Primer Extension	Electrophoresis	ϒSNaPshot™	Tiger species/subspecies	5
	⍰MALDI-TOF MS	ϖMassArray®	Sapelli	6
			Iroko	7
			Salmon	8

NB: table adapted from Sobrino (2005); ⍰FRET = Fluorescence resonance energy transfer, ⍰MALDI_TOF MS = Matrix Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry, ϒAPPLIED BIOSYSTEMS (Foster City, CA, USA), ΦILLUMINA® (San Diego, CA, USA), ϖAGENA BIOSCIENCE™ (San Diego, CA, USA). References: 1 = (Cabezas et al. 2011), 2 = (Martinsohn et al. 2009), 3 = (Correia et al. 2011), 4 = (Ogden 2008), 5 = (Kitpipit et al. 2012), 6 = (Degen et al. 2017), 7 = (Blanc-Jolivet et al. 2017), 8 = (Glover et al. 2010).

Understanding capacity

For any novel species-specific population makers (i.e. STRs and SNPs), there are additional studies that need to be conducted before they can be utilised in forensic timber identification tests. The most important study is in relation to origin verification and individualisation tests for a given species. For these tests, an understanding of the underlying population genetic patterns must be known, as this will indicate the capacity to assign and assess the origins of unknown samples. The discriminatory power of the markers also needs to be assessed and reported. This is done by assigning blinded/unknown timber or leaf samples not already contained in the reference dataset and assessing the capacity of the test to either identify the correct origin of the sample,

or to identify the confidence of detecting samples with falsified origin claims. To assist in the support of the techniques used for assignment testing, the capacity, success and constraints of each technique should be identified and published. Additionally, some of this information is required for the forensic validation of the markers and assignment protocol used (SWGDAM 2016; SWFS 2018; Linacre et al. 2011).

Part 3: Thesis overview

Thesis summary

As outlined above there are various aspects of genetic timber identification that are limiting its application within a legal setting. This thesis consists of research conducted to address some of those limitations and provide evidence as to the legitimacy of genetic methods in timber identification. The studies undertaken as part of this candidature include: optimising, refining and comparing DNA extraction protocols to make them suitable for timber material (*Ch. 2*); using HTS techniques to develop timber DNA appropriate markers for two unrelated timber species (*Ch. 3*); utilising the markers from one species and undertaking a population genetics study across its natural range (*Ch. 4*); demonstrating how those same markers can be used in origin verification testing of blinded timber samples (*Ch. 5*).

The following paragraphs provide a more detailed summary of the research undertaken as part of this candidature.

Chapter 2: BOTAB DNA extraction protocol: optimisation and comparison study

This chapter is presented as a traditional thesis chapter. It revolves around the need for suitable DNA extraction protocols for timber. It begins by providing a more in-depth summary of the limitations arising from timber material, as well as highlighting solutions that have been put forward to facilitating DNA extractions. The main focus of this chapter is around the work undertaken as part of a patent application for a DNA extraction from timber material protocol (the patent has been included in the appendix). Additionally, the chapter includes some of the work that has been undertaken to forensically validate the protocol. The chapter concludes with a summary of the state of play including the current status of the protocol as well as where further work is required.

Chapter 3: Development of novel SNP markers for bigleaf maple and ayous

This chapter incorporates two primer note publications for both bigleaf maple (*Acer macrophyllym*) and ayous (*Triplochiton scleroxylon*). The publications have been presented together primarily because the methodology used to generate the markers was similar with only slight differences. Additionally, both are published in the same journal (Conservation Genetic Resources), so the reported information can be paralleled. For both papers, the publication (including any supplementary material) is presented in the thesis. The papers provide information on the use of new sequencing approaches to develop novel species-specific markers. The papers then provide a description on the markers that were identified and brief summary statistics on level of heterozygosity and fixation for each species.

Chapter 4: Population Genetics of Triplochiton scleroxylon

This chapter is focused around a population genetic analysis of ayous, an economically important timber species from Guineo-Congolian Africa. A total of 911 samples from 43 populations in five countries (Ivory Coast, Ghana, Cameroon, Republic of the Congo and Democratic Republic of the Congo (COD)) across the known distribution of ayous, were genotyped using the SNP markers developed previously (*Ch. 3*$_{II}$). Uninformative loci and samples that failed to genotype sufficiently were then removed. This left a working dataset consisting of 105 loci and 753 individuals. A population clustering analysis revealed there are three genetic populations (clusters) for ayous. There is one cluster in the west, the western region, which consists of samples from both Ivory Coast and Ghana. The second cluster, the central region cluster, encompasses all samples from Cameroon and the Republic of the Congo as well as samples from North Western COD. The third cluster is to the east and is a small cluster that only contains samples from central COD. The formation of distinct western and central region clusters is consistent with previous studies and was to be expected. The forests of tropical equatorial Africa are not continuous, and a single 200 km savannah gap in Togo and Benin divides the forest. Known as the Dahomey gap, it acts as a genetic barrier and limits the amount of geneflow for species. Our results also offer novel insights into genetic clustering patterns with the identification of a genetic separation between individuals from central and north western COD. We found that the north-western samples are more genetically similar to those in Cameroon and Republic of the Congo and that the central COD

samples form their own genetically isolated cluster. We hypothesise that this separation can be attributed the effect of recent historical climatic conditions. We postulate that the grouping of the North Western samples in the central region cluster is the result of forest expansions from coastal Cameroon refugial locations. While the central samples derived from a single inland refugium, situated near their present location.

Chapter 5: Assignment testing in ayous

This chapter is framed around assignment testing of ayous timber samples. Assignment testing has been proposed as a highly suitable technique for assessing the claimed origin of traded species, such as timber, and would fit within the existing legal framework designed to reduce the effect of illegitimate harvesting. Unfortunately for most species, the natural genetic structure rarely reflects something that is legally meaningful. This discrepancy can influence the ability to assign to these artificial geopolitical groups. The lack of geogenetic-geopolitical alignment forces a trade-off between generating a result that is reliable (geogenetic) with one that is jurisdictionally definable (geopolitical). The population genetics chapter (*Ch. 4*) revealed that ayous is a species whose geogenetic-geopolitical populations do not align. This final chapter explores the effect that this misalignment has on assignment testing, by looking at the impact and capability of assessing claims of blinded timber samples. For this chapter, an origin verification test was conducted for 25 blinded timber samples. To assess the claims, the working dataset developed for the population genetic analysis (*Ch. 4*) was used as a reference dataset. In addition to this format, to understand the effect of misalignment, the dataset was structured in two alternative arrangements; one to represent the country of origin, the other the genetic population of origin. Unsurprisingly, the results show that assignment testing was most suitable for genetic population assessments, while results from the country assessments were less reliable. There were even fewer reliable assignments when assigning to the geographical population level. Despite this, in most instances this level of spatial resolution would not be required for testing. The inability to detect cross border falsification events from within a genetic population makes assigning to these groups' problematic. Because of this, we found that while country of origin tests are not the most reliable, they are still the most usable and practical format for performing an assignment test in this context.

Chapter 6: General discussion and Thesis conclusion

This chapter incorporates three areas. Firstly, it provides a summary of the introduction chapter and gives further context for the data chapters (*Ch. 2-5*) and how they fit to the synthesis of the thesis as a whole. The chapter then goes on to discuss limitations and highlighting problems found during the candidature. The discussion concludes with a section looking to the future, identifying additional work directly arising from the research and presenting opportunities for further development in this space. It also includes a statement about the future of the timber identification field and what may be needed to get this technology used more widely.

General Introduction References

The Checklist of CITES Species Website (2018). CITES Secretariat, Geneva, Switzerland. http://checklist.cites.org. Accessed 20/11/2017

[EIA] Environmental Investigation Organisation (2012) Appetite for Destruction: China's trade in illegal timber.

[EIA] Environmental Investigation Organisation (2013) Liquidating the Forests: Hardwood flooring, organised crime and the world's last Siberian tigers.

[FAO] Food and Agriculture Organisation (2014) State of the World's Forests 2014. FAO, Rome

[ITTO] International Tropical Timber Organisation(2015) Biennial Review and Assessment of the World Timber Situation 2013-2014.

[SWGDAM] Scientific Working Group for DNA Analysis Methods (2016) SWGDAM Validation Guidelines for DNA Analysis Methods https://docswixstaticcom/ugd/4344b0_813b241e8944497e99b9c45b163b76bdpdf

[SWFS] Society for Wildllife Forensic Sciences Technical Working Group (2018) Standards and Guidelines for Wildlife Forensic Analysis, Version 3. Ed. Lucy M.I. Webster. Published by the Society for Wildlife Forensic Science, 19th November 2018, pp.21. https://www.wildlifeforensicscience.org/wp-content/uploads/2018/11/SWFS-Standards-and-Guidelines_Version-3_19-11-18.pdf

[UNODC] United Nations Office on Drugs and Crime (2016) Best Practice Guide for Forensic Timber Identication. United Nations, United Nations Office at Vienna, Austria

Alaeddini R (2012) Forensic implications of PCR inhibition—a review. Forensic Science International: Genetics 6:297-305 doi:10.1016/j.fsigen.2011.08.006

Alaeddini R, Walsh SJ, Abbas A (2010) Forensic implications of genetic analyses from degraded DNA—a review. Forensic Science International: Genetics 4:148-157 doi:10.1016/j.fsigen.2009.09.007

Asif MJ, Cannon CH (2005) DNA extraction from processed wood: a case study for the identification of an endangered timber species (Gonystylus bancanus). Plant Molecular Biology Reporter 23:185-192 doi:10.1007/BF02772709

Ballantyne KN, van Oorschot RA, Mitchell RJ (2007) Comparison of two whole genome amplification methods for STR genotyping of LCN and degraded DNA samples. Forensic Science International 166:35-41 doi:10.1016/j.forsciint.2006.03.022

Blanc-Jolivet C, Kersten B, Daïnou K, Hardy OJ, Guichoux E, Delcamp A, Degen B (2017) Development of nuclear SNP markers for genetic tracking of Iroko, *Milicia excelsa* and *Milicia regia*. Conservation Genetics Resources:1-3 doi:10.1007/s12686-017-0716-2

Bortner CD, Hughes FM, Cidlowski JA (1997) A primary role for K+ and Na+ efflux in the activation of apoptosis. Journal of Biological Chemistry 272:32436-32442 doi:10.1074/jbc.272.51.32436

Bradshaw CJ, Sodhi NS, Brook BW (2009) Tropical turmoil: a biodiversity tragedy in progress. Frontiers in Ecology and the Environment 7:79-87 doi:10.1890/070193
Braun B (2013) Wildlife Detector Dogs—A Guideline on the Training of Dogs to Detect Wildlife in Trade.
Browne C, Stafford K, Fordham R (2006) The use of scent-detection dogs. Irish Veterinary Journal 59:97
Butler JM, Shen Y, McCord BR (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. Journal of forensic sciences 48:1054-1064 doi:10.1520/JFS2003043
Cabezas JA et al. (2011) A 48 SNP set for grapevine cultivar identification. BMC Plant Biology 11:1 doi:10.1186/1471-2229-11-153
Campbell NA, Reece JB, Meyers N (2006) Biology. 7 edn. Pearson Education Australia,
Clarke M, Vandenberg N (2010) Dog attack: the application of canine DNA profiling in forensic casework. Forensic Science, Medicine, and Pathology 6:151-157 doi:10.1007/s12024-009-9114-8
Colpaert N, Cavers S, Bandou E, Caron H, Gheysen G, Lowe A (2005) Sampling tissue for DNA analysis of trees: trunk cambium as an alternative to canopy leaves. Silvae Genetica 54:265-269
Cooper A, Poinar HN (2000) Ancient DNA: do it right or not at all. Science 289:1139-1139 doi:10.1126/science.289.5482.1139b
Correia L, Faria D, Grattapaglia D (2011) Comparative assessment of SNPs and microsatellites for fingerprinting, parentage and assignment testing in species of Eucalyptus. BMC Proceedings 5:41 doi:10.1186/1753-6561-5-S7-P41
Cronn R, Knaus BJ, Liston A, Maughan PJ, Parks M, Syring JV, Udall J (2012) Targeted enrichment strategies for next-generation plant biology. American journal of botany 99:291-311 doi:10.3732/ajb.1100356
Cross H, Biffin E, Lowe A, Waycott M (2016) Effective application of next-generation sequencing (NGS) approaches in systematics and population genetics: case studies in Eucalyptus and Acacia. Australian Systematic Botany 29:235-246 doi:10.1071/SB16019
De Filippis L, Magel E (1998) Differences in genomic DNA extracted from bark and from wood of different zones in Robinia trees using RAPD-PCR. Trees 12:377-384 doi:10.1007/PL00009723
Degen B (2012) Practical applications of DNA technologies in timber identification and tracking. Forestry Research Network of Sub-Saharan Africa, http://www.fornis.net/content/forest-health-biodiversity-and-conservation
Degen B, Blanc-Jolivet C, Stierand K, Gillet E (2017) A nearest neighbour approach by genetic distance to the assignment of individual trees to geographic origin. Forensic Science International: Genetics 27:132-141 doi:10.1016/j.fsigen.2016.12.011
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 Science International: Genetics 7:55-62 doi:10.1016/j.fsigen.2012.06.003

Deguilloux M F, Pemonge M H, Petit R J (2002) Novel perspectives in wood certification and forensics: dry wood as a source of DNA. Proceedings of the Royal Society of London B: Biological Sciences 269:1039-1046 doi:10.1098/rspb.2002.1982
Dixon RK, Solomon A, Brown S, Houghton R, Trexier M, Wisniewski J (1994) Carbon pools and flux of global forest ecosystems. Science 263:185-190 doi:10.1126/science.263.5144.185
Dormontt EE et al. (2015) Forensic timber identification: It's time to integrate disciplines to combat illegal logging. Biological Conservation 191:790-798 doi:10.1111/1755-0998.12387
Doyle JJ, Dickson EE (1987) Preservation of plant samples for DNA restriction endonuclease analysis. Taxon:715-722
Doyle JJ, Doyle J (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19:11-15
Dumolin-Lapègue S, Pemonge MH, Gielly L, Taberlet P, Petit R (1999) Amplification of oak DNA from ancient and modern wood. Molecular Ecology 8:2137-2140 doi:10.1046/j.1365-294x.1999.00788.x
Ebert D, Peakall R (2009) A new set of universal de novo sequencing primers for extensive coverage of noncoding chloroplast DNA: new opportunities for phylogenetic studies and cpSSR discovery. Molecular Ecology Resources 9:777-783
Fang G, Hammar S, Grumet R (1992) A quick and inexpensive method for removing polysaccharides from plant genomic DNA. Biotechniques 13:52-54, 56
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. Applications in Plant Sciences 5:1600158 doi:10.3732/apps.1600158
Finkeldey R, Leinemann L, Gailing O (2010) Molecular genetic tools to infer the origin of forest plants and wood. Applied Microbiology and Biotechnolgy 85:1251-1258 doi:10.1007/s00253-009-2328-6
Finkeldey R, Rachmayanti Y, Nuroniah H, Nguyen NP, Cao C, Gailing O Identification of the timber origin of tropical species by molecular genetic markers–the case of dipterocarps. In: Proceeding of the International Workshop “Fingerprinting methods for the identification of timber origins”. Bonn, Germany, 2007. pp 143-158
Fondevila M et al. (2008) Challenging DNA: assessment of a range of genotyping approaches for highly degraded forensic samples. Forensic Science International: Genetics Supplement Series 1:26-28 doi:10.1016/j.fsigss.2007.10.057
Gasson P (2011) How precise can wood identification be? Wood anatomy’s role in support of the legal timber trade, especially CITES. IAWA Journal 32:137-154 doi:10.1163/22941932-90000049
Glover KA, Hansen MM, Lien S, Als TD, Høyheim B, Skaala Ø (2010) A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. BMC Genetics 11:1 doi:10.1186/1471-2156-11-2
Groves M, Rutherford C (2015) CITES and Timber: A guide to CITES-listed tree species. Royal Botanic Gardens, Kew Royal Botanic Gardens, Kew, UK

Gugerli F, Parducci L, Petit RJ (2005) Ancient plant DNA: review and prospects. New Phytologist 166:409-418 doi:10.1111/j.1469-8137.2005.01360.x
Hermanson JC, Wiedenhoeft AC (2011) A brief review of machine vision in the context of automated wood identification systems. IAWA Journal 32:233-250 doi:10.1163/22941932-90000054
Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PloS ONE 6:e19254 doi:10.1371/journal.pone.0019254
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 66:97-104
Jiao L, Yin Y, Cheng Y, Jiang X (2014) DNA barcoding for identification of the endangered species Aquilaria sinensis: comparison of data from heated or aged wood samples. Holzforschung 68:487-494 doi:10.1515/hf-2013-0129
Johnson RN, Wilson-Wilde L, Linacre A (2014) Current and future directions of DNA in wildlife forensic science. Forensic Science International: Genetics 10:1-11 doi:10.1016/j.fsigen.2013.12.007
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 Science International: Genetics 6:487-493 doi:10.1016/j.fsigen.2011.11.002
Jones EL, Daniels LD (2012) Assessment of dendrochronological year-of-death estimates using permanent sample plot data. Tree-Ring Research 68:3-16 doi:10.3959/2010-10.1
Katsuhara M, Kawasaki T (1996) Salt stress induced nuclear and DNA degradation in meristematic cells of barley roots. Plant and Cell Physiology 37:169-173
Kitpipit T, Tobe SS, Kitchener AC, Gill P, Linacre A (2012) The development and validation of a single SNaPshot multiplex for tiger species and subspecies identification—Implications for forensic purposes. Forensic Science International: Genetics 6:250-257 doi:10.1016/j.fsigen.2011.06.001
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 Communications in Mass Spectrometry 26:1147-1156 doi:10.1002/rcm.6215
Linacre A et al. (2011) ISFG: recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. Forensic Science International: Genetics 5:501-505 doi:10.1016/j.fsigen.2010.10.017
Lowe AJ Can we use DNA to identify the geographic origin of tropical timber. In: Proceedings of the international workshop "Fingerprinting methods for the identification of timber origins" October, 2007. pp 8-9
Lowe AJ, Cross HB (2011) The application of DNA methods to timber tracking and origin verification. IAWA Journal 32:251 doi:10.1163/22941932-90000055
Lowe AJ et al. (2016) Opportunities for Improved Transparency in the Timber Trade through Scientific Verification. BioScience 66:990-998 doi:10.1093/biosci/biw129

Lowe AJ, Wong KN, Tiong YS, Iyerh S, Chew FT (2010) A DNA Method to Verify the Integrity of Timber Supply Chains; Confirming the Legal Sourcing of Merbau Timber From Logging Concession to Sawmill. Silvae Genetica 59:263
Martinsohn JT, Ogden R, FishPopTrace Consortium (2009) FishPopTrace—Developing SNP-based population genetic assignment methods to investigate illegal fishing. Forensic Science International: Genetics Supplement Series 2:294-296 doi:10.3354/esr00144
Menotti-Raymond MA, David VA, O'brien SJ (1997) Pet cat hair implicates murder suspect. Nature 386:774 doi:10.1038/386774a0
Migone A, Howlett M (2012) From paper trails to DNA barcodes: Enhancing traceability in forest and fishery certification. Natural Resources Journal:421-441
Muellner A, Schaefer H, Lahaye R (2011) Evaluation of candidate DNA barcoding loci for economically important timber species of the mahogany family (Meliaceae). Molecular Ecology Resources 11:450-460 doi:10.1111/j.1755-0998.2011.02984.x
Nellemann C, INTERPOL (2012) Green carbon, black trade: illegal logging, tax fraud and laundering in the world's tropical forests. A Rapid Response Assessment. United Nations Environment Programme, GRID-Arendal,
Nellemann C, Henriksen R, Raxter P, Ash N, Mrema E, (Eds.) (2014) The environmental crime crisis: threats to sustainable development from illegal exploitation and trade in wildlife and forest resources. A UNEP Rapid Response Assessment. United Nations Environment Programme (UNEP) and GRID-Arendal, Nairobi and Arendal
Ng CH, Lee SL, Tnah LH, Ng KKS, Lee CT, Diway B, Khoo E (2017) Geographic origin and individual assignment of *Shorea platyclados* (Dipterocarpaceae) for forensic identification. PloS ONE 12:e0176158 doi:10.1371/ journal.pone.0176158
Ng KKS et al. (2016) Forensic timber identification: a case study of a CITES listed species, *Gonystylus bancanus* (Thymelaeaceae). Forensic Science International: Genetics 23:197-209 doi:10.1016/j.fsigen.2016.05.002
Nowakowska JA, Oszako T, Tereba A, Konecka A (2015) Forest tree species traced with a DNA-based proof for illegal logging case in Poland. In: Evolutionary biology: biodiversification from genotype to phenotype. Springer, pp 373-388. doi:10.1007/978-3-319-19932-0_19
Ogden R (2008) Fisheries forensics: the use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. Fish and Fisheries 9:462-472 doi:10.1111/j.1467-2979.2008.00305.x
Ogden R (2011) Unlocking the potential of genomic technologies for wildlife forensics. Molecular Ecology Resources 11:109-116 doi:10.1111/j.1755-0998.2010.02954.x
Ogden R, Linacre A (2015) Wildlife forensic science: A review of genetic geographic origin assignment. Forensic Science International: Genetics 18:152-159 doi:10.3354/esr00144
Pääbo S et al. (2004) Genetic analyses from ancient DNA. Annual review of genetics 38:645-679 doi:10.1146/annurev.genet.37.110801.143214

Pirttilä AM, Hirsikorpi M, Kämäräinen T, Jaakola L, Hohtola A (2001) DNA isolation methods for medicinal and aromatic plants. Plant Molecular Biology Reporter 19:273-273 doi:10.1007/BF02772901
Rachmayanti Y, Leinemann L, Gailing O, Finkeldey R (2009) DNA from processed and unprocessed wood: factors influencing the isolation success. Forensic Science International: Genetics 3:185-192 doi:10.1016/j.fsigen.2009.01.002
Ramsey CB (2008) Radiocarbon dating: revolutions in understanding. Archaeometry 50:249-275 doi:10.1111/j.1475-4754.2008.00394.x
Raven PH, Evert RF, Eichhorn SE (2005) Biology of plants. Macmillan,
Richter HG, Grosser D, Heinz I, Gasson P (2004) IAWA list of microscopic features for softwood identification. IAWA Journal 25:1-70 doi:10.1163/22941932-90000349
Ruffinatto F, Crivellaro A, Wiedenhoeft AC (2015) Review of macroscopic features for hardwood and softwood identification and a proposal for a new character list. IAWA Journal 36:208-241 doi:10.1163/22941932-00000096
Sasaki N et al. (2016) Sustainable management of tropical forests can reduce carbon emissions and stabilize timber production. Frontiers in Environmental Science 4:50 doi:10.3389/fenvs.2016.00050
Schmerer WM, Hummel S, Herrmann B (1999) Optimized DNA extraction to improve reproducibility of short tandem repeat genotyping with highly degraded DNA as target. Electrophoresis 20:1712-1716
Schneider PM (1997) Basic issues in forensic DNA typing. Forensic Science International 88:17-22 doi:10.1016/S0379-0738(97)00079-0
Schoon GAA (1996) Scent identification lineups by dogs (*Canis familiaris*): experimental design and forensic application.. Applied Animal Behaviour Science 49:257-267
Smirnov DY, Kabanets AG, Milakovsky BJ, Lepeshkin EA, Sychikov D (2013) Illegal logging in the Russian Far East: global demand and taiga destruction.
Sobrino B, Brión M, Carracedo A (2005) SNPs in forensic genetics: a review on SNP typing methodologies Forensic Science International 154:181-194 doi:10.1016/j.forsciint.2004.10.020
Speer JH (2010) Fundamentals of tree-ring research. University of Arizona Press,
Stolle E, Moritz RF (2013) RESTseq–Efficient Benchtop Population Genomics with RESTriction Fragment SEQuencing. PloS ONE 8:e63960
Taberlet P et al. (2006) Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. Nucleic acids research 35:e14-e14 doi:10.1093/nar/gkl938
Tang X, Zhao G, Ping L (2011) Wood identification with PCR targeting noncoding chloroplast DNA. Plant Molecular Biology 77:609-617 doi:10.1007/s11103-011-9837-2
Telle S, Thines M (2008) Amplification of *cox2* (~ 620 bp) from 2 mg of up to 129 years old herbarium specimens, comparing 19 extraction methods and 15 polymerases. PLoS ONE 3:e3584 doi:10.1371/journal.pone.0003584
Tereba A, Woodward S, Konecka A, Borys M, Nowakowska JA (2017) Analysis of DNA profiles of ash (Fraxinus excelsior L.) to provide evidence of illegal

logging. Wood Science and Technology 51:1377-1387 doi:10.1007/s00226-017-0942-5
Tnah LH, Lee SL, Ng KK, Tani N, Bhassu S, Othman RY (2009) Geographical traceability of an important tropical timber (*Neobalanocarpus heimii*) inferred from chloroplast DNA. Forest Ecology and Management 258:1918-1923 doi:10.1016/j.foreco.2009.07.029
Ueland M, Ewart K, Troobnikoff AN, Frankham G, Johnson RN, Forbes SL (2016) A rapid chemical odour profiling method for the identification of rhinoceros horns. Forensic Science International 266:e99-e102 doi:10.1016/j.forsciint.2016.05.011
van Orsouw NJ et al. (2007) Complexity reduction of polymorphic sequences (CRoPS™): a novel approach for large-scale polymorphism discovery in complex genomes. PloS ONE 2:e1172 doi:10.1371/journal.pone.0001172
Varma A, Padh H, Shrivastava N (2007) Plant genomic DNA isolation: an art or a science. Biotechnology Journal 2:386-392 doi:10.1002/biot.200600195
Verbylaite R, Beisys P, Rimas V, Kuusiene S (2010) Comparison of Ten DNA Extraction Protocols from Wood of European Aspen (Populus tremula L.). Baltic Forestry 16:35-42
Vlam M et al. (2018) Developing forensic tools for an African timber: Regional origin is revealed by genetic characteristics, but not by isotopic signature. Biological Conservation 220:262-271 doi:10.1016/j.biocon.2018.01.031
Weising K, Beyermann B, Ramser J, Kahl G (1991) Plant DNA Fingerprinting with Radioactive and Digoxigenated Oligonucleotide Probes Complementary to Simple Repetitive DNA-Sequences. Electrophoresis 12:159-169 doi:10.1002/Elps.1150120211
Wheeler EA, Baas P, Gasson PE (1989) IAWA list of microscopic features for hardwood identification
White E, Hunter J, Dubetz C, Brost R, Bratton A, Edes S, Sahota R (2000) Microsatellite markers for individual tree genotyping: application in forest crime prosecutions. Journal of Chemical Technology and Biotechnology 75:923-926
White NE, Dawson R, Coghlan ML, Tridico SR, Mawson PR, Haile J, Bunce M (2012) Application of STR markers in wildlife forensic casework involving Australian black-cockatoos (Calyptorhynchus spp.). Forensic Science International: Genetics 6:664-670 doi:10.1016/j.fsigen.2011.10.003
Willerslev E, Cooper A (2005) Review Paper. Ancient DNA. Proceedings of the Royal Society of London B: Biological Sciences 272:3-16 doi:10.1098/rspb.2004.2813
Yamamoto Y, Takeuchi K, Shinkuma T (2014) Is there a price premium for certified wood? Empirical evidence from log auction data in Japan. Forest Policy and Economics 38:168-172 doi:10.1016/j.forpol.2013.07.002
Yoon CK (1993) Botanical witness for the prosecution. Science 260:894-896

Chapter 2: BOTAB DNA extraction protocol: optimisation and comparison study

Authors: Duncan Iain Jardine[1], Lasse Schindler[2,3], Bernd Degen[2], Andrew Lowe[1],

Affiliations: [1]School of Biological Sciences & Environment Institute, University of Adelaide, Adelaide, SA 5005, Australia; [2]Thünen Institute of Forest Genetics, Sieker Landstraße 2, 22927 Großhansdorf, Germany; [3]LABiTec LAbor BioMedical Technologies GmbH, An der Strusbek 6, 22926 Ahrensburg, Germany.

Statement of Authorship

Title of Paper	BOTAB DNA extraction protocol: optimisation and comparison study
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	N/A

Principal Author

Name of Principal Author (Candidate)	*Duncan Iain Jardine*		
Contribution to the Paper	co-planned experiments; co-conducted lab work; conducted data analysis; conducted literature review; wrote manuscript as principal author		
Overall percentage (%)	40%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	30.07.2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	*Lasse Schindler*		
Contribution to the Paper	co-planned experiments; co-conducted lab work; edited manuscript (40%)		
Signature		Date	29.06.2019

Name of Co-Author	*Bernd Degen*		
Contribution to the Paper	obtained funding for the research; supervised experiments; edited manuscript (10%)		
Signature		Date	13.07.2019

Name of Co-Author	*Andrew J Lowe*
Contribution to the Paper	as principal supervisor; obtained funding for the research; supervised experiments; advised on manuscript structure; edited manuscript (10%)

Signature		Date	10.07.2019

BOTAB DNA extraction protocol: optimisation and comparison study

Abstract

DNA can be used in a range of identification methods and is often used in wildlife and forest forensics to identify trade in protected species. However, despite illegal logging representing the world's third largest transnational illicit trade, the application of genetic methods to identify the species and origin (geographical region) of timber products has been limited. Typically because routine extraction protocols are unsuitable for working with timber. The success of commercial DNA extraction kits when used on timber material can potentially be enhanced by incorporating appropriate modifications. Alternatively, in-house extraction protocols can be developed. One such in-house method, has been developed in collaboration between our laboratories in Australia and Germany. Known as the BOTAB protocol it is based on a CTAB protocol. The current study was conducted as part of the patenting process and also allowed further optimisation of the BOTAB protocol and comparison with other commercial DNA extraction kits. We found that the quantity and quality of DNA extracted using the BOTAB protocol is comparable to that obtained from commercial kits, which is a key forensic validation finding. Through this work, we also found additional reagents, not previously identified in the patent, which could be used as an alternative for some existing extraction buffer reagents. Similarly, additional temperature, time or reagent concentration conditions of the extraction buffer were found to be viable for inclusion in the patent. This work led to an update of the wording of the patent document that better reflects the optimised BOTAB protocol. Finally, the BOTAB protocol was updated to incorporate a variation to the procedure. The substitute steps – an overnight incubation of the extraction buffer, followed by a short (<1 hour) -80^{o}C incubation of the precipitation solution – provide additional options for laboratory timetabling without impacting on time or capacity constraints. The outcome of this study is an increased understanding of parameters that effect successful DNA extraction from timber material, the result of which will be a greater capacity to utilise genetic analysis for the identification of the origin (species and region) of timber products.

Key Words: Timber tracking, Experimental design, DNA extraction, Degraded tissue, Illegal logging

Introduction

Genetic identification methods have been proposed as a mechanism to support the legal frameworks that control and monitor the trade of timber species (Dormontt et al. 2015). Genetic methods are well suited to this end, as they have the capacity to provide information at multiple levels (e.g. species name, geographic origin and match (individualisation) testing) (Lowe & Cross 2011). For these methods to be applied, the extraction of usable DNA from timber material is a fundamental prerequisite (Verbylaite et al. 2010). Most plant DNA extraction protocols are based on either the CTAB (Murray & Thompson 1980; Doyle & Doyle 1987) or Alkaline (i.e. SDS) (Dellaporta et al. 1983; Fang et al. 1992) methods (Varma et al. 2007; Tan & Yiap 2009; Verbylaite et al. 2010). These in-house protocols have evolved into some commercially available plant DNA extraction kits, such as the DNeasy Plant Mini Kit (QIAGEN; Hilden, Germany), the NucleoSpin® Plant II (MACHEREY-NAGEL; Düren, Germany) and the innuPREP Plant DNA (ANALYTIK JENA; Jena, Germany). By providing a product that is easy to use, at a cost-effective price, these kit-based protocols have revolutionised the field of genetic analysis (Varma et al. 2007; Verbylaite et al. 2010). Nevertheless, in-house extraction methods remain an important procedure for many DNA labs due to their flexibility and transparency regarding reagents, and often lower consumable cost. However, they tend to be more labour intensive, taking longer on average than kit-based offerings.

Many of the commercially available kits have been designed to work with fresh tissue, which presents several complications when attempting to extract from timber material. Firstly, as the quantities of DNA in timber is typically much less than that found in fresh leaf material, a greater amount of wood tissue is usually required to generate sufficient template DNA. If standard weights are used as per the protocols, then DNA yields may be insufficient for use in genetic analyses (Tang et al. 2011). DNA yield can be improved by using an increased amount of tissue, however, in most cases the extraction buffer volumes will need to be increased also (Deguilloux et al. 2003; Rachmayanti et al. 2006),

or else the mixture becomes too viscous to be used in proceeding steps. Alternatively, multiple extractions can be performed on a sample and pooled together in one of the final steps (Speirs et al. 2009). But this increases the cost of the extraction, as multiple sets of reagents and tubes are needed.

DNA yield may also be suboptimal because the incubation steps can be brief (e.g. for extraction buffer or precipitation solution steps) or non-existent (e.g. elution step) (Drábková et al. 2002). Appropriate incubation parameters, such as time and temperature, allow for the optimal interaction between the buffer/solution and tissue. Therefore, steps with these actions should be extended (or initiated if absent) to facilitate the highest probability of obtaining DNA from timber (Rachmayanti et al. 2006; Finkeldey et al. 2010). Another problem with kits is that the elution (also known as the re-suspension) volumes are usually too large for timber samples, diluting the concentration of the extracted DNA even further as a consequence. This setback can be resolved by decreasing the elution volume (Rachmayanti et al. 2006). Despite their impact, all of these limitations can be rectified without having to alter the protocol too drastically.

In some instances, even with appropriate modifications, DNA extractions may still not work. This can be attributed to inhibitory chemicals or compounds such as polyphenols (e.g. tannins), polysaccharides, proteins and other secondary metabolites. While useful for the plant when alive, they are unfortunately co-extracted with the DNA which inhibits downstream genetic analysis (Porebski et al. 1997; Tan & Yiap 2009). Additionally, the processing and treatment of timber material may further degrade the DNA and incorporate more inhibitory chemicals that prevent a successful DNA extraction (Rachmayanti et al. 2009; Tnah et. al 2012). In order to facilitate the successful use of DNA extracted from timber, these chemicals must be neutralised, separated and/or removed from DNA during the extraction process (Varma et al. 2007).

An important component of any DNA extraction protocol is the extraction buffer (also known as the lysis buffer). Specifically, the reagents incorporated within the buffer that aim to limit, neutralise and counteract the effect of inhibitors, as well as maximising DNA yield (Varma et al. 2007; Tan & Yiap 2009). While steps incorporating

precipitation and wash solutions are useful, their role is to separate the DNA from any inhibitor chemicals and compounds once they have been neutralised by the extraction buffer reagents (Varma et al. 2007; Tan & Yiap 2009). Although the exact reagents used in an extraction buffer may differ between protocols, similar types of reagents are nearly always used, as each performs a particular function, which can be categorised into broad functional groups. For a description of these reagent functional groups and some common reagents for each group, see Table S1.

In most instances, for various proprietary or copyright reasons, the exact reagents and concentrations used in the extraction buffers of commercially available kits are not known. This means they cannot be readily modified. Whereas in-house protocols are transparent, and reagents and their concentrations are known and can be easily adjusted, making them more suitable to working with difficult samples such as timber (Verbylaite et al. 2010). For in-house protocols developed for timber and many other degraded tissues, the most widely used detergent in the extraction buffer is CTAB (De Filippis & Magel 1998; Allen et al. 2006; Verbylaite et al. 2010; Jiao et al. 2014). In our laboratories we have been working on optimising the CTAB protocol for use on timber samples. Known as the BOTAB extraction protocol, it is modified by the inclusion of Boric acid in the extraction buffer (Wolfe et al. 2010). Boric acid assists with the removal of carbohydrate impurities and forms complexes with polyphenols (Manning 1991; Linder et al. 2000).

Recently the BOTAB protocol, alongside the application of the extracted DNA in downstream methods for genetic identification, was recognised as a unique commercial product, and was patented (Lowe et al. 2015) (see Thesis Appendix V for the complete patent document). Under the rules of intellectual property, non-commercial use of the protocol is still allowed, so the protocol is available to be used and modified by other researchers. The patent is primarily framed around the extraction buffer, focusing on two areas; firstly, around the types (functional groups) and concentrations of reagents included in the buffer; and secondly, around the time and temperature of the incubation requirements for the extraction buffer. Within each of these areas the current chapter outlines the standard (reagent type and concentration, incubation time and temperature) as well as identification of suitable alternatives. As part of the patent

development, the patent agents (Wrays, Perth, Australia) produced an initial patent document.

The aim of this chapter is to assess if the variations and alternatives presented were sufficient, and to identify if any additional material needed to be included in the final patent document.

The BOTAB protocol has also been used for the forensic identification of timber to support legal proceedings (Dormontt et.al. 2020). As such, certain aspects of the protocol required validation (for example: repeatability, sensitivity, universal application, comparison to other existing protocols), to verify that the results generated are reliable (SWFS 2018; SWGDAM 2016). This chapter provides supporting evidence towards the forensic validation of the BOTAB protocol.

This study also includes additional experiments for the benefit of optimising the protocol. While there were no validation or patent assessment requirements for those experiments, results from these in-house experiments have been able to advance the BOTAB protocol and to gain a greater understanding of the flexibility of the protocol. However, it is important to point out that these optimisations are for internal use only, and a correct validation of any alterations would need to be conducted before they could be incorporated into any forensic casework studies.

Materials and Methods

This section contains a brief summary of the materials and methods used in this study. Further information for many sections is available (and referred to) within the Supplementary Information (hereafter SI) (see Chapter appendix I).

For this study, a range of components of the BOTAB extraction protocol were tested. There are ten experimental sections to this chapter, with each experiment undertaken for one of three requirements; patent development, forensic validation, or in-house optimisation. Nine of the experiments (*Experiments I-IX*) investigate how variation to

the protocol affects DNA extraction success. The final experiment (*Experiment X*) is a comparison of the updated BOTAB protocol against several commercial kits, which is an important requirement for the forensic validation of the BOTAB protocol (SWFS 2018; SWGDAM 2016).

For a summary of all ten experiments with information about each including: justification (patent development, forensic validation or in-house optimisation), the variables, timber samples and the parameters used to assess the results see Table 1. For more detailed information on each of the experiments, see SI (*Methods: Experiments I-X*).

Table 1: *overview of the experiments conducted for this study*

About experiments			Additional information				Experiment details				
#	title	overview	Experiment type	focal area	part of BOTAB protocol used in experiment (*variation experiments only*)[1] section	step #	description of experiment[2]	Samples used[3]	Negatives used	measure of success[4]	Table/figure of results
I	*Tissue weight*	assess the validity of the current starting tissue weight	variation	Forensic validation	1	1	Compare the results of BOTAB extractions undertaken with standard tissue weight (100 mg) to those of three alternatives (200 mg, 50 mg and 10 mg) ▪ NB: 10 mg alternative tissue weight utilised only one tube compared to other weights that use two tubes	▪ jarrah ▪ English oak ▪ mahogany[C]	Yes	▪ DNA concentration ▪ DNA purity ratios	▪ Figure S1 ▪ Table S6
II	*Tissue grinding*	determine if more suitable grinding parameters can be identified	variation	In-house optimisation	1	3	Vary the parameters used to grind different tissue types (either lathe or scalpel prepared) including: beads (number, size, ratio), grinding (speed, number of cycles, duration) ▪ See *Table S7* for description on the grinding parameters used	▪ merbau ▪ European white oak	No	▪ visual[5]	▪ Table S7 & S8
III	*BOTAB extraction buffer reagent*	assess if the original extraction buffer reagents and concentration are adequate	variation	Patent development	1	4	Multiple experiments: • Compare samples where the standard reagent is used to when removed. • Assess if the concentration of the current reagents is important by comparing to two additional concentrations • Compare standard reagents to alternatives using three different concentrations ▪ See *Table S3* for a list of standard and alternative reagents tested; and *Table S4* for the three concentrations used for each reagent	▪ English oak ▪ mahogany[C] ▪ merbau ▪ zebrawood ▪ wenge ▪ Baltic pine	Yes	▪ DNA concentration ▪ DNA purity ratios ▪ statistical significance (for some tests only)	▪ Figure S2-S5 ▪ Table S9-S14
IV	*Additional extraction buffer NaCl (salt) concentration*	further testing focused on the extraction buffer salt concentration	variation	Patent development & In-house optimisation	1	4	In addition to the work from *Experiment III*, a broader assessment of the NaCl (salt) concentration	▪ merbau ▪ European white oak	No	▪ DNA purity ratios ▪ agarose gel results	▪ Table S15 & S16
V	*Additional extraction buffer antioxidant*	additional experiments on alternative extraction buffer reagents	variation	Internal optimisation	1	4	Following on from the work in experiment III, this experiment was designed to further test the suitability of the alternative antioxidant reagents acetic acid and sodium acetate. Tests were done using a broader range of concentrations, and also compared to extractions using standard antioxidant reagents (DTT and βME)	▪ mahogany[C & J] ▪ merbau ▪ European white oak ▪ larch	No	▪ DNA purity ratios ▪ agarose gel results	▪ Table S17-20
VI	*Extraction buffer incubation time*	identify suitable alterative extraction buffer incubation times	variation	Patent development & In-house optimisation	1	5	Undertake extractions using a wide range of extraction buffer incubation times from 1hour through to 3 days (~72 hours)	▪ English oak ▪ mahogany[C] ▪ merbau ▪ American white oak	No	▪ agarose gel results	▪ Table S21
VII	*Extraction buffer incubation temperature*	assess the suitability of alternative extraction buffer incubation temperatures	variation	Patent development	1	5	Incubate the extraction buffer at three alternative temperatures: Room temp (~25°C), 45°C and 65°C	▪ Mahogany[C] ▪ merbau ▪ white oak ▪ American white oak	No	▪ DNA purity ratios ▪ agarose gel results	▪ Table S22
VIII	*Precipitation solution salt*	testing of alternative precipitation solution salt reagents	variation	In-house optimisation	2	9	Perform extractions using three alternative precipitation solution salts (ammonium acetate, potassium acetate and lithium chloride). Each salt was tested at three concentrations ▪ See *Table S5* for further information on the salts and the concentrations in which they were tested at	▪ merbau ▪ European white oak	No	▪ DNA purity ratios ▪ agarose gel results	▪ Table S23 & 24
IX	*Precipitation Solution incubation time and temperature*	investigate the potential incorporation of alternative precipitation solution incubation conditions	variation	In-house optimisation	2	11	Compare incubations at two temperatures -20°C (standard) and -80°C, using a range of incubation times (30mins – overnight (~16 hours))	▪ mahogany[C] ▪ merbau	No	▪ agarose gel results	▪ Table S25
X	*Comparison of BOTAB protocol to commercial kits*	compare the BOTAB extraction protocol to commercial kits	comparison	Forensic validation	-	-	Compare the BOTAB protocol (with any updates acquired from these experiments) to three commercial extraction kits. ▪ NB the kits were modified to suit timber material ▪ See SI (*Methods: Experiment X*) for a description of the protocol modifications for each kit	▪ jarrah ▪ English oak ▪ mahogany[C] ▪ wenge ▪ Baltic pine ▪ bigleaf maple	Yes	▪ DNA concentration ▪ DNA purity ratios	▪ Figure 1 ▪ Table S26

NB: [1]see chapter appendix II for entire BOTAB DNA extraction protocol; [2]see SI (methods) for full description of each experiment; [3]letters next to mahogany samples (either C or J) are the unique identifier, see Table S2 for more information; [4]see SI (Methods) for information on measuring results; [5]see table S7 for scoring metric

This study was conducted using only timber material. See SI (*Methods: timber material*) for details regarding the timber samples used in the experiments, a summary of which can also be found in Table S2. The outer surface of each timber sample was UV treated to neutralise surface contaminants. Tissue used for extractions was collected using either a lathe or a scalpel. For further information on the decontamination and sample preparation techniques see SI (*Methods: Tissue preparation, storage and use).*

Unless otherwise stated DNA extractions were undertaken using the standard BOTAB protocol (see appendix II). For each experiment, only the component of the protocol being tested was varied, with all other steps remaining as standard (see Table 1 for description of how each experiment varies the protocol). Unless stated otherwise, extractions were conducted with a five hour incubation of the extraction buffer and overnight incubation of the precipitation solution at -20°C (referred to hereafter as the BOTAB day variant). The alternative to this protocol is the BOTAB overnight variant, which consists of an overnight incubation of the extraction buffer and a short (<1 hour) incubation of the precipitation solution at -80°C.

Negative control DNA extractions were performed for select experiments only. They are a requirement for the two forensic validation experiments (*Experiments I & X*) as per (SWFS 2018; SWGDAM 2016). See the SI (*Methods*) for each experiment for information regarding the inclusion of negative extractions.

To measure the resulting extractions, several strategies were used, DNA concentration and/or purity ratios were recorded. For further information regarding the details and equipment used for concentration and purity reading see SI (*Methods*). For certain experiments, species-specific microsatellite markers were utilised, and DNA purity was assessed via PCR amplification success using agarose gels (see SI (*Methods: Scoring amplification*) for more information). Finally, for several experiments, a non-parametric Friedman test was used to detect treatment differences. For further information regarding the statistical analyses see SI (*Methods: Statistical analysis*). Statistical tests were

only conducted on experiments where six timber samples and at least three variables were incorporated. Summaries of each experiment and application of Friedman tests can be found in *SI (Methods)*.

Results and Discussion

This section contains a summary of the results for each experiment. For more detailed results of each experiment see SI (Results).

This study was broken into ten experiments, with each conducted to fulfil the requirements for one of three areas: forensic validation, patent development or in-house optimisation. Two of the experiments (*Experiments I & X*) are framed around the forensic validation of the BOTAB protocol (SWFS 2018; SWGDAM 2016). The most important of these experiments was the comparison of the BOTAB protocol to three commercial kits (*Experiment X*). The median DNA concentrations for all four extraction protocols was low ($0.036\ ng/\mu l$–$0.20\ ng/\mu l$). Despite this, the median DNA concentration of the BOTAB protocol was the highest of the four ($0.20_{ng/\mu l}$).

The median A_{260}/A_{280} DNA purity ratio of the BOTAB protocol extractions (1.95) was also higher than the three commercial kits (<0.58). Yet the median A_{260}/A_{230} DNA purity ratio were lower (<0.03) than any of the kits (0.639 (JENA), 0.898 (MO-BIO), 1.475 (QIAGEN). For each of the protocols, it appears that there is a trade-off between yield and purity. Yet, it was not the intention of this experiment to identify the best extraction protocol, but to identify that similar results can be obtained when extracting DNA using the BOTAB protocol compared to the commercial kits. A boxplot of the DNA Concentration and purity for each extraction technique can be seen below in Figure 1.

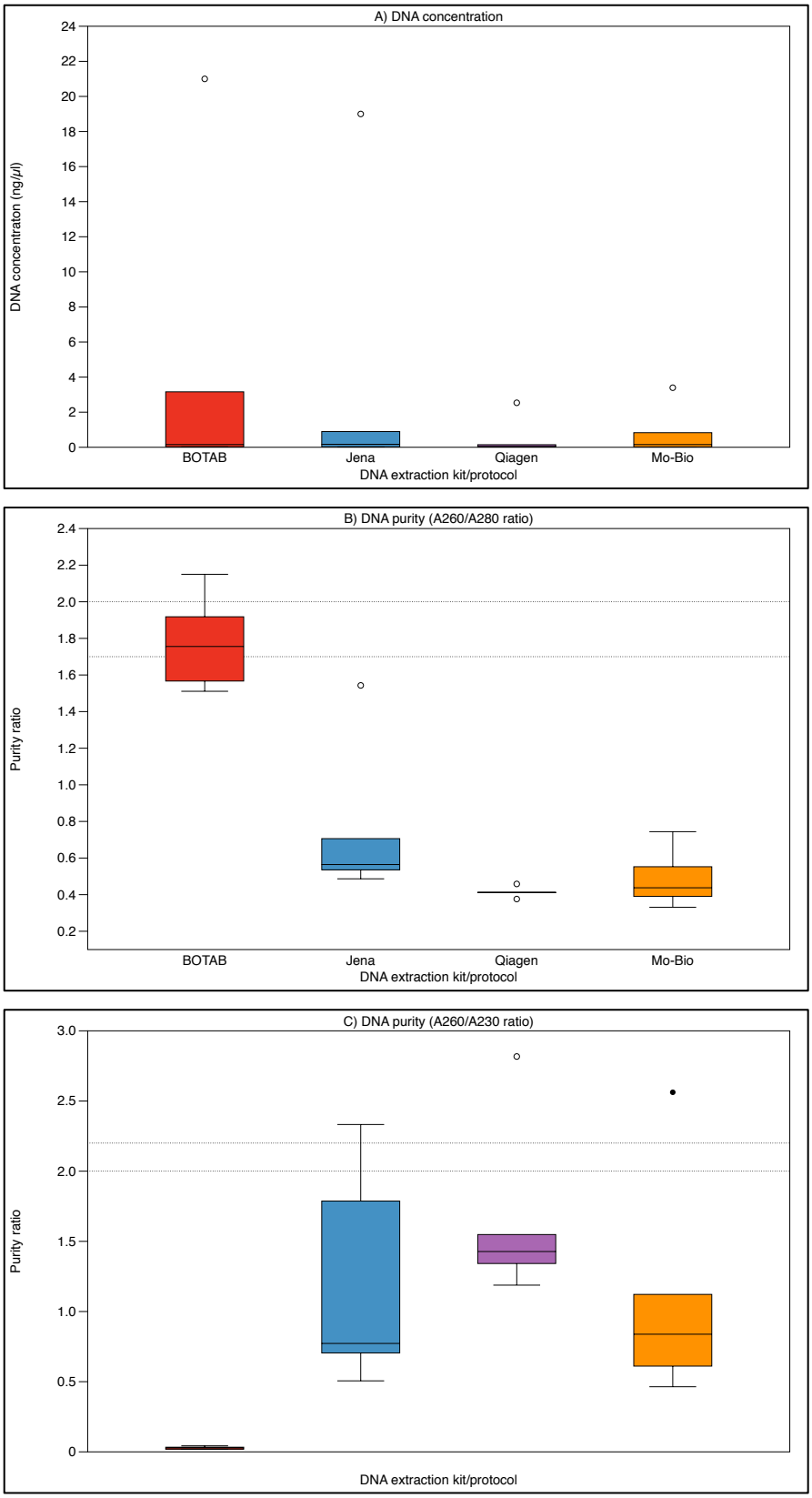


Figure 1: *Results from DNA extraction protocol comparison study (Experiment X)*. Boxplot graphs of DNA concentration (A) and DNA purity ratios (B & C) results for two BOTAB extractions (Initial and Final) and three commercial extraction kits (JENA, QIAGEN & MO-BIO). *NB: filled circles signify outlier samples, open circles signify extreme outliers (that are three times the IQR).*

Regardless of small differences in DNA quantity and quality, usable DNA was extracted from timber samples using all three commercial kits, as well as the in-house BOTAB protocol. This result demonstrates that by using appropriate modifications designed for timber material, commercial kits can be optimised for timber templates with a high degree of reliability (Telle & Thines 2008; Verbylaite et al. 2010; Särkinen et al. 2012).

Commercial DNA extraction kits have come a long way in recent years and are being designed to extract DNA from increasingly difficult material. In fact, kits (with their respective modifications as described in this chapter) have become the go-to extraction protocol for many timber extractions undertaken in our facility (see thesis appendix II); predominantly due to their ease of use and shorter time requirements (several hours) compared to the BOTAB protocol (two days). Additionally, the BOTAB protocol is more complex. As such it is only used certain circumstances, e.g. when commercial kit extractions don't work and when trying to extract from difficult (degraded) samples, or for casework samples.

The other forensic validation experiment was to determine the impact that starting tissue weight had on DNA concentration and purity ratios (*Experiment I*). We found that the current tissue weight (~100 mg) is sufficient to generate a usable result; there was no clear drop off in yield when a lesser amount of tissue was used. Despite the higher (200 mg) tissue weight having the highest median DNA yield ($2.30_{ng/\mu l}$), the viscosity of the tissue/extraction buffer mixture made it more difficult to handle. Additionally, unless the material is generated using a lathe, the time requirements for generating such volume of tissue may be prohibitory.

The second focal area of this study was conducting experiments to support the development of the patent (*Experiments III - VII*). While the patent encompasses the entire BOTAB protocol, it is the extraction buffer and how it is utilised, where the wording of the patent document is most articulate. This is why there were a number of experiments focused on the extraction buffer. In particular the

reagents used in the buffer (*Experiment III*) and the concentration (*Experiments III – V*) as well as the incubation time (*Experiment VI*) and temperature at which the buffer is used (*Experiment VII*).

The importance of DNA concentration of the extraction buffer reagents were assessed in four reagents (*DTT, Boric Acid, CTAB and Proteinase K*). In all instances except one (varying proteinase K), There was a significant difference ($p < 0.05$) in DNA concentration when the reagent concentrations were either altered (higher or lower than standard) or removed. For all four reagents, the DNA concentration from timber samples when using the standard concentration of each reagent were the lowest and the A_{260}/A_{280} DNA purity ratios of the standard reagent concentrations were the highest. Post-hoc tests revealed that there was a significant difference between both of the alternative reagent concentrations to the standard (see Tables S9-S11 (DNA concentration and purity scores; Figures S2-S5 (the four reagents where statistical tests were performed)).

This result agrees with the comparison testing (*Experiment X*) findings, where by a trade-off between DNA yield and DNA quality occurs. Yet, without any PCR amplification it is difficult to determine which is more important, i.e. which has the least/greatest effect on DNA amplification success. Based on the results from this experiment, all of the reagents used in the BOTAB buffer will continue to be included. Furthermore, apart from the concentration of the NaCl (the salt reagent), the current concentrations of the reagents will not be altered, as there was insufficient evidence to justify changing them. While a universal concentration of NaCl was unable to be conclusively determined (from *experiment IV*), previous work has identified that the optimum concentration can be species specific (Varma et al. 2007). We recommend that further work should seek to explore how best to use this information, perhaps by determining if any taxonomic or chemical inhibitor correlations can be identified, with the aim of either further optimisations to the protocol universally, or capacity to tailor extractions to particular samples/species.

When comparing alternative extraction buffer reagents to those currently used in BOTAB protocol (*Experiments III & VI*), similar results were obtained for some of the reagents (Ascorbic acid, sodium ascorbate). For these reagents, they were considered suitable alternatives to the original reagent, and added to the final patent document if not already included (see Table S3 for novel reagents). Regarding the two "safe" alternative antioxidant reagents (Ascorbic acid and Sodium ascorbate), further work to ascertain the application of these alternative reagents more routinely in the BOTAB protocol should be undertaken in the future.

The only other alternative reagents that could be worth investigating further are the PVP excipient reagents. PVP is categorised by Molecular Weight ($_{MW}$). Alongside the standard reagent (PVP-K40) ($40{,}000_{MW}$), the high molecular weight PVP360 ($360{,}000_{MW}$) was tested in this study. We detected no clear difference in the median DNA concentration or purity between the two reagents, indicating that molecular weight may not be important when selecting the best excipient reagent. To support this finding, both reagents have been used previously in timber and degraded material DNA extraction buffers (Colpaert et al. 2005; Speirs et al. 2009). Furthermore, some publications have not disclosed the molecular weight of the PVP (Rachmayanti et al. 2009; Jiao et al. 2012; Jiao et al. 2014), or suggest using different molecular weights for various troubleshooting approaches in difficult samples (Allen et al. 2006). These inconsistencies create uncertainty and confusion as to the best molecular weight of PVP to use in extractions. Further experiments regarding the impact (or lack of) that molecular weight of the PVP reagent has on extraction success are recommended. This should include additional molecular weights of PVP, such as the low molecular weight PVP10 ($10{,}000_{MW}$) which one study found to be the better PVP excipient reagent (Puchooa & Khoyratty 2004).

The two other sections of the BOTAB patent that required particular focus were the incubation (time and temperature) requirements of the extraction buffer. For the temperature tests (*Experiment VI*), the alternative incubation temperatures (Room Temperature (~25^{o}C), 40^{o}C or 64^{o}C) generated DNA of reasonable

quality. The tests identified that there was no detectable difference between the temperatures, suggesting that any of them might be suitable for incubating the extraction buffer. However, because of capacity restraints (limited number of heat incubators), this experiment did not include the current incubation temperature (55^{o}C), so further comparisons are required before justifying the amendment of the incubation temperature in the BOTAB protocol.

For the extraction buffer incubation time test (*Experiment VII*), there were insufficient results to draw robust conclusions, so the existing incubation time (5 hours) will remain. Yet, further work regarding suitable incubation times should be considered. Because these time/temperature sensitive steps require specialist equipment, any alterations to the incubation time may assist in improving the capacity of the laboratory.

Shorter times have been used by others successfully in CTAB extractions (e.g. 30 minutes (Allen et al. 2006), 30-120 minutes (Verbylaite et al. 2010), 60 minutes (Wolfe et al. 2010; Särkinen et al. 2012). These short time options could free up equipment, whilst longer times would enable equipment to run during quiet times in labs (e.g. overnight). Overnight incubations have been successfully used for BOTAB protocol extractions by both the Thünen Institute and University of Adelaide previously. It has been employed as a dovetail protocol that can be used alongside an extraction using the standard incubation time to maximise the efficiency of existing lab equipment. This time period was not listed in the initial patent document, and despite not being able to generate a significant result, was included in the final document. Additionally, the overnight extraction buffer incubation time was incorporated into the updated extraction protocol as a permanent option (see further in discussion for more information).

The final focal area of this study was for in-house optimisation of the protocol only (not updates to the patent). There were three experiments in this section, two focusing on the precipitation solution of the protocol (*Experiments VIII & IX*), and one looking at the sample tissue grinding parameters (*Experiment II*).

Results from the grinding test (*Experiment II*) identified an effective grinding strategy that was different to that which had been used previously. To ensure the highest proportion of tissue is homogenised there are several measures that can be employed. Firstly, the initial shavings generated should be as fine as possible, and the easiest way of doing this is by using a lathe. If this is not available, then a scalpel can be used instead, ensuring that the shavings are as fine as possible. Secondly, for scalpel prepared samples, we found that a combination of small (1.4 mm) and large (2.8 mm) beads was the best for maximising the grinding success. A bead combination was not important for lathe prepared samples, presumably due to the fineness of the shavings. Furthermore, lathe prepared samples only needed one or two grinding cycles to successfully grind the tissue, whereas scalpel prepared samples required up to four cycles to maximise the grinding success. After this number of cycles there was no detectable change in the grinding outcome. Finally, it was identified that the grinding time doesn't need to be long, as similar results were found between 20 and 60 second grinding cycles.

This was an important finding, as excessive grinding strategies (speed and number of cycles) can negatively impact the DNA yield of a sample because DNA can be sheared/fragmented by the same forces used in the grinding (Varma et al. 2007). To maintain DNA integrity, we concluded that the highest grinding speed should be $4.0ms^{-1}$ and a maximum of four grinding cycles be employed. Nevertheless, for samples where the tissue is not completely ground after four cycles, a result may still be possible. A similar study comparing tissue weight and grinding parameters together (Shepherd et al. 2002), identified that degree of grinding did not impact the yield of extracted DNA. So long as sufficient material was ground then usable DNA could be obtained. While our study did not compare grinding and tissue weight together (they were considered independently), an assumption can still be inferred, that so long as more than 50% of the starting material is ground then a usable yield of DNA can be obtained. Therefore, for samples where the shavings are too thick, or the tissue is quite tough, using a slightly larger tissue weight should be considered to compensate for the reduced proportion of grinding.

The remaining two experiments of this study were an assessment of two components of the precipitation solution. The first (*Experiment VIII*) was to compare three alternative salt reagents (Ammonium acetate, lithium chloride & potassium acetate). For this experiment, we used DNA purity and PCR amplification to determine the success of the alternative reagents & concentrations. DNA purity readings were reported for all samples (see table S23). Limited information could be garnered from the PCR scoring, with only three results (2.1_M Ammonium acetate & 1_M Lithium Chloride for merbau, 0.8_M for European white oak). These markers have been used successfully before on these samples in other experiments (see experiments II & IV-VI) where the standard reagent (Sodium acetate) was used as the precipitation salt. We chose not to include Sodium acetate in this test as this was only a pilot experiment, with no patent or forensic requirements. In hindsight, this restricted the importance of the findings or ability to determine success/failure. Yet, considering the context of the experiment (i.e.to scope if it was worth investigating the alternative reagents), we believe this decision was justified. Further testing on these alternative reagents, alongside Sodium acetate may be worth considering in the future. Additionally, other reagents not yet tested might also be worth incorporating. For example, based on the MSDS documents for both the Qiagen[14] and Jena[15] kits, acetic acid is used as a precipitation salt and could present a viable alternative in the BOTAB protocol.

The aim of the second precipitation stage experiment (*Experiment IX*), was to assess the suitability of alternative incubation conditions. The results from these experiments revealed that similar DNA yields could be generated using an alternative incubation procedure to the original one. The standard procedure is an overnight incubation at -20^{o}C, and this test found comparable results when samples were incubated at -80^{o}C for ~1 hour. These alternative incubation conditions tie in well with the other alterations to the standard protocol, namely an overnight incubation of the extraction buffer. Traditionally, the overnight

[14]https://sds.qiagen.com/ehswww/QIAGENwww/result/report.jsp?P_LANGU=E&P_SYS=4&P_SSN=191698&P_REP=00000000000000000082&P_RES=325785
[15]www.analytik-jena.de/fileadmin/content/pdf_life_science/MSDS/MSDS_innuPREP_Plant_DNA_Kit_en.pdf

incubation of the extraction buffer was run alongside the standard overnight incubation (at -20°C) of the precipitation solution, the impact of this was that the duration of the protocol was extended to three days. This extra day can be avoided using a short (<1hr) incubation of the precipitation solution at -80°C for any overnight extraction buffer incubated samples. Known as the overnight variant of the BOTAB protocol, it allows extractions to be completed in a similar timeframe to the standard protocol (known as the day variant) (see chapter appendix II for details for the procedure of both variants). Furthermore, as both variants use unique incubation schedules (BOTAB buffer and precipitation solution steps) that do not conflict with each other, the equipment and labour requirements of the two variants of the BOTAB protocol dovetail well and can be used in combination to increase the capacity of the lab if needed.

Finally, in addition to the modifications to the BOTAB protocol that were determined via the designated experiments, the protocol was also updated as an indirect result of the sheer volume of extractions undertaken for this study. For one part, the chloroform centrifuge step (*step 8*), the spin speed was altered, and reduced from 16,000 G to 10,000 G. When centrifuging at the lower speed, samples were less likely to generate a mucous aqueous phase, which was more typically formed at the higher speeds in the same sample. The chloroform mixing and centrifuge steps were deliberately not experimented on, because of the toxicity of chloroform. However, if any novel experiments into the extraction protocol were to be undertaken, it would be important to consider these steps, especially in attempts to remove the need for chloroform entirely.

Conclusion

The work undertaken for this study satisfied requirements in three focal areas (forensic validation, patent development or in-house optimisation). We were able to demonstrate that the BOTAB protocol performs as well, if not better, than commercial kits, which is a key metric for its forensic validation. The document developed for the patent was robust with only minor changes being needed to be incorporated (addition of buffer incubation times and reagents not previously

listed). Finally, the BOTAB protocol was optimised, by utilising an overnight incubation of the BOTAB buffer alongside a short (~1 hour) incubation of the precipitation solution at -80°C, a combination that dovetails with the existing procedure to increase the capacity of a facility.

This study has helped to demonstrate that regardless of the protocol used, DNA extraction from timber material is becoming more routine. This can be attributed to the advances in many of the DNA extraction protocols (including the BOTAB protocol) in recent times. Additionally, a greater understanding of how best to adapt existing protocols to make them more suitable to timber material is being gained. Further innovations and advancements in timber DNA extraction methods in the future will hopefully strengthen their future usability. Having reliable DNA extraction protocols means that the barrier to using genetic methods for timber sample identification as part of legal requirements is diminishing. Alongside other scientific techniques, genetic analysis has the capacity to support laws designed to monitor and control the trade of timber. The more that these laws can be enforced to lessen the impact of illegal logging activities, the better.

Acknowledgements

Funding: University of Adelaide; Commercial Accelerator Scheme (CAS), Thünen institute for forest Genetics, International Tropical Timber Organisation (ITTO)

References

[SWGDAM] Scientific Working Group for DNA Analysis Methods (2016) SWGDAM Validation Guidelines for DNA Analysis Methods. https://docswixstaticcom/ugd/4344b0_813b241e8944497e99b9c45b163b76bdpdf.

[SWFS] Society for Wildllife Forensic Sciences Technical Working Group (2018) Standards and Guidelines for Wildlife Forensic Analysis, Version 3. Ed. Lucy M.I. Webster. Published by the Society for Wildlife Forensic Science, 19th November 2018, pp.21. https://www.wildlifeforensicscience.org/wp-content/uploads/2018/11/SWFS-Standards-and-Guidelines_Version-3_19-11-18.pdf

Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nature protocols, 1, 2320.
Colpaert N, Cavers S, Bandou E, Caron H, Gheysen G, Lowe A (2005) Sampling tissue for DNA analysis of trees: trunk cambium as an alternative to canopy leaves. Silvae Genetica, 54, 265-269.
De Filippis L, Magel E (1998) Differences in genomic DNA extracted from bark and from wood of different zones in Robinia trees using RAPD-PCR. Trees, 12, 377-384.
Deguilloux MF, Pemonge MH, Bertel L, Kremer A, Petit RJ (2003) Checking the geographical origin of oak wood: molecular and statistical tools. Molecular Ecology, 12, 1629-1636.
Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant molecular Biology Reporter, 1, 19-21.
Dormontt EE, Boner M, Braun B, Breulmann G, Degen B, Espinoza E, Gardner S, Guillery P, Hermanson JC, Koch G, Leek SL, Kanashirol M, Rimbawantom A, Thomas D, Wiedenhoeft AC, Yin Y, Zahnenq J, Lowe AJ (2015) Forensic timber identification: It's time to integrate disciplines to combat illegal logging. Biological Conservation, 191, 790-798.
Dormontt EE, Jardine DI, van Dijk K-J, Dunker BF, Dixon RRM, Hipkins VD, Tobe S, Linacre A, Lowe AJ (2020) Forensic validation of a SNP and INDEL panel for individualisation of timber from bigleaf maple (Acer macrophyllum Pursch). Forensic Science International: Genetics 46: e102252 doi: 10.1016/j.fsigen.2020.102252
Doyle JJ, Doyle J (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 19, 11-15.
Drábková L, Kirschner J, Vlček Ĉ (2002) Comparison of seven DNA extraction and amplification protocols in historical herbarium specimens of Juncaceae. Plant Molecular Biology Reporter, 20, 161-175.
Fang G, Hammar S, Grumet R (1992) A quick and inexpensive method for removing polysaccharides from plant genomic DNA. Biotechniques, 13, 52-54, 56.
Finkeldey R, Leinemann L, Gailing O (2010) Molecular genetic tools to infer the origin of forest plants and wood. Applied Microbiology and Biotechnolgy, 85, 1251-1258.
Jiao L, Yin Y, Cheng Y, Jiang X (2014) DNA barcoding for identification of the endangered species Aquilaria sinensis: comparison of data from heated or aged wood samples. Holzforschung, 68, 487-494.
Jiao L, Yin Y, Xiao F, Sun Q, Song K, Jiang X (2012) Comparative analysis of two DNA extraction protocols from fresh and dried wood of *Cunninghamia lanceolata* (*Taxodiaceae*). IAWA Journal, 33, 441-456.
Linder C, Moore L, Jackson R (2000) A universal molecular method for identifying underground plant parts to species. Molecular Ecology, 9, 1549-1559.
Lowe AJ, Cross HB (2011) The application of DNA methods to timber tracking and origin verification. IAWA Journal, 32, 251.
Lowe AJ, Jardine DI, Cross HB, Degen B, Schindler L, Höltken AM (2015) A method of extracting plant nucleic acids from lignified plant tissue.

Manning K (1991) Isolation of nucleic acids from plants by differential solvent precipitation. Analytical biochemistry, 195, 45-50.
Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic acids research, 8, 4321-4326.
Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant molecular Biology Reporter, 15, 8-15.
Puchooa D, Khoyratty S-USS (2004) Genomic DNA extraction from Victoria amazonica. Plant Molecular Biology Reporter, 22, 195a.
Rachmayanti Y, Leinemann L, Gailing O, Finkeldey R (2006) Extraction, amplification and characterization of wood DNA from Dipterocarpaceae. Plant Molecular Biology Reporter, 24, 45-55.
Rachmayanti Y, Leinemann L, Gailing O, Finkeldey R (2009) DNA from processed and unprocessed wood: factors influencing the isolation success. Forensic Science International: Genetics, 3, 185-192.
Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT (2012) How to open the treasure chest? Optimising DNA extraction from herbarium specimens. PloS ONE, 7, e43808.
Shepherd M, Cross M, Stokoe RL, Scott LJ, Jones ME (2002) High-throughput DNA extraction from forest trees. Plant Molecular Biology Reporter, 20, 425-425.
Speirs AK, McConnachie G, Lowe A (2009) Chloroplast DNA from 16th century waterlogged oak in a marine environment: initial steps in sourcing the Mary Rose timbers. Archaeological science under a microscope: studies in residue and ancient DNA analysis in honour of Thomas H Loy. p.^pp. 175-189.
Tan SC, Yiap BC (2009) DNA, RNA, and protein extraction: the past and the present. Journal of Biomedicine and Biotechnology, 2009.
Tang X, Zhao G, Ping L (2011) Wood identification with PCR targeting noncoding chloroplast DNA. Plant Molecular Biology, 77, 609-617.
Telle S, Thines M (2008) Amplification of *cox2* (~ 620 bp) from 2 mg of up to 129 years old herbarium specimens, comparing 19 extraction methods and 15 polymerases. PLoS ONE, 3, e3584.
Tnah LH, Lee SL, Ng KKS, Bhassu S, Othman RY (2012) DNA extraction from dry wood of Neobalanocarpus heimii (Dipterocarpaceae) for forensic DNA profiling and timber tracking Wood Science and Technology 46:813-825 doi:10.1007/s00226-011-0447-6
Varma A, Padh H, Shrivastava N (2007) Plant genomic DNA isolation: an art or a science. Biotechnology Journal, 2, 386-392.
Verbylaite R, Beisys P, Rimas V, Kuusiene S (2010) Comparison of Ten DNA Extraction Protocols from Wood of European Aspen (Populus tremula L.). Baltic Forestry, 16, 35-42.
Wolfe BE, Richard F, Cross HB, Pringle A (2010) Distribution and abundance of the introduced ectomycorrhizal fungus Amanita phalloides in North America. New Phytologist, 185, 803-816.

Appendix for Chapter 2

Appendix 1: Supplementary Information

Supplementary Table 1: *Common DNA extraction buffer reagents and their function*. Description of the reagents contained within a typical DNA extraction buffer that can be grouped into broad functional groups

Functional Group	Example Reagents	Functional purpose of group
Detergent	CTAB SDS	• Disruption of cells or tissue and dissolve cellular membrane • Inactivation of nucleases (DNase/RNase) • Denaturation of nucleoprotein complexes • Assist in the removal of contaminants • Collate and emulsify lipids
Buffer	Tris	• Maintain the pH of the solution during the extraction
Chelating Agent	EDTA	• Binding metal ions (e.g. magnesium, calcium, Iron). Preferably, these agents will chelate the ions to reduce their availability to act as co-factors to endogenous nucleases that may cause damage to the extracted DNA • Inactivation of nucleases (DNase/RNase)
Salt	NaCl	• Maintain osmoregularity of the solution and assist in nucleic acid stabilisation • Disruption of cells or tissue and dissolve cellular membrane • Denaturation of nucleoprotein complexes • Assist in the removal of contaminants
Excipient	PVP*	• Allow the components of the solution to react with lignified tissue • enhance precipitation of compounds other than the nucleic acids such as phenolic compounds. • Denaturation of nucleoprotein complexes • Assist in the removal of contaminants
Biological Antioxidant (*Reducing Agent*)	DTT βME	• Reduce the disulphide bonds (or at least nucleic acid) in the solution and/or to reduce the tannins and other polyphenols present on the extraction mixture. • Destroy structural organisation of proteins. • Inactivation of nucleases (DNase/RNase)
Protease	Proteinase K	• Assist in the breakdown of cellular wall material in the lignified plant tissue and inactivate nucleases. • Break peptide bonds. • Denaturation of nucleoprotein complexes • Assist in the removal of contaminants

NB: Table is a summary of the following: BOTAB patent (see Thesis appendix V), Varma (2007) & Tan (2009)*. For reagent abbreviations (except SDS= Sodium dodecyl sulphate) see Table S3.*=PVP is sold in a variety of molecular weights, which can vary between protocols.*

Materials and methods

Timber material

The timber samples were chosen to represent a range of species, geographical locations and forest types. Where possible the samples were morphologically verified by the Thünen Institute of wood research xylarium (Hamburg, Germany). Specific information pertaining to each sample's history (e.g. heartwood or softwood/ fresh aged timber) is unknown. There was particular emphasis placed on the oak specimens because they are an important group to further understand. They are well traded globally and can be passed off as each other during timber importations. Additionally, the group also includes *Quercus molgolica* (Mongolian oak) that is a CITES (appendix III) listed species. At the time of this research the Thünen laboratory was in the process of working on microsatellite markers that could differentiate between the species, so they were readily available to us during my time at the facility. Specimen information, along with which experiments they were used in can be found in Table S2. When referring to the two mahogany samples they will be differentiated by their sample code (see Table S2).

Supplementary Table 2: Information on the timber specimens used in the study

Sample code	Collection code	name		other information				Experiment[3]									
		Scientific	common	Family	Forest type	Native Range[1]	IUCN Status[2]	I	II	III	IV	V	VI	VII	VIII	IX	X
A	DJ_jar_2014	*Eucalyptus marginata*	jarrah	Myrtaceae	Temperate	Au	-	X[4(a)]									X
B	DJ_oak_2014	*Quercus robur*	english oak	Fagaceae	Temperate	EU	LC	X		X[4(b)]							
C	DJ_mah_2014	*Swietenia sp.*	mahogany	Meliaceae	Tropical	SA	VU	X		X		X	X	X		X	X
D	DJ_mer_2014	*Intsia palembanica*	merbau	Caesalpiniaceae	Tropical	SEA	VU		X	X[4(b)]	X	X	X	X	X	X	
E	FGNR_09_2013	*Quercus sp.*	European white oak	Fagaceae	Temperate	NA	-		X		X	X[4(c)]	X	X	X		
F	DJ_zeb_2014	*Microberlinia brazzavillensis*	zebrano	Fabaceae	Tropical	Af	VU			X							
G	DJ_wen_2014	*Millettia laurentii*	wenge	Fabaceae	Tropical	Af	EN			X[4(b)]							X
H	DJ_pin_2014	*Pinus sylvestris*	Baltic pine	Pinaceae	Temperate	EU	LC			X[4(b)]							X
I	FGNR_10_6_2014	*Larix sp.*	larch	Pinaceae	Temperate	EU or NA	LC					X					
J	TN1	*Swietenia sp.*	mahogany	Meliaceae	Tropical	SA	VU					X					
K	T11C	*Quercus sp.*	American white oak	Fagaceae	Temperate	NA	-					X	X	X			
L	LD210073	*Acer macrophyllum*	bigleaf maple	Sapindaceae	Temperate	NA	LC										X

NB: [1]continent (or region) where species is naturally found, shorthand's as follows (Au = Australia, Af = Africa, EU = Europe, NA = North America, SA = South America, SEA = South East Asia. [2]IUCN (international Union for Conservation of Nature) redlist highlighting species closeness to extinction, shorthand's as follows (LC = Least Concern, VU = Vulnerable, EN = Endangered). [3]Experiments where timber sample was used, experiment numbers are as follows (I: tissue weight, II: tissue grinding, III: BOTAB extraction buffer reagents, IV: additional extraction buffer salt (NaCl), V: additional extraction buffer antioxidant reagents, VI: extraction buffer incubation time, VII: extraction buffer incubation temperature, VIII: precipitation solution salt alternatives, IX: precipitation solution incubation time and temperature, X: comparison of BOTAB protocol to commercial kits).[4]some tests from within an experiment are missing or were unable to obtain results:[(a)]No 50mg tissue weight tested; [(b)]not tested for all extraction buffer tests due to time constraints and limited availability of ground tissue; [(c)]samples not DNA Purity tested for reagent Sodium ascorbate.

Tissue preparation, storage and use

The outside surfaces of all samples were UV sterilised then removed with a scalpel to expose clean uncontaminated tissue underneath for sampling. For all testing, ground tissue was used, obtained predominantly using a lathe with heat-treated (180 minutes @ 120°C) Forstner bits, or on occasion using a scalpel. For the reagent testing experiments (*Experiment III*), due to the scale of the study, ground tissue was prepared in bulk using the lathe and collected in sterilized jars (stored at -80°C until needed) prior to starting. Unless stated 100 mg of tissue was used in each tube. Tissue samples were ground using an *OMNItech Bead Ruptor Elite system*. Tissue was pre-frozen in liquid nitrogen and also between grinding cycles. The results from Experiment II (*Tissue grinding experiment*) would determine the grinding speed and number of cycles.

Quantification of DNA concentration and calculation of DNA purity

The concentration and purity of DNA were measured only for experiments where stated. DNA concentrations were measured using a *Quantus fluorometer* (Promega Corporation, Madison WI, USA). DNA purity was measured using either a *NanoDrop 2000* (Thermo Scientific, Wilmington DE, USA) or *Synergy H1 microplate reader* (Biotek, Winooski, VT, USA). To assess the purity of DNA the A_{260}/A_{230} and A_{260}/A_{280} ratios were measured (Varma, Padh et al. 2007). DNA was considered pure when within the following ranges: 1.7-2.0 for $A_{260/280}$ as per (Särkinen, et al. 2012) and 2.0-2.2 for $A_{260/230}$ as per (Matlock and Thermo Fisher Scientific 2015)[16].

Scoring amplification

DNA purity was also assessed via PCR amplification success. This is an approach that has been used on timber extraction previously (Jiao, et al. 2012; Särkinen, et al. 2012). Success (or failure) is measured by the intensity of the band on the gel. A strong band indicated a highly successful amplification and is associated with clean DNA that is free of contaminants, whereas a weak band indicated that some contaminants passed through the extraction procedure but not enough to

[16] https://tools.thermofisher.com/content/sfs/brochures/TN52646-E-0215M-NucleicAcid.pdf

fully inhibit the amplification. The failure of a sample to amplify signified that the extraction had failed. Strong bands were given a score of 2, weak bands a score of 1 and no amplifications were given a score of 0.

Statistical analysis

The Friedman test calculations were performed using either the *r* package *PMCMR* (Pohlert 2014) or the *astatsa* online program (Vasavada 2016). For either approach, three post-hoc tests were also utilised to identify pairwise relationships (Conover (with *p-values* adjusted by either: 1) Family Wide Adjusted Rate (FWER) (i.e. Holm procedure), or 2) False Discovery Rate (FDR) (i.e. Benjaminyi-Hochberg procedure) 3) Nemenyi (with no *p-value* adjustments)).

Experiment I: Tissue weight experiment

The foundational step of any DNA extraction protocol is the addition of tissue to a tube. The aim of this experiment was to determine whether the tissue weight is critical or if similar results are obtained if the weight is varied. Furthermore, the test included variables designed to assess the maximum and minimum tissue weight detection limits. This is a sensitivity study and is a forensic validation requirement (SWFS 2018, SWGDAM 2016). For the BOTAB protocol the typical weight of tissue per tube is ~100 mg. For this test four weights were used, the standard (100 mg) and three alternatives (200 mg, 50 mg and 10 mg). For the 10 mg weights only one tube was used compared to two for all others (the comparable two tube weight would be 5mg in each tube). Extractions were conducted on one sample each from three timber species (jarrah, English oak and mahogany[C]) (*NB: superscript letter refers to the sample code, see Table S2 for more information*). Apart from the tissue weight, this experiment was conducted using the standard BOTAB (day) protocol. As this is a forensic validation experiment, negative extractions were included and reported. For the 10 mg tissue weight samples, only one tube was used instead of two. DNA concentration and DNA purity were calculated for this experiment.

Experiment II: Tissue grinding experiments

The grinding of tissue is a step of most extraction protocols that is largely overlooked. Yet, for timber samples especially, it is a critical one. Timber material can be tough; the rigid polysaccharide cell walls can make it difficult to break them open and allow the DNA to be extracted (Varma, Padh et al. 2007). The most efficient way of breaking the cell walls is via mechanical grinding (using ceramic or metal beads and shaking them rapidly in specialised equipment). However, care must be taken when grinding not to damage DNA within the cells as the forces required to disrupt cell walls are also the same for shearing high molecular weight DNA (Bürgmann, Pesaro et al. 2001). This experiment was designed to assess whether the current grinding schedule could be optimised to maximise the proportion of tissue that is homogenised whilst also limiting the potential impact of DNA fragmentation. The focus of this experiment was for internal optimisation purposes. Prior to this experiment, both the University of Adelaide (UA) and Thünen Institute for Forest Genetics (TIFG) laboratories used different approaches to tissue grinding. Typically, UA used a bead ratio of 20 small (1.4 mm) beads to three large (2.8 mm) ones (*this can be simplified to 20:3, the format used hereafter*) and grinding for 60secs @ $4ms^{-1}$ in two cycles; whereas the TIFG facility used a bead ratio of 100:0 and grinding for 20secs @ $5ms^{-1}$ in two cycles. Both facilities used the same grinding schedule for both lathe and scalpel prepared samples, despite the fineness of the prepared tissue being different between the techniques. As such, this experiment was also undertaken to identify the requirements for beads when tissue was prepared using either a scalpel or lathe.

The first experiment was to compare both grinding approaches. The next experiment compared different bead combinations for both lathe and scalpel prepared samples (See Table S7 for bead combinations), with samples ground using the TIFG approach. A final experiment then focused on optimising the grinding for scalpel prepared samples only. For this, two different bead ratios (20:3 or 40:3) were used at two different speeds ($3.55ms^{-1}$ or $4.0ms^{-1}$) with up to four cycles. As previously, samples were ground for 20 seconds per cycle. The same two timber samples (merbau and European white oak) were used for all

tissue grinding experiments. For each sample 100 mg of tissue was used. No DNA extractions were undertaken for these experiments and results were measured via visual assessment only. This was done by scorings the samples based on the proportion of tissue ground. See Table S7 for scores and measures evaluating them.

Experiment III: BOTAB extraction buffer reagent experiments

As outlined previously, the BOTAB extraction buffer is a significant focus of the patent (Thesis Appendix V). The patent lists the presently used reagents and concentrations, as well as the alternatives that may be substituted. The aim of this experiment was to determine whether the parameters listed in the patent generate the same result, or whether modification of the patent was required. For each of the functional groups of the extraction buffer, a range of concentrations and alternative reagents were tested and compared against the standards. A list of which reagents, including standard and alternatives, that were tested can be found in Table S3.

Supplementary Table 3: List of standard and alternative BOTAB extraction buffer reagents tested *(Experiment III)*

Functional Group[1]	Full Chemical name	Short hand	CAS Number[2]	Status[3]
Detergent	Cetrimonium bromide	CTAB	57-09-0	standard
	Cetrimonium chloride	CTAC	112-02-7	novel
	Benzalkonium chloride	BAC	63449-41-2	novel
	Benzethonium chloride	BZT	121-54-0	novel
Buffer	tris(hydroxymethyl)methylamine	Tris	77-86-1	standard
	piperazine-N,N′-bis(2-ethanesulfonic acid)	PIPES	5625-37-6	alternative
	2-(N-morpholino)ethanesulfonic acid	MES	4432-31-9	novel
	2(R)-2-(methylamino) succinic acid	Succonic Acid	110-15-6	novel
Chelating agent	Ethylenediaminetetraacetic acid	EDTA	60-00-4	standard
	Ethylene glycol tetraacetic acid	EGTA	67-42-5	alternative
	Ethylenediamine-N,N'-disuccinic acid	EDDS	178949-82-1	novel
	diethylene triamine pentaacetic acid	DTPA	67-43-6	novel
Salt	Sodium chloride	NaCl	7647-14-5	standard
	tri-Sodium Citrate Dihydrate	Urisol	6132-04-3	alternative
Excipient	Polyvinylpyrrolidone 360	PVP 360	9003-39-8	alternative
	Polyvinylpyrrolidone K30	PVP K30	9003-39-8	standard
	Polyvinylpolypyrrolidone	PVPP	25249-54-1	alternative
	Polyethylene glycol 200	PEG 200	25322-68-3	alternative
Biological antioxidant	β-mercaptoethanol	βME	60-24-2	standard
	Dithiothreitol	DTT	3483-12-3	alternative
	Ascorbic acid	AA	50-81-7	alternative
	(+) -Sodium L-ascorbate	NaA	134-03-2	alternative
	tris(2-carboxyethyl)phosphine	TCEP	51805-45-9	alternative
Protease	Proteinase K			standard

NB: [1]broad grouping of chemicals based on their function in the extraction buffer (for descriptions of each functional group and the function of the reagents within it see Table S1; [2]CAS number refers to the unique identifier for every chemical; [3]status infers whether the reagent is the standard chemical used in the BOTAB extraction buffer, an alternative reagent was listed in the initial patent document, or a novel reagent not found in the initial patent document, but added to the final patent.

Varying concentrations of standard and alternative reagents were also used in this experiment. Three concentrations were tested for each reagent, including the concentration that is presently used for standard reagents for each functional group (optimal concentration), as well as two (low and high) alternative concentrations (see Table S4 for the three concentrations tested for reagents in each functional group). For each of the standard BOTAB extraction buffer reagents, to identify their importance in the buffer they were also excluded (removed).

Supplementary Table 4: the three concentrations tested for reagents in each functional group of the BOTAB extraction buffer

	Concentration			
Functional Group	**low**	**Optimal[1]**	**high**	**Units**
Boric acid	0.5	1	2	% (w/v)
Detergent	45	55	65	mM
Buffer	70	100	130	mM
Chelating agent	10	20	30	mM
Salt	1	1.4	2	M
Excipient	1	2	3	% (w/v)
Biological antioxidant	30	47	65	mM
Protease	1	1.86	2.4	%

NB: [1]for standard BOTAB buffer reagents, the reagent was removed (to assess the significance of removing). "optimal" is the term used in the patent document (thesis appendix V)

Testing was done in a randomised order and experiments were carried out using a one variable at a time experimental design, whereby only a single reagent at one concentration was varied at a time. This was to not overcomplicate the experiments. To measure the effect of storing tissue over time (see tissue prep section (*SI: Tissue preparation, storage and use*) for further information), extractions using the standard BOTAB buffers were performed intermittently throughout the testing and results compared at the end. Apart from the extraction buffer variants, the protocol was conducted as per the standard BOTAB protocol. Six timber samples were used for this testing (English oak, mahogany[C], merbau, Zebrano, wenge and Baltic pine). Negative extractions were performed throughout this component. DNA concentrations and purity were measured for these tests. For standard reagents, the three variables (low or high concentration or removed) were compared to standard BOTAB extractions. For alternative reagents, the variable concentrations were pooled, and compared to standard BOTAB extraction results. Friedman tests were conducted for several of the standard reagents to assess the impact of either changing or removing the concentration or removing it. For all other results, only comparisons of DNA concentrations or purity were done.

Experiment IV: Additional extraction buffer NaCl (salt) concentration experiments

While the present concentration (1.4_M) of the BOTAB buffer salt (NaCl) is standard of many extraction buffers, it has been reported that the optimal

concentration may depend upon the study species (Varma, Padh et al. 2007). In addition to this work fulfilling patent requirements, this experiment aimed to identify a more suitable salt concentration for internal purposes. For the experiment five concentrations of NaCl were tested, the standard (1.4_M), three alternatives (1_M, 1.8_M and 2.2_M) and absence (0_M). Similarly, to *Experiment III* parameter, only the NaCl concentration was varied, and the remainder of the protocol was conducted as per the standard BOTAB protocol. The test was undertaken on two timber samples (merbau and European white oak). No negative extractions were included. The DNA purity of the extracted products was measured, and agarose gel scoring of PCR amplifications were recorded.

Experiment V: Additional extraction buffer antioxidant experiments

The alternative extraction buffer antioxidant reagents (Ascorbic Acid (AA) and Sodium Ascorbate (NaA) have been used previously in DNA extractions (Colpaert, Cavers et al. 2005). They are also considerably less toxic than other antioxidant reagents studied (βME, DTT and TCEP), so they are ideal candidate reagents for any realistic options for reducing the toxicity and potential harmfulness of the BOTAB protocol. While the purpose of the original experiment (*Experiment III*) was for the patent, this experiment it was conducted for internal optimisation reasons. For this experiment, the two alternative antioxidant reagents (AA and NaA) were tested at five different concentrations to ascertain if an optimal concentration could be identified. These concentrations encompassed the lower (30_{mM}) and upper (60_{mM}) concentrations used previously (*Experiment III*), a similar concentration to the standard (45_{mM} instead of 47_{mM}) as well as broad low (15_{mM}) and high (75_{mM}) concentration alternative. For comparison the experiment also included samples with no antioxidant reagent (i.e. 0_{mM}) as well as samples extracted with the standard antioxidant reagent and concentration (DTT @ 45_{mM}). A second experiment to compare the most commonly used antioxidants (βME and DTT) to the alternatives (AA and NaA) using 45_{mM} concentrations was then conducted. For both experiments, three timber samples were used (merbau, European white oak and larch), and for the second experiment, a further three samples (mahogany[C & J] and American white oak) were included. No negative extractions

were included. For all samples, the experiments were evaluated by comparing the DNA purity of the extracts. A Friedmann test of the test using four reagents at 45_{mM} concentration was performed. Additionally, for the merbau and both oak samples, PCR amplifications were undertaken, and DNA purity was assessed by reviewing the PCR products on agarose gels.

Experiment VI: Extraction buffer incubation time experiments

The incubation of the extraction buffer could be argued as the most crucial step of any DNA extraction protocol. It facilitates the interaction between the reagents in the extraction buffer and the DNA containing tissue. The incubation time is optimal when it allows for the greatest amount of DNA to be extracted from the tissue. The duration should also be sufficient enough to allow for inhibiting compounds to be neutralised, yet not too long that it may degrade the DNA. Because of its importance in the extraction procedure, the patent includes a section on the incubation time of the extraction buffer. Similarly to the wording of the extraction buffer reagent sections, a range of incubation times are included in the patent document. As such, this experiment was conducted for the purpose of the patent to determine if the range of incubation times listed were sufficient to encompass all incubation times that would yield a result. Additionally, considering that the identification of a novel optimal incubation time would be beneficial for the protocol, this experiment was conducted for internal optimisation purposes as well.

For this experiment a range of incubation times were tested: 1, 2, 3, 4, 4.5, 5, 5.5, 6, and 7 hours, overnight (O/N, ~16 hours), 2-days (~48 hours) and 3-days (~72 hours). Four timber samples were used in this incubation time experiment (mahogany^{C}, merbau, European white oak, and American white oak), however, only merbau and mahogany samples were tested for 2 or 3-day incubation times. The extraction buffer incubations were conducted using the standard temperature (55^{o}C). Samples incubated for short time periods (≤7hrs) were processed independently to the precipitation incubation step (*step 11*) where stored (@ -20^{o}C), then washed as a batch. Samples incubated for longer times (O/N, 2 & 3-days) were always processed independently according to the

protocol, except for the oak O/N samples, which were incubated for the precipitation solution incubation step at -80°C for 1 hour instead, then processed alongside the shorter incubation time samples. No negative extractions were included in this test due to equipment capacity constraints. This experiment was evaluated using agarose gel scores only.

Experiment VII: Extraction buffer incubation temperature experiments
The incubation temperature of the extraction buffer is also an important consideration for attaining optimal DNA concentrations. The range of temperatures that the BOTAB extraction buffer can be used at has been included in the patent. While primary focus of this experiment is for assessing the patent range, the successful identification of a new optimal incubation temperature would benefit the protocol, so this experiment was also for the purpose of internal optimisation.

The standard incubation temperature for the BOTAB protocol extraction buffer is 55°C, yet CTAB extractions have been used successfully at a range of temperatures from 37°C - 65°C (Allen et al. 2006; Särkinen et al. 2012; Verbylaite et al. 2010; Wolfe et al. 2010). For this experiment, three alternative incubation temperatures were compared (room temperature (~26°C), 40°C and 64°C). The incubation of samples at room temperature is of particular interest. If successful it would demonstrate that the extraction buffer incubation does not require equipment (such as heat blocks or water baths) to maintain the temperature, which is a major bottleneck for most laboratories. Four timber samples were used in this experiment (mahogany[C], merbau, European white oak and American white oak). The extraction buffer incubations were conducted as per the standard time (5 hours). Due to capacity constraints, neither the standard incubation temperature, nor negative extractions were included in this experiment. The results were evaluated by comparing the DNA purity of the extractions.

Experiment VIII: Precipitation solution salt experiments

The precipitation salt of the BOTAB protocol is sodium acetate (NaOAc), which is also the most commonly used precipitation salt for CTAB extractions, yet other precipitation salts have been used successfully for plant DNA extractions. The aim of this test was to identify if any of the alternative precipitation salts are suited for timber sample DNA extractions. Alternative precipitation salts are not discussed in the patent, so this experiment was for internal optimisation purposes only. This experiment was conducted using three alternative reagents; Ammonium acetate (NH_4OAc), Lithium Chloride (LiCl) and Potassium acetate (KOAc). AmOAc has been used in CTAB extractions (Alexander L 2016; Särkinen et al. 2012) KOAc is widely used in SDS extractions, (Dellaporta, Wood et al. 1983, Csaikl, Bastian et al. 1998) and LiCl, while not specifically used as a precipitation salt, it has been utilised in extraction buffers for similar reasons (Varma, Padh et al. 2007). For each reagent we tested three different concentrations (Table S5). Two timber samples (merbau and European white oak) were used in this experiment. Neither negative extractions nor samples extracted using the standard precipitation salt (NaOAc) were incorporate in this experiment. Apart from the reagent alterations, the extractions were conducted as per the protocol with the exception of the use of $800_{\mu l}$ of isopropanol (instead of $600_{\mu l}$) for the NH_4OAc samples. To evaluate the results, DNA purity of the extractions was calculated and compared. Additionally, for merbau only, agarose gel assessments were made.

Supplementary Table 5: Alternative precipitation salts tested in experiment VIII

Regent			concentration tested		
name	**Short hand**	**CAS number[1]**	**low**	**medium**	**high**
Ammonium acetate[2]	NH_4OAc	631-61-8	2.1_M	2.5_M	2.8_M
Lithium chloride	LiCl	7447-41-8	0.57_M	0.8_M	1_M
Potassium acetate	KOAc	127-08-2	0.57_M	0.8_M	1_M

NB: [1]CAS number refers to the unique identifier for every chemical; [2]additional volume of isopropanol used (to maintain ratio)

Experiment IX: Precipitation Solution incubation time and temperature experiments

The final step of the BOTAB protocol to be specifically evaluated in this study was for the incubation of the precipitation solution. This experiment was only for the purposes of internally optimising the BOTAB protocol. It was conducted to assess whether the time required for this stage could be reduced (the standard

procedure has an incubation overnight (~16 hours at -20°C)) and whether the temperature had any influence on DNA purity and PCR success. For this test a multivariable experiment comparing both incubation time and temperature together was used. The experiments incorporated four different incubation times (30min, 1hr, 2hr, 4hr & O/N (~24hrs)) using two incubation temperatures (-20°C or -80°C). Absolute isopropanol (100%) has a freezing point of -89°C, so solutions with ~40-50% isopropanol, the expected concentration in the precipitation solution, will freeze at approximately -20°C. At present the solution never freezes at -20°C, regardless of time, so there are no critical time constraints at this temperature. However, for incubations in lower temperature freezers, such as -80°C, then time considerations need to be factored in. Hence the emphasis on shorter rather than longer times in this experiment. Experiments were conducted using two timber samples (mahogany[C] and merbau). To facilitate the time considerations of these experiments, this experiment was conducted using an O/N incubation time for the extraction buffer. Negative extractions were not included. The extracts from these experiments were amplified and DNA purity was assessed by reviewing agarose gels.

Experiment X: Comparison of BOTAB protocol to commercial kits

The final component of this chapter was to compare the BOTAB protocol to commercially available kits. This test forms a significant part of the requirements for forensic validation of the BOTAB protocol (SWFS 2018, SWGDAM 2016). The aim of this experiment was not to determine the best protocol, instead, it was to determine if comparable results could be obtained using the BOTAB protocol to those of the commercial kits. The BOTAB protocol was compared to three widely available commercial plant DNA extraction Kits, Qiagen: DNeasy Plant Mini Kit (hereafter Qiagen kit), Analytik Jena: InnuPREP Plant DNA Kit (hereafter Jena kit) and the MoBio: Power Plant Pro DNA Isolation Kit (hereafter Mobio kit). The kit protocols were modified to make them more suitable for timber samples, and the specific alterations for each kit is listed below. The study was undertaken by performing BOTAB extractions at the beginning and the end of the study, with the kits being used in the middle. The experiments were undertaken using six

timber samples (jarrah, English oak, mahogany[C], wenge, Baltic pine and bigleaf maple). BOTAB extractions were undertaken using the updated BOTAB (day) protocol. Samples were evaluated by comparing DNA concentration and purity.

Qiagen kit

The Qiagen kit has been used successfully used on timber samples previously (Rachmayanti, Leinemann et al. 2009, Jiao, Yin et al. 2014), and is considered the benchmark for kit based DNA extractions for timber samples. The extractions were undertaken using the Rachmayanti (2006) modified protocol with the following alterations. 100 mg of starting tissue used. 2%$^{(w/v)}$ PVP included in AP1 lysis buffer with 800$_{\mu l}$ AP1 lysis buffer and 8$_{\mu l}$ RNase A added to each sample. Prior to addition of the AP2 buffer, samples centrifuged (for 30 seconds @ 17500 g) before 400$_{\mu l}$ of the buffer/sample mixture was transferred to a new 1.5$_{ml}$ tube and mixed with 260$_{\mu l}$ of AP2 buffer and incubated as per protocol. Elutions were performed using 15$_{\mu l}$ of AE buffer and incubated (@ room temperature) for five minutes and were pooled together rather than remaining separate.

Jena Kit

The Jena kit has also been used successfully on timber samples previously using an unmodified protocol (Telle and Thines 2008; Verbylaite et al. 2010). For this experiment, DNA extractions were performed using the OPT protocol, and were undertaken as per the manufacturer's protocol with following changes: 800$_{\mu l}$ of OPT Lysis Solution was added to the sample tube and mixed well. The incubation time was extended to 2 hours (65°C, mixing at 500 rpm). After the incubation of the lysis buffer, samples were centrifuged (for 30 seconds @ 17500 g) before 400$_{\mu l}$ of the buffer/sample mixture was transferred to a new 1.5$_{ml}$ tube and mixed with 100$_{\mu l}$ of Precipitation buffer. The RNA removal step was left out. The final elution was done twice, with 20$_{\mu l}$ used for each elution. The elution incubation time was extended to 3 minutes.

NB: the OPT lysis buffer is an SDS lysis buffer (as per the MSDS) and contains between 1-2.5% SDS.

MoBio Kit

The capability of the MoBio kit for working with timber material is relatively unknown. The MoBio kit extractions were performed as per manufacturers' protocol, with following changes: 100 mg of tissue along with 900$_{\mu l}$ of solution PD1 and 100$_{\mu l}$ of solution PD2 were added to our own grinding tubes (with our own beads). Sample were incubated for 45 minutes at 65°C prior to homogenization. 20$_{\mu l}$ of solution PD7 was loaded to the centre of the filter and incubated for 5 minutes and was repeated a second time (20$_{\mu l}$ elution volume and 5 minutes incubation).

Results

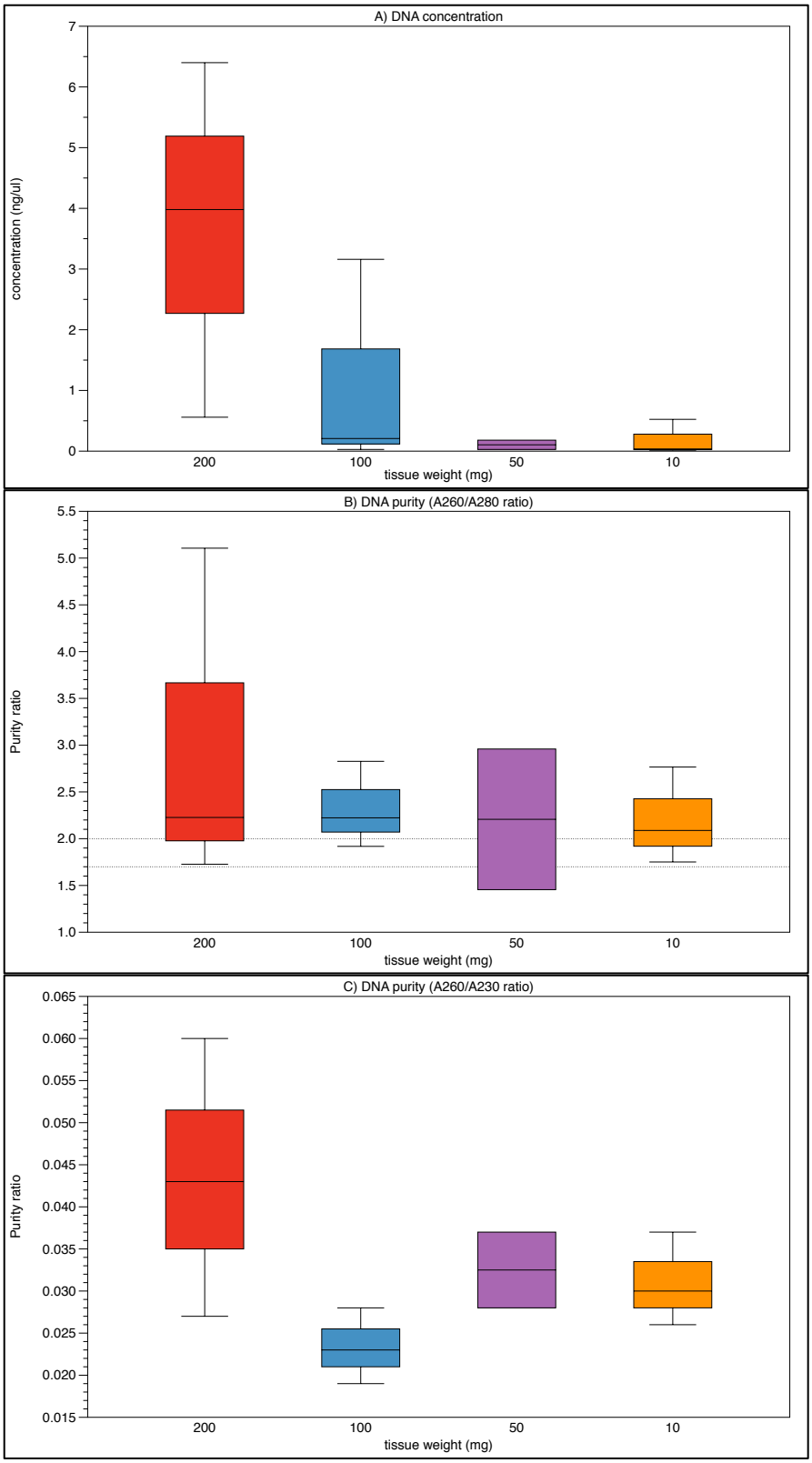
Experiment I: Tissue weight experiments

Due to space availability and to ensure a negative extraction was included in this experiment, there were no 50mg tubes for jarrah. The highest median DNA yield was found when 200 mg of tissue was used (2.30$_{ng/\mu l}$). This median concentration was much higher than for the other three weights (100 mg = 0.21$_{ng/\mu l}$, 50 mg = 0.10$_{ng/\mu l}$ and 10 mg =0.04$_{ng/\mu l}$). The A_{260}/A_{280} ratios were comparable between three of the weights (100 mg, 200 mg and 10 mg) ranging from 2.04 to 2.06. The absence of jarrah samples with 50 mg tissue weight meant the median A_{260}/A_{280} ratio (1.49) was not similar to the other weights. The median A_{260}/A_{230} ratios of the four tissue weights were comparable, yet were all very low (<0.05) and similar to the negative extraction results. Figure S1 displays the median tissue weight for all timber samples at each tissue weight, and the median for each timber sample specifically can be seen in Table S6.

Supplementary table 6: Median (per timber samples) DNA concentration and purity for specific tissue weights.

	tissue weight (mg)			
Sample	**100**	**200**	**50**	**10**
DNA concentration (ng/µl)				
jarrah	3.16	6.40		0.52
English oak	0.03	3.98	0.03	0.02
mahogany[C]	0.21	0.56	0.18	0.04
median	*0.21*	*2.30*	*0.10*	*0.04*
negative extraction	0.00		0.02[1]	
DNA purity (A_{260}/A_{280} ratio)				
jarrah	2.83	1.73		2.77
English oak	1.92	2.23	1.46	2.09
mahogany[C]	2.22	5.11	2.96	1.75
median	*2.04*	*2.06*	*1.49*	*2.05*
negative extraction	4.21		4.23[1]	
DNA purity (A_{260}/A_{230} ratio)				
jarrah	0.03	0.04		0.03
English oak	0.02	0.03	0.04	0.03
mahogany[C]	0.02	0.06	0.03	0.04
median	*0.02*	*0.05*	*0.03*	*0.03*
negative extraction	0.03		0.03[1]	

NB: jarrah samples not tested for 50 mg weight test, [1]alternative tissue weight tests conducted at same time, only one negative/extraction used for test



Supplementary Figure 1: Median (of all samples) DNA concentration (A) and purity (B & C) for specific starting tissue weights.
NB: Appropriate range of DNA purity ratio for the A260/A280 are indicated by dotted lines (i.e. 1.7 – 2.0 as per (Sarkinen et al.2012). DNA purity ratio A260/A230 was all well below appropriate range. filled circles signify outlier samples, open circles signify extreme outliers (that are three times the IQR).

Experiment II: Tissue grinding experiments

The initial test identified that sufficient grinding can be achieved using the TIFG schedule (i.e. 2 cycles for 20 seconds each @ $5ms^{-1}$), and the UA grinding time (60 seconds) generated the same result. However, the TIFG approach is to use only using small beads (Combination G6), and not all tissue from scalpel prepared samples was ground up (score of two for all 4 samples). Problems also occurred with the samples only using large beads (Combination G6) as some of the sample tubes cracked after the first grinding cycle. Yet despite this setback, for the lathe prepared samples at least, all tissue was completely ground after the one cycle. Furthermore, for all lathe prepared samples, complete grinding success was achieved using any of the bead combinations. However, for scalpel prepared samples, the results were not as successful. None of the bead combinations was sufficient to completely grind all tissue. For both merbau and oak, the best results occurred when either no large beads were used (combination G1) of if used, in a ratio with smaller beads (combination G5 (20:3) for both samples or G4 (25:3) for oak only). If only large beads were used (combination G6) then no change to the scalpel prepared samples occurred. The results for the first experiment can be seen in Table S7.

Supplementary table 7: Bead combinations/mixtures grinding results (Experiment II)

	# beads per size		merbau		European white oak		**sample**
Combination #	1.4_{mm}	2.8_{mm}	lathe	scalpel	lathe	scalpel	**tissue prep method**
G1	100	0	+++	++	+++	++	
G2	50	0	+++	+	+++	+	
G3	30	2	+++	+	+++	+	
G4	25	3	+++	+	+++	++	
[1]G5A	20	3	+++	+	+++	++	
[1]G5T	20	3	+++	++	+++	++	
[2]G6	0	4	+++	-	+++	-	

NB: 1.4 mm=small beads, 2.8 mm=large beads. Bead combination referred to as shorthand in text (e.g. combination G1 (100 small beads and 0 large beads is written shorthand as 100:0). [1]bead combination run at two variable speed and time grinding schedules: G5A = 60 seconds @ $4ms^{-1}$ (the original UA approach) or G5T = 20 seconds @ $5ms^{-1}$ (the TIFG approach). [2]G6 samples only ground in one cycle. Scoring: +/- represent proportion of tissue ground (-=no change from input tissue, +=1/3 of tissue ground, ++=2/3 of tissue ground, +++=all tissue ground).

In the final grinding experiment, neither the bead mixtures nor the grinding speeds had an influence on grinding effectiveness in the scalpel prepared samples, and similar proportions of ground tissue were seen in all combinations tested for both oak and merbau. Nevertheless, what did influence the capability

for grinding was the number of cycles. With the proportion of ground tissue increasing as the number of cycles increased. For both samples, the biggest change in grinding proportions occurred after three cycles (score of 2), with limited change identified after four cycles (score 2). The results from this combination experiment can be seen in Table S8.

Supplementary table 8: Grinding test speed, bead ratio and cycle number combination results (*Experiment II*)

sample		merbau				European white oak			
Bead combination (s:l)		20:3		40:3		20:3		40:03	
Grinding speed (ms^{-1})		3.55	4.00	3.55	4.00	3.55	4.00	3.55	4.00
# grinding cycles	1	-	-	-	-	+	+	+	+
	2	+	+	+	+	+	+	+	+
	3	++	++	++	++	++	++	++	++
	4	++	++	++	++	++	++	++	++

NB: ratio refers to number of each size bead used (s=small beads (1.4 mm) l= large beads (2.8 mm)).
Scoring: +/- represent proportion of tissue ground (-=no change from input tissue, +=1/3 of tissue ground, ++=2/3 of tissue ground, +++=all tissue ground).

Experiment III: BOTAB extraction buffer reagent experiments

For this study, we were only able to perform extractions on all buffer reagents and concentrations for two timber samples (mahogany[C] and zebrano). For the other four samples (merbau, English oak, Baltic pine, wenge), because of time constraints and limited available ground tissue, only a few of the tests could be completed. See Tables S9-11 for DNA yields and purity of all standard reagents at different concentrations. Statistical analyses could be done for four of the standard BOTAB buffer reagents (DTT (Biological antioxidant), Boric Acid, CTAB (Detergent) and Proteinase K (Protease)). For these reagents, there were results for the six timber samples at three additional reagent concentrations to the standard. The DNA yield and purity for these was collated and assessed. All were found to be significant except one (*effect of Proteinase K concentration on DNA yield*). This was found in all three of the measurements. Boxplot figures of all of these assessable tests have been included in the experiment can be seen below (Figures S2 (DTT), S3 (Boric Acid), S4 (CTAB) & S5 (Proteinase K)).

Supplementary Table 9: DNA concentration (ng/ul) of standard BOTAB buffer reagents at standard concentration, two alternatives (higher or lower) or removed

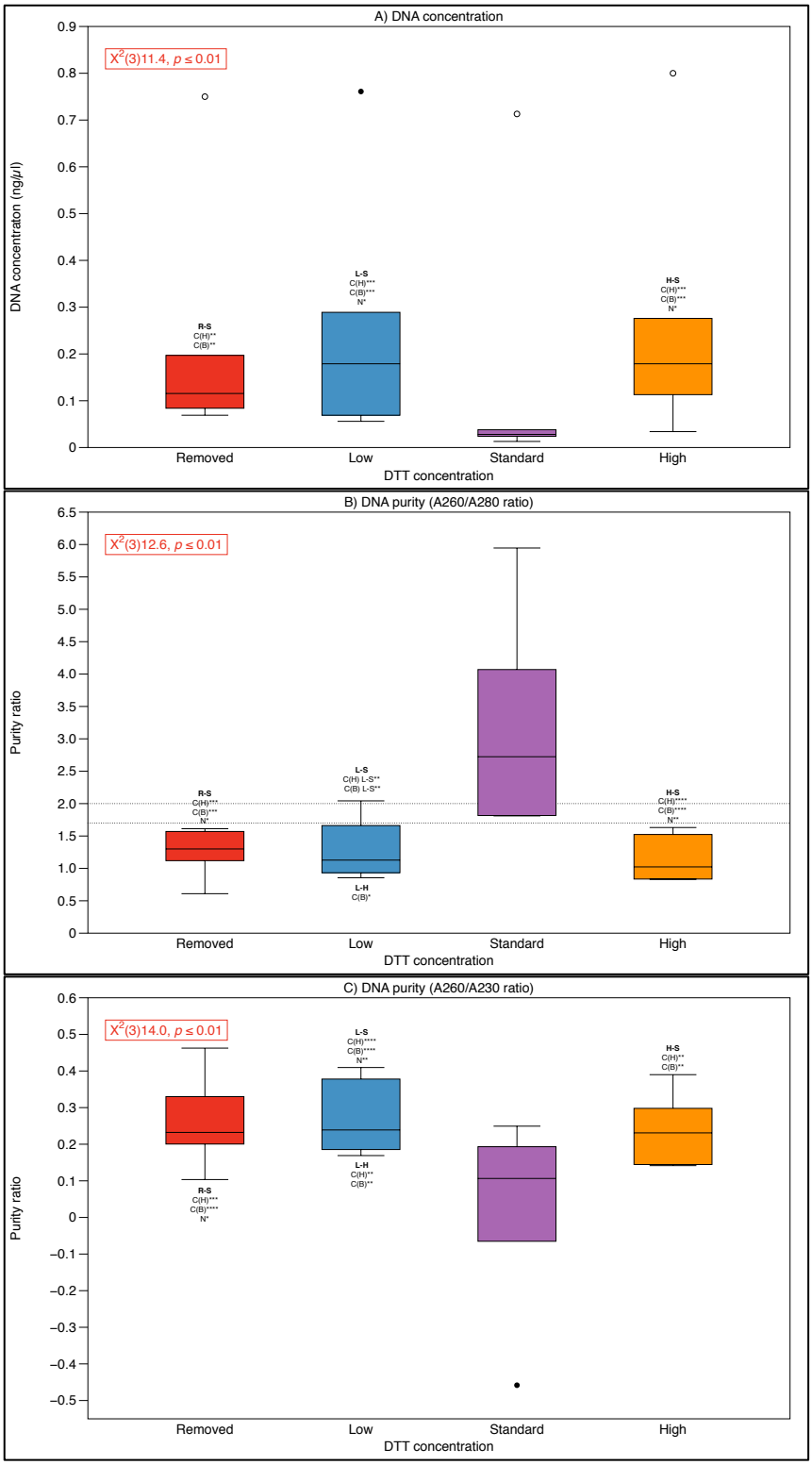
		reagent concentration			
reagent	**species**	**Standard**	**removed**	**low**	**high**
DTT					
	English oak	0.030	0.197	0.289	0.276
	mahogany[C]	0.038	0.084	0.247	0.196
	merbau	0.026	0.094	0.111	0.162
	Wenge	0.024	0.137	0.069	0.113
	zebrano	0.713	0.750	0.761	0.800
	Baltic Pine	0.013	0.069	0.056	0.034
	median	*0.028*	*0.115*	*0.179*	*0.179*
Boric Acid					
	English oak	0.030	0.201	0.419	0.579
	mahogany[C]	0.038	0.476	0.133	0.244
	merbau	0.026	0.090	0.056	0.064
	zebrano	0.713	0.914	0.722	0.981
	wenge	0.024	0.415	0.093	0.094
	Baltic Pine	0.013	0.227	0.046	0.043
	median	*0.028*	*0.321*	*0.113*	*0.169*
TRIS					
	English oak	0.030	0.325		
	mahogany[C]	0.038	0.312	0.026	0.139
	merbau	0.026	0.039		
	zebrano	0.713	0.441	0.606	0.522
	wenge	0.024	0.111		
	Baltic Pine	0.013	0.040		
	median	*0.028*	*0.211*	*0.316*	*0.330*
EDTA					
	mahogany[C]	0.038	0.070	0.358	0.102
	zebrano	0.713	1.108	0.673	0.892
	median	*0.376*	*0.589*	*0.515*	*0.497*
CTAB					
	English oak	0.030	2.506	0.404	0.165
	mahogany[C]	0.038	0.686	0.159	0.152
	merbau	0.026	0.060	0.127	0.070
	zebrano	0.713	0.663	0.742	0.759
	wenge	0.024	0.207	0.078	0.129
	Baltic Pine	0.013	0.136	0.035	0.062
	median	*0.028*	*0.435*	*0.143*	*0.141*
PVP					
	mahogany[C]	0.038	0.015	0.077	0.114
	zebrano	0.713	0.416	0.571	0.403
	median	*0.376*	*0.215*	*0.324*	*0.258*
Proteinase K					
	English oak	0.030	0.172	0.126	0.137
	mahogany[C]	0.038	0.181	0.067	0.066
	merbau	0.026	0.094	0.052	0.018
	zebrano	0.713	0.499	0.493	0.629
	wenge	0.024	0.121	0.064	0.286
	Baltic Pine	0.013	0.214	0.027	0.024
	median	*0.028*	*0.176*	*0.066*	*0.101*
NaCl					
	mahogany[C]	0.038	0.036	0.103	0.014
	zebrano	0.713	0.013	0.526	0.637
	median	*0.376*	*0.024*	*0.314*	*0.326*

Supplementary Table 10: A_{260}/A_{280} DNA purity ratio of standard BOTAB buffer reagents at standard concentration, two alternatives (higher or lower) or removed

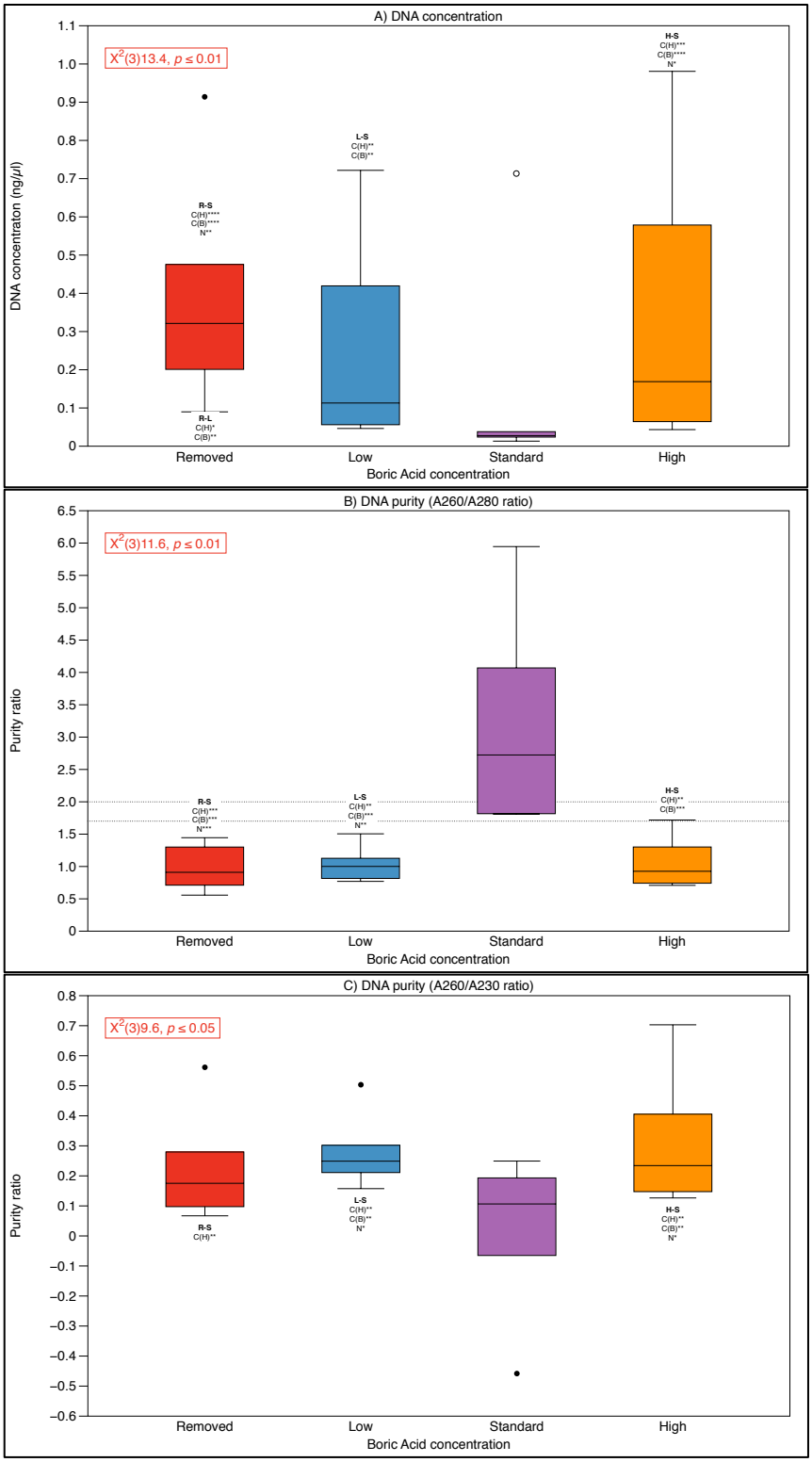
		reagent concentration			
reagent	**species**	**Standard**	**removed**	**low**	**high**
DTT					
	English oak	1.816	1.613	1.662	1.631
	mahogany[C]	1.811	1.139	1.245	0.998
	merbau	2.517	1.461	1.012	0.833
	wenge	5.946	1.119	0.933	0.836
	zebrano	2.930	1.570	2.043	1.524
	Baltic Pine	4.070	0.610	0.856	1.049
	median	*2.724*	*1.300*	*1.128*	*1.023*
Boric Acid					
	English oak	1.816	1.445	1.505	1.719
	mahogany[C]	1.811	1.301	0.816	0.709
	merbau	2.517	0.711	0.770	0.833
	zebrano	2.930	0.557	1.127	1.302
	wenge	5.946	0.936	0.896	1.020
	Baltic Pine	4.070	0.884	1.106	0.742
	median	*2.724*	*0.910*	*1.001*	*0.926*
TRIS					
	English oak	1.816	1.637		
	mahogany[C]	1.811	0.781	1.600	0.841
	merbau	2.517	0.964		
	zebrano	2.930	0.815	1.400	0.984
	wenge	5.946	0.999		
	Baltic Pine	4.070	1.029		
	median	*2.724*	*0.981*	*1.500*	*0.912*
EDTA					
	mahogany[C]	1.811	1.082	0.731	1.245
	zebrano	2.930	1.195	1.258	1.566
	median	*2.371*	*1.138*	*0.995*	*1.405*
CTAB					
	English oak	1.816	1.161	1.740	1.526
	mahogany[C]	1.811	0.618	0.686	0.615
	merbau	2.517	0.787	1.331	0.698
	zebrano	2.930	0.866	1.285	1.471
	wenge	5.946	0.949	0.986	0.945
	Baltic Pine	4.070	0.775	0.563	0.835
	median	*2.724*	*0.826*	*1.135*	*0.890*
PVP					
	mahogany[C]	1.811	1.122	1.556	1.099
	zebrano	2.930	1.456	1.429	1.241
	median	*2.371*	*1.289*	*1.492*	*1.170*
Proteinase K					
	English oak	1.816	1.607	1.640	1.485
	mahogany[C]	1.811	1.099	1.537	4.152
	merbau	2.517	1.477	0.685	0.927
	zebrano	2.930	1.546	2.323	1.600
	wenge	5.946	0.831	0.895	0.923
	Baltic Pine	4.070	1.057	1.270	1.203
	median	*2.724*	*1.288*	*1.404*	*1.344*
NaCl					
	mahogany[C]	1.811	1.223	0.933	1.190
	zebrano	2.930	1.503	1.970	1.644
	median	*2.371*	*1.363*	*1.451*	*1.417*

Supplementary Table 11: A_{260}/A_{230} DNA purity ratio of standard BOTAB buffer reagents at standard concentration, two alternatives (higher or lower) or removed

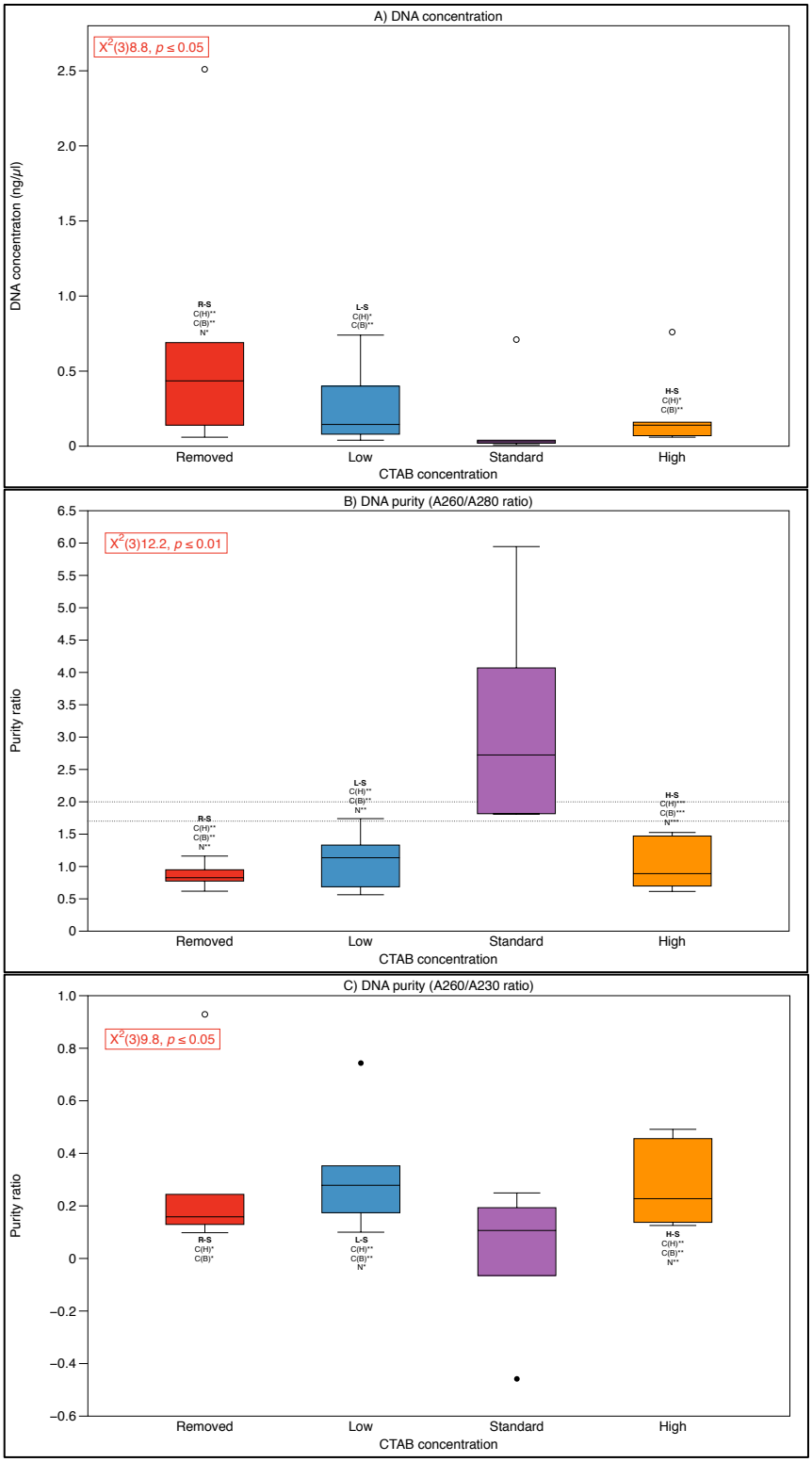
		reagent concentration			
reagent	**species**	**Standard**	**removed**	**low**	**high**
DTT					
	English oak	0.193	0.463	0.410	0.390
	mahogany[C]	0.062	0.212	0.217	0.212
	merbau	-0.458	0.201	0.169	0.142
	wenge	0.151	0.253	0.262	0.250
	zebrano	0.249	0.330	0.378	0.298
	Baltic Pine	-0.065	0.104	0.186	0.145
	median	*0.106*	*0.232*	*0.239*	*0.231*
Boric Acid					
	English oak	0.193	0.562	0.504	0.703
	mahogany[C]	0.062	0.280	0.211	0.184
	merbau	-0.458	0.149	0.158	0.148
	zebrano	0.249	0.067	0.302	0.406
	wenge	0.151	0.202	0.263	0.285
	Baltic Pine	-0.065	0.098	0.235	0.127
	median	*0.106*	*0.175*	*0.249*	*0.234*
TRIS					
	English oak	0.193	0.428		
	mahogany[C]	0.062	0.182	0.198	0.187
	merbau	-0.458	0.186		
	zebrano	0.249	0.143	0.426	0.446
	wenge	0.151	0.226		
	Baltic Pine	-0.065	0.108		
	median	*0.106*	*0.184*	*0.312*	*0.317*
EDTA					
	mahogany[C]	0.062	0.222	0.206	0.345
	zebrano	0.249	0.383	0.199	0.335
	median	*0.156*	*0.302*	*0.202*	*0.340*
CTAB					
	English oak	0.193	0.929	0.744	0.492
	mahogany[C]	0.062	0.165	0.174	0.164
	merbau	-0.458	0.152	0.288	0.126
	zebrano	0.249	0.130	0.353	0.456
	wenge	0.151	0.244	0.269	0.291
	Baltic Pine	-0.065	0.098	0.100	0.138
	median	*0.106*	*0.159*	*0.279*	*0.228*
PVP					
	mahogany[C]	0.062	0.189	0.253	0.217
	zebrano	0.249	0.452	0.395	0.298
	median	*0.156*	*0.321*	*0.324*	*0.258*
Proteinase K					
	English oak	0.193	0.452	0.422	0.478
	mahogany[C]	0.062	0.229	0.301	0.181
	merbau	-0.458	0.160	0.137	0.139
	zebrano	0.249	0.299	0.231	0.363
	wenge	0.151	0.227	0.253	0.274
	Baltic Pine	-0.065	0.159	0.220	0.200
	median	*0.106*	*0.228*	*0.242*	*0.237*
NaCl					
	mahogany[C]	0.062	0.136	0.179	0.068
	zebrano	0.249	0.130	0.277	0.353
	median	*0.156*	*0.133*	*0.228*	*0.211*



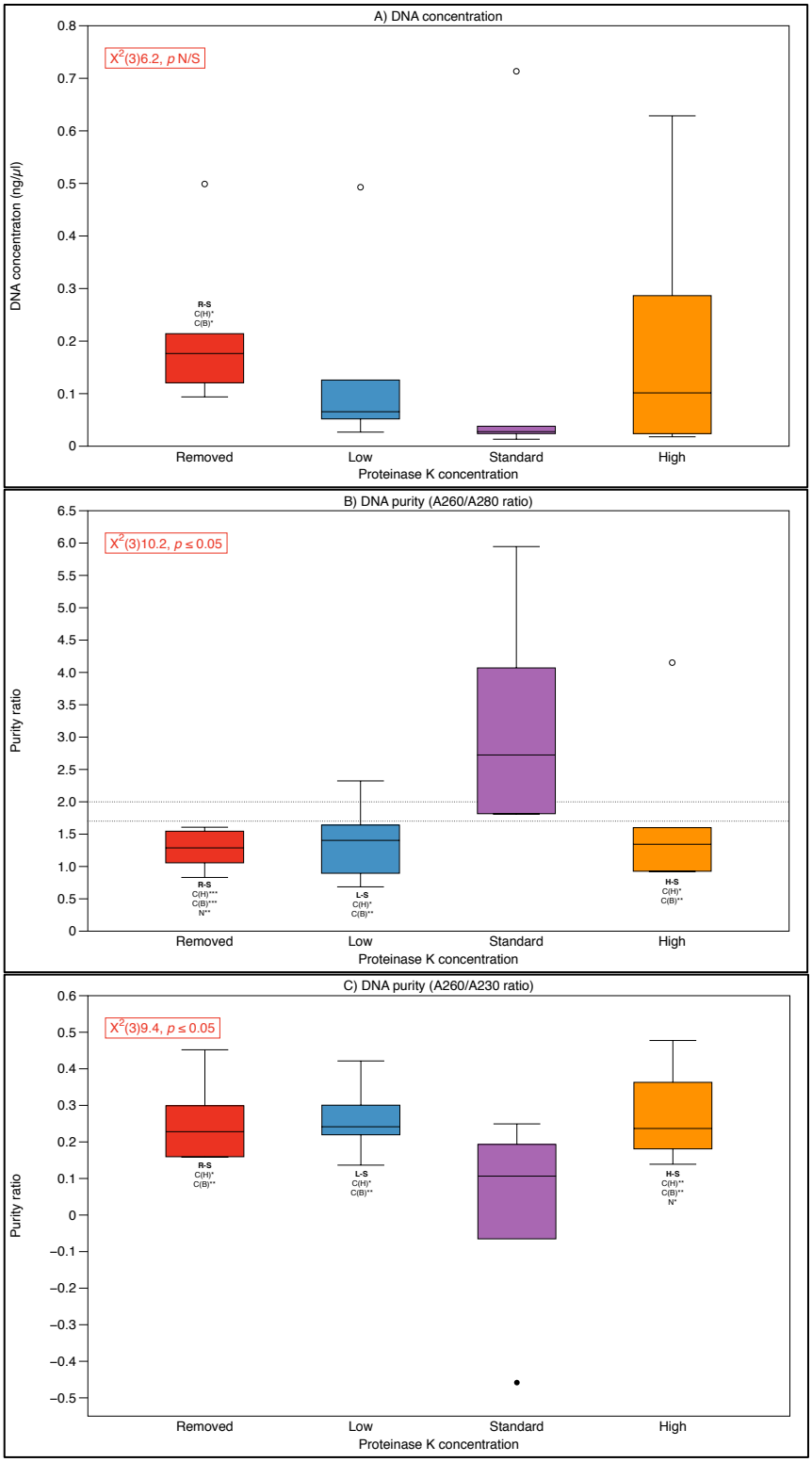
Supplementary Figure 2: Boxplots of standard BOTAB buffer reagent DTT at four different concentrations for DNA concentration (A) and purity (B&C). *NB: Dotted lines infer the DNA purity range for that ratio (i.e. 1.7-2.0 for $A_{260/280}$ as per (Särkinen, et al. 2012 and 2.0-2.2 for $A_{260/230}$ as per (Matlock and Thermo Fisher Scientific 2015), missing dotted line indicates that purity ratios were not within the range. Red box indicates the chi square and p-value result, bold letters in the figure refer to a significant difference between two concentrations (R=Removed, L=Low, S=Standard, H=High), letters below bold text refer to post-hoc tests (C(H)=Covoner (Holm procedure), C(B)=Conover (Benjaminyi-Hochberg procedure), N=Nemenyi procedure, stars following these letters refer to p-value scores from the post-hoc tests (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001) filled circles signify outlier samples, open circles signify extreme outliers (that are three times the IQR).*



Supplementary Figure 3: Boxplots of standard BOTAB buffer reagent Boric Acid at four different concentrations for DNA concentration (A) and purity (B&C). *NB: Dotted lines infer the DNA purity range for that ratio (i.e. 1.7-2.0 for $A_{260/280}$ as per (Särkinen, et al. 2012) and 2.0-2.2 for $A_{260/230}$ as per (Matlock and Thermo Fisher Scientific 2015), missing dotted line indicates that purity ratios were not within the range. Red box indicates the chi square and p-value result, bold letters in the figure refer to a significant difference between two concentrations (R=Removed, L=Low, S=Standard, H=High), letters below bold text refer to post-hoc tests (C(H)=Covoner (Holm procedure), C(B)=Conover (Benjaminyi-Hochberg procedure), N=Nemenyi procedure, stars following these letters refer to p-value scores from the post-hoc tests (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001) filled circles signify outlier samples, open circles signify extreme outliers (that are three times the IQR).*



Supplementary Figure 4: Boxplots of standard BOTAB buffer reagent CTAB at four different concentrations for DNA concentration (A) and purity (B&C). *NB: Dotted lines infer the DNA purity range for that ratio (i.e. 1.7-2.0 for $A_{260/280}$ as per (Särkinen, et al. 2012) and 2.0-2.2 for $A_{260/230}$ as per (Matlock and Thermo Fisher Scientific 2015), missing dotted line indicates that purity ratios were not within the range. Red box indicates the chi square and p-value result, bold letters in the figure refer to a significant difference between two concentrations (R=Removed, L=Low, S=Standard, H=High), letters below bold text refer to post-hoc tests (C(H)=Covoner (Holm procedure), C(B)=Conover (Benjaminyi-Hochberg procedure), N=Nemenyi procedure, stars following these letters refer to p-value scores from the post-hoc tests (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001). filled circles signify outlier samples, open circles signify extreme outliers (that are three times the IQR).*



Supplementary Figure 5: Boxplots of standard BOTAB buffer reagent Proteinase K at four different concentrations for DNA concentration (A) and purity (B&C). *NB: Dotted lines infer the DNA purity range for that ratio (i.e. 1.7-2.0 for $A_{260/280}$ as per (Särkinen, et al. 2012) and 2.0-2.2 for $A_{260/230}$ as per (Matlock and Thermo Fisher Scientific 2015), missing dotted line indicates that purity ratios were not within the range. Red box indicates the chi square and p-value result, bold letters in the figure refer to a significant difference between two concentrations (R=Removed, L=Low, S=Standard, H=High), letters below bold text refer to post-hoc tests (C(H)=Covoner (Holm procedure), C(B)=Conover (Benjaminyi-Hochberg procedure), N=Nemenyi procedure, stars following these letters refer to p-value scores from the post-hoc tests (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001). filled circles signify outlier samples, open circles signify extreme outliers (that are three times the IQR).*

Of the alternative reagents tested, TCEP had the highest median DNA concentration ($1.131_{ng/ul}$), and Sodium acetate (NaA) the highest median DNA purity scores for both ratios (A_{260}/A_{280} = 1.676, A_{260}/A_{230} = 1.668). The lowest combined median DNA concentration was urisol (0.024) but only one concentration for one species (zebrano @ high concentration) had sufficient DNA to be quantified. The lowest median DNA concentration where all samples and concentrations were tested was for CTAC ($0.035_{ng/ul}$), and the lowest DNA purity rations were PVPK30 (A_{260}/A_{280} ratio (0.739)) and urisol (A_{260}/A_{230} ratio (0.006)). Typically the median DNA concentrations for the alternative reagents was higher than that when standard reagents were used. Yet, the A_{260}/A_{280} ratio DNA purity scores were always lower. See below for a full summary of the DNA concentration scores (Table S12) and purity ratios (Tables S13 & S14).

Supplementary Table 12: DNA concentration (ng/ul) results for alternative BOTAB buffer reagents (*Experiment III*)

			Reagent concentration			
Functional group	**Reagent**	**species**	**low**	**med**	**high**	*median*
Biological antioxidant	BME					
		mahogany[C]	0.446	0.437	0.333	*0.437*
		zebrano	0.851	0.806	0.908	*0.851*
		median	*0.648*	*0.621*	*0.621*	*0.626*
	AA					
		mahogany[C]	0.206	0.139	0.184	*0.184*
		zebrano	0.602	0.772	0.716	*0.716*
		median	*0.404*	*0.456*	*0.450*	*0.404*
	NaA					
		mahogany[C]	0.069	0.055	0.046	*0.055*
		zebrano	0.758	0.917	1.015	*0.917*
		median	*0.413*	*0.486*	*0.531*	*0.413*
	TCEP					
		mahogany[C]	1.216	0.133	0.154	*0.154*
		zebrano	1.119	1.311	1.142	*1.142*
		median	*1.168*	*0.722*	*0.648*	*1.131*
Buffer	PIPES					
		mahogany[C]	0.045	0.068	0.060	*0.060*
		zebrano	0.371	0.493	0.300	*0.371*
		median	*0.208*	*0.281*	*0.180*	*0.184*
	MES					
		mahogany[C]	0.914	0.872	0.631	*0.872*
		zebrano	0.514	0.622	0.674	*0.622*
		median	*0.714*	*0.747*	*0.652*	*0.652*
Chelating agent	EGTA					
		mahogany[C]	0.321	0.227	0.129	*0.227*
		zebrano	0.706	0.908	1.285	*0.908*
		median	*0.514*	*0.567*	*0.707*	*0.514*
	EDDS					
		mahogany[C]	0.004	0.083	0.043	*0.043*
		zebrano	0.059	0.487		*0.273*

Functional group	Reagent	species	Reagent concentration low	med	high	*median*
		median	*0.032*	*0.285*	*0.043*	*0.059*
	DTPA					
		mahogany[C]	0.304	0.278	0.285	*0.285*
		zebrano	1.969	0.477	2.010	*1.969*
		median	*1.137*	*0.377*	*1.148*	*0.390*
Detergent	CTAC					
		English oak	0.173	0.013	0.032	*0.032*
		mahogany[C]	0.031	0.005	0.019	*0.019*
		merbau	0.091	0.017	0.057	*0.057*
		zebrano	0.294	0.296	0.369	*0.296*
		wenge	0.033	0.041	0.036	*0.036*
		Baltic Pine	0.039	0.016	0.034	*0.034*
		median	*0.039*	*0.017*	*0.036*	*0.035*
	BAC					
		mahogany[C]	0.147	0.184	0.151	*0.151*
		Zebrano	0.697	0.660	0.592	*0.660*
		median	*0.422*	*0.422*	*0.371*	*0.388*
	BZT					
		English oak	0.586	0.819	0.472	*0.586*
		mahogany[C]	0.101	0.117	0.160	*0.117*
		merbau	0.056	0.023	0.070	*0.056*
		zebrano	0.529	0.416	0.592	*0.529*
		Baltic Pine	0.017	0.012	0.034	*0.017*
		wenge	0.058	0.058	0.099	*0.058*
		median	*0.058*	*0.058*	*0.099*	*0.100*
Excipient	PVP-K30					
		mahogany[C]	0.076	0.132	0.075	*0.076*
		zebrano	0.163	0.355	0.158	*0.163*
		median	*0.119*	*0.243*	*0.116*	*0.145*
	PVPP					
		mahogany[C]	0.054	0.069	0.047	*0.054*
		zebrano	0.204	0.376	0.610	*0.376*
		median	*0.129*	*0.223*	*0.328*	*0.137*
	PEG 200					
		mahogany[C]	0.054	0.024	0.046	*0.046*
		zebrano	0.383	0.572	0.179	*0.383*
		median	*0.218*	*0.298*	*0.113*	*0.117*
Salt	Urisol					
		mahogany[C]				
		zebrano			0.024	*0.024*
		median			*0.024*	*0.024*

Supplementary Table 13: A_{260}/A_{280} DNA purity ratio results for alternative BOTAB buffer reagents (*Experiment III*)

			Reagent concentration			
Functional group	**Reagent**	**species**	**low**	**med**	**high**	*median*
Biological antioxidant	BME					
		mahogany[c]	0.637	1.016	1.017	*1.016*
		zebrano	3.055	1.513	1.423	*1.513*
		median	*1.846*	*1.265*	*1.220*	*1.220*
	AA					
		mahogany[c]	1.416	1.475	1.546	*1.475*
		zebrano	1.771	1.802	1.731	*1.771*
		median	*1.593*	*1.639*	*1.638*	*1.638*
	NaA					
		mahogany[c]	1.114	1.135	1.586	*1.135*
		zebrano	1.766	1.775	1.786	*1.775*
		median	*1.440*	*1.455*	*1.686*	*1.676*
	TCEP					
		mahogany[c]	1.094	1.015	1.178	*1.094*
		zebrano	1.416	1.367	1.021	*1.367*
		median	*1.255*	*1.191*	*1.100*	*1.136*
Buffer	PIPES					
		mahogany[c]	0.950	1.231	1.205	*1.205*
		zebrano	1.355	1.330	1.511	*1.355*
		median	*1.152*	*1.280*	*1.358*	*1.280*
	MES					
		mahogany[c]	0.697	0.522	0.684	*0.684*
		zebrano	1.870	1.000	1.972	*1.870*
		median	*1.283*	*0.761*	*1.328*	*0.849*
Chelating agent	EGTA					
		mahogany[c]	1.071	1.169	1.000	*1.071*
		zebrano	1.450	0.440	1.178	*1.178*
		median	*1.260*	*0.804*	*1.089*	*1.120*
	EDDS					
		mahogany[c]	0.594	0.824	0.673	*0.673*
		zebrano	1.000	1.254	1.348	*1.254*
		median	*0.797*	*1.039*	*1.010*	*0.912*
	DTPA					
		mahogany[c]	0.966	1.313	0.760	*0.966*
		zebrano	1.025	1.103	1.169	*1.103*
		median	*0.995*	*1.208*	*0.964*	*1.064*
Detergent	CTAC					
		English oak	1.476	1.603	1.665	*1.603*
		mahogany[c]	3.898	1.542	1.026	*1.542*
		merbau	0.902	1.138	1.122	*1.122*
		zebrano	1.964	1.565	1.569	*1.569*
		wenge	1.032	0.699	1.104	*1.032*
		Baltic Pine	1.135	0.903	1.049	*1.049*
		median	*1.135*	*1.138*	*1.104*	*1.136*
	BAC					
		mahogany[c]	0.991	0.917	0.928	*0.928*
		zebrano	1.359	1.465	1.319	*1.359*
		median	*1.175*	*1.191*	*1.123*	*1.155*
	BZT					
		English oak	1.681	1.679	1.587	*1.679*
		mahogany[c]	1.101	1.023	0.985	*1.023*
		merbau	0.926	1.078	0.751	*0.926*
		zebrano	1.653	1.498	1.321	*1.498*
		Baltic Pine	1.263	0.862	0.948	*0.948*
		wenge	0.979	0.918	0.620	*0.918*
		median	*1.101*	*1.023*	*0.948*	*1.050*

			Reagent concentration			
Functional group	**Reagent**	**species**	**low**	**med**	**high**	*median*
Excipient	PVP-K30					
		mahogany[c]	0.737	1.630	0.724	*0.737*
		zebrano	1.037	-2.885	0.741	*0.741*
		median	*0.887*	*-0.628*	*0.732*	*0.739*
	PVPP					
		mahogany[c]	1.139	7.600	1.067	*1.139*
		zebrano	1.456	1.429	1.261	*1.429*
		median	*1.297*	*4.515*	*1.164*	*1.345*
	PEG 200					
		mahogany[c]	1.024	2.599	-4.527	*1.024*
		zebrano	1.367	1.458	1.356	*1.367*
		median	*1.196*	*2.028*	*-1.586*	*1.361*
Salt	Urisol					
		mahogany[c]	0.607	0.734	0.840	*0.734*
		zebrano	-7.645	2.016	2.944	*2.016*
		median	*-3.519*	*1.375*	*1.892*	*0.787*

Supplementary Table 14: A_{260}/A_{230} DNA purity ratio results for alternative BOTAB buffer reagents (*Experiment III*)

			Reagent concentration			
Functional group	**Reagent**	**species**	**low**	**med**	**high**	*median*
Biological antioxidant	BME					
		mahogany[C]	0.120	0.163	0.152	*0.152*
		zebrano	0.288	0.295	0.311	*0.295*
		median	*0.204*	*0.229*	*0.231*	*0.225*
	AA					
		mahogany[C]	0.720	0.632	0.976	*0.720*
		zebrano	2.147	1.712	1.848	*1.848*
		median	*1.434*	*1.172*	*1.412*	*1.344*
	NaA					
		mahogany[C]	0.083	0.269	1.282	*0.269*
		zebrano	2.498	2.054	2.460	*2.460*
		median	*1.290*	*1.161*	*1.871*	*1.668*
	TCEP					
		mahogany[C]	0.195	0.178	0.253	*0.195*
		zebrano	0.431	0.433	0.237	*0.431*
		median	*0.313*	*0.305*	*0.245*	*0.245*
Buffer	PIPES					
		mahogany[C]	0.223	0.1585	0.239	*0.223*
		zebrano	0.2695	0.255	0.1865	*0.255*
		median	*0.246*	*0.207*	*0.213*	*0.231*
	MES					
		mahogany[C]	0.1435	0.1425	0.161	*0.144*
		zebrano	0.192	0.1855	0.1705	*0.186*
		median	*0.168*	*0.164*	*0.166*	*0.166*
Chelating agent	EGTA					
		mahogany[C]	0.1865	0.211	0.2	*0.200*
		zebrano	0.2415	0.356	0.294	*0.294*
		median	*0.214*	*0.284*	*0.247*	*0.226*
	EDDS					
		mahogany[C]	0.114	0.09	0.05	*0.090*
		zebrano	0.006	0.1435	0.012	*0.012*
		median	*0.060*	*0.117*	*0.031*	*0.070*
	DTPA					
		mahogany[C]	0.2135	0.213	0.1265	*0.213*
		zebrano	0.3245	0.359	0.2995	*0.325*
		median	*0.269*	*0.286*	*0.213*	*0.257*
Detergent	CTAC					
		English oak	0.473	0.526	0.471	*0.473*
		mahogany[C]	0.160	0.260	0.305	*0.260*
		merbau	0.145	0.200	0.160	*0.160*
		zebrano	0.355	0.168	0.442	*0.355*
		wenge	0.247	0.238	0.271	*0.247*
		Baltic Pine	0.171	0.189	0.163	*0.171*
		median	*0.171*	*0.200*	*0.271*	*0.242*
	BAC					
		mahogany[C]	0.174	0.187	0.228	*0.187*
		zebrano	0.305	0.386	0.337	*0.337*
		median	*0.239*	*0.286*	*0.282*	*0.266*
	BZT					
		English oak	0.446	0.409	0.395	*0.409*
		mahogany[C]	0.267	0.224	0.149	*0.224*
		merbau	0.155	0.195	0.123	*0.155*
		zebrano	0.455	0.295	0.138	*0.295*
		Baltic Pine	0.174	0.153	0.089	*0.153*
		wenge	0.261	0.266	0.190	*0.261*
		median	*0.261*	*0.224*	*0.138*	*0.209*

Functional group	Reagent	species	Reagent concentration low	med	high	median
Excipient	PVP-K30					
		mahogany[c]	0.156	0.219	0.121	*0.156*
		zebrano	0.136	-0.062	0.085	*0.085*
		median	*0.146*	*0.079*	*0.103*	*0.129*
	PVPP					
		mahogany[c]	0.112	0.318	0.306	*0.306*
		zebrano	0.351	0.399	0.485	*0.399*
		median	*0.231*	*0.359*	*0.395*	*0.335*
	PEG 200					
		mahogany[c]	0.169	0.200	0.593	*0.200*
		zebrano	0.377	0.527	0.451	*0.451*
		median	*0.273*	*0.364*	*0.522*	*0.414*
Salt	Urisol					
		mahogany[c]	0.011	-0.025	0.064	*0.011*
		zebrano	-0.003	0.001	0.109	*0.001*
		median	*0.004*	*-0.012*	*0.086*	*0.006*

Experiment IV: Additional extraction buffer NaCl (salt) concentration experiments
The median A_{260}/A_{280} DNA purity ratios for both the merbau and oak samples was between 1.23 and 1.53, and always <1 for the A_{260}/A_{230} ratios. The purity scores from the salt test can be seen in Table S15. When the products were amplified, the strongest bands were recorded when using 1.0_M NaCl for both merbau and oak. There was also a strong band in merbau for 2.2_M NaCl samples. For both species, a weak band was detected when NaCl was not used (i.e. 0_M). For all other concentrations no bands were detected. All agarose gel scores can be seen in Table S16.

Supplementary table 15: median DNA purity from extraction buffer salt experiment (*Experiment IV*).

	salt (naCl) concentration (M) 0	1	1.4	1.8	2.2
DNA purity (A_{260}/A_{280} ratio)					
merbau	1.31	1.43	0.94	1.52	1.24
European white oak	1.54	1.64	1.52	1.49	1.21
median	*1.42*	*1.54*	*1.41*	*1.49*	*1.24*
DNA purity (A_{260}/A_{230} ratio)					
merbau	0.38	0.41	0.11	0.47	0.39
European white oak	0.51	0.88	0.88	0.89	1.76
median	*0.45*	*0.64*	*0.60*	*0.66*	*0.63*

Supplementary table 16: Agarose gel scores for range of extraction buffer Salt (NaCl) concentrations (*Experiment IV*)

concentration (M)	merbau	European white oak
0	+	+
1.0	++	++
1.4	-	-
1.8	-	-
2.2	++	-

NB: +/- represent agarose gel scoring (- no band detected, + weak band detected, ++ strong band detected)

Experiment V: Additional extraction buffer antioxidant experiments

From the initial extraction buffer reagent study (*Experiment III*) it was found that the A_{260}/A_{230} ratios of both ascorbic acid (AA) and sodium ascorbate (NaA) were the highest of the antioxidant reagents (AA = 1.344, NaA = 1.668 (see Table S14). This was the justification for conducting a more detailed experiment. From this experiment the highest median DNA purity ratios for AA was 75_{mM} for A_{260}/A_{280} (1.57) and 60_{mM} for A_{260}/A_{230} (2.08). For NaA only the merbau and larch readings were available (purity readings of the European white oak samples were not carried out). The highest A_{260}/A_{280} ratio was for 60_{mM} (1.47) and 75_{mM} for A_{260}/A_{230} (1.38). The median DTT scores for both ratios was much less than for all other concentrations/reagents including the 0_{mM} concentration (A_{260}/A_{280} = 0.85, A_{260}/A_{230} = 0.25). All median DNA purity results can be found in Table S17. When the extracts were amplified, the amplifications were most successful in AA for 15_{mM} and 45_{mM} concentrations for merbau and 30_{mM} and 60_{mM} for European white oak. For NaA, only weak products were detected for both species at 15_{mM}, 45_{mM}, and 60_{mM} for merbau and 15_{mM} and 30_{mM} for European European white oak. For both reagents no bands or peaks were seen for either 0_{mM} or 75_{mM} concentrations and also when DTT was used. The agarose gel scores of this experiment can be seen in Table S18.

Supplementary table 17: Median DNA purity reading for alternative extraction buffer antioxidant reagent at various concentrations (*Experiment V)*

reagent	**Ascorbic acid (AA)**					**DTT**	**None**	**Sodium ascorbate (NaA)**				
Concentration (mM)	**15**	**30**	**45**	**60**	**75**	**47**	**0**	**15**	**30**	**45**	**60**	**75**
DNA purity (A_{260}/A_{280} ratio)												
merbau	1.24	1.51	1.57	1.57	1.57	1.31	1.44	1.41	1.45	1.52	1.57	1.44
European white oak	1.06	1.47	1.50	1.52	1.58		3.44					
larch	0.88	1.40	1.41	1.38		0.48	0.52	0.53	0.90	1.02	1.31	1.35
median	*1.03*	*1.47*	*1.50*	*1.52*	*1.57*	*0.85*	*1.44*	*0.96*	*1.22*	*1.05*	*1.47*	*1.36*
DNA purity (A_{260}/A_{230} ratio)												
merbau	0.57	2.02	1.65	2.58	2.91	0.38	0.39	0.43	0.56	0.97	1.37	1.38
European white oak	0.20	0.80	0.95	1.06	1.27		0.38					
larch	1.85	1.55	1.84	2.08		0.14	0.18	0.46	0.56	0.78	1.26	1.38
median	*0.57*	*1.48*	*1.53*	*2.08*	*1.90*	*0.25*	*0.31*	*0.43*	*0.56*	*0.96*	*1.30*	*1.38*

Supplementary table 18: Agarose gel scores of alternative extraction buffer antioxidant reagents (*Experiment V*)

	merbau		European white oak	
Concentration (mM)	**AA**	**NaA**	**AA**	**NaA**
0	-		-	
15	++	+	-	+
30	+	-	++	+
45	++	+	-	-
60	+	+	++	-
75	-	-	-	-
DTT (47)	+		-	

NB: AA= ascorbic acid, NaA= sodium ascorbate. Number in brackets next to DTT is the concentration that reagent was used at in experiment (i.e. 47$_{mM}$). +/- represent agarose gel scores (- no band detected, + weak band detected, ++ strong band detected)

The second part of this experiment was to compare the four antioxidant reagents at a 45$_{mM}$ concentration. The DNA purity results were found to be similar to those in the first part of this experiment. The highest median DNA purity for both ratios was for AA and lowest for DTT (see Table S19 for all scores). There was a significant difference ($p < 0.05$) between the reagents for both ratio scores, with the post-hoc tests identifying pairwise differences between many of the reagents. The agarose gels from the amplified extracts found weak bands for both oak species only for DTT, while strong bands were detected in merbau for AA, NaA and DTT (see Table S20 for scores).

Supplementary table S19: Median DNA purity of extraction buffer antioxidant reagents at 45mM concentration (*Experiment V*).

sample	AA	BME	DTT	NaA	
DNA purity (A_{260}/A_{280} ratio)					*p value* <0.05
mahogany[C]	1.48	1.36	1.05	1.5	
merbau	1.58	1.36	1.16	1.55	
European white oak	1.53	-0.14	1.42	0.52	
larch	1.44	0.58	0.52	1.35	
mahogany[J]	1.5	1.24	0.97	1.47	
American white oak	1.47	2.15	1.18	1.65	
median	*1.51*	*1.25*	*1.13*	*1.48*	
	A**, C*, E*	C*	A**, B*, D**, E*	B*, D**	*Significant differences*
DNA purity (A_{260}/A_{230} ratio)					*p value* <0.05
mahogany[C]	1.61	0.47	0.37	1.12	
merbau	2.04	0.34	0.35	1.53	
European white oak	1	0.11	1.08	-0.11	
larch	1.84	0.21	0.29	1.45	
mahogany[J]	2.07	0.4	0.3	1.03	
American white oak	1.42	0.7	0.53	0.94	
median	*1.59*	*0.38*	*0.4*	*0.99*	
	F**, G**H*, I*, J*	F**, I*	G**, J*	H*	*Significant differences*

*NB: reagent shorthand's, AA= Ascorbic Acid, NaA= Sodium Ascorbate. Single letters indicate significant difference between the reagent pair identified from post-hoc tests, stars indicate level of significance (*p<0.05, **p<0.01). Significant differences identified from either Conovor (Holm) (pairs: A, B, F, G), Conovor (Benjaminyi-Hochberg) (pairs: A, C, D, F, G, H) or Nemenyi (pairs: E, I, J).*

Supplementary table 20: Agarose gel scores for the different extraction buffer antioxidant reagents at a common concentration (45mM) (*Experiment V*).

	Reagent		
sample	AA	NaA	DTT
European white oak	-	-	+
merbau	++	++	++
American white oak	-	-	+

NB: AA = ascorbic Acid, NaA = Sodium ascorbate. +/- represents agarose gel scoring (- no band detected, + weak band detected, ++ strong band detected)

Experiment VI: Extraction buffer incubation time experiments

For this experiment, results were only generated for the two oak samples as the amplification of merbau was unsuccessful. For American white oak weak bands were detected for three of the ten time periods (2, 4.5 and 5 hours). For the European white oak sample, bands were detected for nine of the ten time periods, with strong bands detected for 1, 2, 6, 7 and O/N time periods and weak bands detected for the 4-5.5hr times range. Only the 3-hour incubation time failed to amplify (see Table S21 for all scores).

Supplementary table 21: Agarose gel scores of various extraction buffer incubation times (*Experiment VI*).

sample

Incubation time (hours)	European white oak	American white oak
1	++	-
2	++	+
3	-	-
4	+	-
4.5	+	+
5	+	+
5.5	+	-
6	++	-
7	++	-
~16 (O/N)	++	-
48 (2 days)	X*	X*
72 (3 days)	X*	X*

NB: X samples not tested at that time period. +/- represents agarose gel scoring (- no band detected, + weak band detected, ++ strong band detected)*

Experiment VII: Extraction buffer incubation temperature experiments

Only the DNA purity scores could be reported for this experiment. The amplification and subsequent sequencing revealed no usable results. For the DNA purity scores, both the median A_{260}/A_{280} and A_{260}/A_{230} ratios were similar between the three incubation temperatures. Additionally, no detectable trend was evident (see Table S22 for DNA purity scores).

Supplementary table 22: DNA purity scores for three alternative extraction buffer incubation temperatures (*Experiment VII*)

	Temperature (°C)		
sample	RT (~25)	40	64
DNA purity (A_{260}/A_{280} ratio)			
mahogany[C]	1.43	1.47	1.48
merbau	1.48	1.52	1.44
European white oak	1.36	1.42	1.34
American white oak	1.44	1.45	1.47
median	*1.44*	*1.47*	*1.46*
DNA purity (A_{260}/A_{230} ratio)			
mahogany[C]	0.60	0.60	0.63
merbau	0.85	0.77	1.02
European white oak	0.48	0.64	0.48
American white oak	0.42	0.46	0.44
median	*0.54*	*0.60*	*0.49*

NB: RT = room temperature

Experiment VIII: Precipitation solution salt alternatives experiments

The experiment into alternative precipitation solution salts was unable to yield any conclusive results. For European white oak the DNA purity was more within the good quality range DNA at the A_{260}/A_{280} ratio (1.7-2) for all three precipitation salts and most concentrations, with potassium acetate (KOAc) at 0.57_M (2.34) and 1_M (2.59), lithium chloride (LiCl) at 0.57_M (2.53) and 1_M (4.39) being the exceptions. For merbau the ratios were low for all samples at both ratios (See Table S23 for purity scores for both oak and merbau). The

amplifications were mostly unsuccessful. For merbau, a weak band was only detected for two salts for one concentration each, LiCl at 1_M and Ammonium acetate (NH_4OAc) at 2.1_M. For oak, only one band was detected for all variables, a strong band for KOAc at 0.8_M (see Table S24 for amplification results for both species).

Supplementary table 23: DNA purity ratios for alternative precipitation solution salt reagents (*Experiment VIII*)

reagent	NH_4OAc			LiCl			KOAc		
concentration (M)	2.1	2.5	2.8	0.57	0.8	1	0.57	0.8	1
DNA purity (A_{260}/A_{280} ratio)									
merbau	1.14	1.14	1.30	1.40	1.30	1.20	1.20	1.25	1.31
European white oak	1.72	1.92	1.90	2.53	1.75	4.39	2.34	1.75	2.59
median	*1.14*	*1.20*	*1.36*	*1.42*	*1.32*	*1.28*	*1.24*	*1.29*	*1.36*
DNA purity (A_{260}/A_{230} ratio)									
merbau	0.33	0.31	0.37	0.42	0.37	0.35	0.32	0.35	0.37
European white oak	0.32	0.55	0.53	1.81	0.40	5.46	1.77	0.28	1.92
median	*0.32*	*0.32*	*0.39*	*0.42*	*0.39*	*0.38*	*0.32*	*0.31*	*0.39*

Supplementary table 24: Agarose gel scores for alternative precipitation solution salts (*Experiment VIII*)

reagent	NH_4OAc			LiCl			KOAc		
concentration (M)	2.1	2.5	2.8	0.57	0.8	1	0.57	0.8	1
Merbau	+	-	-	-	-	+	-	-	-
European White Oak	-	-	-	-	-	-	-	++	-

NB: +/- represents agarose gel scoring (- no band detected, + weak band detected, ++ strong band detected)

Experiment IX: Precipitation solution incubation time and temperature experiments

For this experiment, it was found that while incubation of the precipitation solution at -20°C did not need to consider the freezing times of the buffer, it was important when at -80°C, with samples incubated at this temperature freezing after ~1hour. Samples incubated for the 2hr, 4hr and O/N time periods had to be defrosted in the fridge (+4°C) until thawed, before proceeding. The agarose gel for mahogany revealed that incubations worked for both incubation temperatures. For -20°C weak bands were seen for 2 and 4hr times and a strong band for the O/N incubation and strong bands were also seen for two incubation times when incubating at -80°C (0.5hr and O/N). See Table S25 for agarose gel scores

Supplementary table 25: Agarose gel amplification scores for incubation time and temperature testing of the precipitation solution for mahogany[C] timber sample.

	Incubation temperature (°C)	
Incubation time (hrs)	-20	-80

0.5	-	++
1	-	-
2	+	-
4	+	-
~16 (O/N)	++	++

NB: +/- represents agarose gel scoring (- no band detected, + weak band detected, ++ strong band detected)

Experiment X: Comparison of BOTAB protocol to commercial kits

For the most part the DNA yields from the samples was low. Apart from bigleaf maple (for all protocols) and jarrah (for both BOTAB extractions) the DNA concentration was $<1_{ng/\mu l}$. This resulted in the median DNA concentrations for all protocols being low ($<0.21_{ng/\mu l}$), with both BOTAB extractions having the highest medians ($0.21_{ng/\mu l}$ (initial) & $0.19_{ng/\mu l}$ (final)). The median DNA purity ratios between the protocols identified that both BOTAB extractions had the purest A_{260}/A_{280} ratios (2.50 (initial) & 1.49 (final)) compared to the three commercial kits (<0.58). The A_{260}/A_{230} ratios identified that the opposite was occurring. The median ratio was highest in the Qiagen kit (1.38) and <1 for all other protocols. The lowest median ratios were for the two BOTAB extractions (0.02 (initial) & 0.04 (final)). See Figure 1 and Table S26 for DNA concentration results and median readings.

Supplementary table 26: Median DNA concentration (ng/ul) and purity results for the different extraction protocols (*Experiment X*)

	Extraction protocol				
Sample	BOTAB (initial)	Jena Kit	Qiagen Kit	MoBio Kit	BOTAB (final)
DNA concentration (ng/µl)					
jarrah	3.16	0.90	0.08	0.83	5.30
english oak	0.03	0.06	0.02	0.01	-
mahogany[C]	0.21	0.28	0.15	0.27	0.20
wenge	0.01	0.03	0.01	0.02	0.03
Baltic pine	0.05	0.02	0.02	0.04	0.13
bigleaf maple	12.83	19.00	2.54	3.40	25.50
median	*0.21*	*0.15*	*0.08*	*0.04*	*0.19*
negative extraction	0.00	0.01	0.01	0.01	0.01
DNA purity (A_{260}/A_{280} ratio)					
jarrah	2.83	0.55	0.41	0.74	2.10
english oak	1.92	0.71	0.41	0.39	-
mahogany[C]	2.22	0.58	0.41	0.47	1.35
wenge	2.85	0.54	0.38	0.41	1.44
Baltic pine	2.70	0.49	0.41	0.33	2.41
bigleaf maple	2.30	1.54	0.46	0.55	1.34
median	*2.50*	*0.58*	*0.41*	*0.44*	*1.46*
negative extraction	4.21	0.45	0.40	0.36	1.78
DNA purity (A_{260}/A_{230} ratio)					
jarrah	0.03	0.82	1.38	0.46	0.04
english oak	0.02	0.70	1.19	1.12	-
mahogany[C]	0.02	0.51	1.34	0.61	0.02
wenge	0.03	0.72	2.82	0.84	0.04
Baltic pine	0.02	1.79	1.48	2.56	0.02
bigleaf maple	0.04	2.33	1.55	0.84	0.05
median	*0.02*	*0.84*	*1.38*	*0.78*	*0.04*
negative extraction	0.03	0.52	2.07	2.58	0.04

References

[SWGDAM] Scientific Working Group for DNA Analysis Methods (2016). "SWGDAM Validation Guidelines for DNA Analysis Methods." https://docs.wixstatic.com/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf.

[SWFS] Society for Wildllife Forensic Sciences Technical Working Group (2018) Standards and Guidelines for Wildlife Forensic Analysis, Version 3. Ed. Lucy M.I. Webster. Published by the Society for Wildlife Forensic Science, 19th November 2018, pp.21. https://www.wildlifeforensicscience.org/wp-content/uploads/2018/11/SWFS-Standards-and-Guidelines_Version-3_19-11-18.pdf

Alexander L (2016) Rapid, effective DNA isolation from osmanthus via modified alkaline lysis. Journal of Biomolecular Techniques 27:53 doi:10.7171/jbt.16-2702-001

Bürgmann, H., et al. (2001). "A strategy for optimizing quality and quantity of DNA extracted from soil." Journal of microbiological methods **45**(1): 7-20.

Colpaert, N., et al. (2005). "Sampling tissue for DNA analysis of trees: trunk cambium as an alternative to canopy leaves." Silvae Genetica **54**(6): 265-269.

Csaikl, U., et al. (1998). "Comparative analysis of different DNA extraction protocols: a fast, universal maxi-preparation of high quality plant DNA for genetic evaluation and phylogenetic studies." Plant Molecular Biology Reporter **16**(1): 69-86.

Dellaporta, S. L., et al. (1983). "A plant DNA minipreparation: version II." Plant Molecular Biology Reporter **1**(4): 19-21.

Jiao, L., et al. (2014). "DNA barcoding for identification of the endangered species Aquilaria sinensis: comparison of data from heated or aged wood samples." Holzforschung **68**(4): 487-494.

Jiao, L., et al. (2012). "Comparative analysis of two DNA extraction protocols from fresh and dried wood of *Cunninghamia lanceolata* (*Taxodiaceae*)." IAWA Journal **33**(4): 441-456.

Matlock, B. and Thermo Fisher Scientific (2015). Assessment of Nucleic Acid Purity. Technical Note 52646. Wilmington, MA, USA, Thermo Fisher Scientific.

Pohlert, T. (2014). The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR). R package.

Rachmayanti, Y., et al. (2006). "Extraction, amplification and characterization of wood DNA from Dipterocarpaceae." Plant Molecular Biology Reporter **24**(1): 45-55.

Rachmayanti, Y., et al. (2009). "DNA from processed and unprocessed wood: factors influencing the isolation success." Forensic Science International: Genetics **3**(3): 185-192.

Särkinen, T., et al. (2012). "How to open the treasure chest? Optimising DNA extraction from herbarium specimens." PloS one **7**(8): e43808.

Tan, S. C. and B. C. Yiap (2009). "DNA, RNA, and protein extraction: the past and the present." Journal of Biomedicine and Biotechnology **2009**.

Telle, S. and M. Thines (2008). "Amplification of *cox2* (~ 620 bp) from 2 mg of up to 129 years old herbarium specimens, comparing 19 extraction methods and 15 polymerases." PloS one **3**(10): e3584.
Varma, A., et al. (2007). "Plant genomic DNA isolation: an art or a science." Biotechnology Journal **2**(3): 386-392.
Vasavada, N. (2016). "Friedman test for correlated multiple samples." Online Web Statistical Calculators for Categorical Data Analysis. from http://astatsa.com/FriedmanTest/.
Verbylaite, R., et al. (2010). "Comparison of Ten DNA Extraction Protocols from Wood of European Aspen (Populus tremula L.)." BALTIC FORESTRY **16**(1): 35-42.

Chapter 3: Development of novel SNP markers for bigleaf maple and ayous

Statement of Authorship

Title of Paper	*A set of 204 SNP and INDEL markers for bigleaf maple (Acer macrophyllum Pursch)*
Publication Status	Published
Publication Details	Conservation Genetic Resources 7(4) 797-801, DOI: 10.1007/s12686-015-0486-7

Principal Author

Name of Principal Author (Candidate)	Duncan Iain Jardine		
Contribution to the Paper	assisted with sample preparation; conducted lab work & data analysis; conducted literature review; co-wrote manuscript as co-principal author		
Overall percentage (%)	40%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	12.08.2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

iv. the candidate's stated contribution to the publication is accurate (as detailed above);

v. permission is granted for the candidate in include the publication in the thesis; and

vi. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Eleanor E Dormontt		
Contribution to the Paper	organised sample collections; assisted with sample preparation; assisted with data analysis; co-wrote manuscript as co-principal author 35%		
Signature		Date	05.08.2019

Name of Co-Author	*Kor-Jent Van Dijk*		
Contribution to the Paper	conducted parts of the labwork & data analysis; planning of manuscript theme; edited manuscript 10%		
Signature		Date	16.07.2019

Name of Co-Author	*Rainbo RM Dixon*		
Contribution to the Paper	assisted with sample preparation; planning of manuscript theme; edited manuscript 5%		
Signature		Date	30.07.2019

Name of Co-Author	*Bianca Dunker*		
Contribution to the Paper	assisted with sample preparation; planning of manuscript theme; edited manuscript 5%		
Signature	B. D…	Date	12.12.18

Name of Co-Author	*Andrew J Lowe*		
Contribution to the Paper	as principal supervisor; obtained funding for the research; supervised experiments; advised on focus & structure of manuscript; edited manuscript 5%		
Signature		Date	10.07.2019

Conservation Genet Resour
DOI 10.1007/s12686-015-0486-7

TECHNICAL NOTE

A set of 204 SNP and INDEL markers for Bigleaf maple (*Acer macrophyllum* Pursch)

D. I. Jardine[1] · E. E. Dormontt[1] · K.-J. van Dijk[1] · R. R. M. Dixon[1] · B. Dunker[1] · A. J. Lowe[1]

Received: 1 August 2015 / Accepted: 12 August 2015

Abstract Novel single nucleotide polymorphisms (SNPs) and insertion/deletions (INDELs) were identified for the Bigleaf maple (*Acer macrophyllum* Pursch) using a combined next generation sequencing approach on the Ion Torrent PGM system with genotyping on the MassARRAY® iPLEX™ platform. Five hundred and ninety-eight putative loci were identified through sequencing of DNA fragments following a double restriction enzyme digest method. Two hundred and four polymorphic loci (199 SNPs and five INDELs) were successfully amplified across 65 individuals from seven populations across the native range of the species. Twenty-nine loci showed evidence of deviation from Hardy–Weinberg Equilibrium, and 85 were significantly linked. Expected heterozygosity ranged from 0.015 to 0.472 and F_{ST} from 0.011 to 0.359. These genetic resources will prove useful for future studies into the population genetics and phylogeography of this important and iconic timber species.

Keywords Single nucleotide polymorphism · Insertion/deletion · MassARRAY

Bigleaf maple (*Acer macrophyllum* Pursch) is an economically important hardwood tree species from the Pacific Northwest of North America. It is most distinguishable by its short stature (just 15–30 m tall) and aptly given common name, with leaves often 30 cm across at maturity. Bigleaf maple is found in coastal regions, within ~300 km of the ocean and often associated with streams and rivers (Minore and Zasada 1990; Iddrisu and Ritland 2004). The timber from Bigleaf maple is commonly used to make furniture, and trees that produce a 'figured' grain pattern are increasingly sought after for musical instruments.

Previous population genetic analysis of Bigleaf maple has used isozymes (Iddrisu and Ritland 2004); we sought to develop a suite of genetic markers suitable for modern population genetic and phylogeographic analyses to enable more detailed determination of the dynamics of natural Bigleaf maple populations. Here we present a set of 204 genetic markers, comprising 199 polymorphic single nucleotide polymorphisms (SNPs) and five insertion/deletions (INDELs) for Bigleaf maple.

For marker discovery, 31 individuals from 18 geographic locations (Table S1) were sequenced; DNA was extracted from cambium using the Nucleospin Plant II Kit (Machery-Nagel, Düren, Germany) with the PL2/PL3 buffer system. Reduced representation libraries were prepared using a modified double restriction enzyme digest method (Vos et al. 1995; van Orsouw et al. 2007). Extracted DNA was diluted to 20 ng/μL then digested with restriction enzymes *Eco*RI-HF® and *Mse*I (New England Biolabs, Beverly, MA). To each digested sample, double stranded adapters (*Eco*RI and *Mse*I) and T4 ligase (New England Biolabs) were added and incubated overnight. Initial pre-selective PCR (25 cycles) used an optimized polymerase for difficult templates (DyNAzyme™ EXT

D. I. Jardine and E. E. Dormontt have contributed equally.

Electronic supplementary material The online version of this article (doi:10.1007/s12686-015-0486-7) contains supplementary material, which is available to authorized users.

✉ A. J. Lowe
andrew.lowe@adelaide.edu.au

[1] School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

Published online: 19 August 2015

DNA polymerase, Finnzyme, Thermo Fisher Scientific, Waltham, MA) and the ligated adapters (*Eco*RI +A and *Mse*I +C) as priming sites, with the additional 'selective' base added to reduce the amplicon pool to ~1/16th. The second PCR (25 cycles) used fusion primers and Amplitaq Gold® (Life Technologies, Carlsbad, CA) polymerase. The fusion primer consists of the sequence used in the first PCR followed by an additional 6 bp multiplex identifier sequence or MID tag and the sequencing specific primer keys Ion A and Ion P1 that attach to the 5′ end. PCR products were pooled and purified using AMPure™ XP (Agencourt, Beckman Coulter, Inc., Brea, CA) to remove leftover primer and primer dimers. The library was optimised for the Ion Torrent PGM Sequencing 400 Kit (Life Technologies) by selecting 350–400 bp products with either a Pippin Prep™ (Sage Science, Beverly, MA) or an E-Gel® (Life Technologies). The size-selected amplification pool was quantified using a 2200 TapeStation™ (Agilent, Santa Clara, CA) with the High-Sensitivity D1000 ScreenTape. These quantifications were used to adjust the library dilution to 9–14 pmol/L. Emulsion PCR and enrichment were conducted according to manufacturer's specifications. Sequencing followed on the Ion Torrent PGM™ (Life Technologies). Sequencing reads were analysed using CLC-Genomic Workbench (Qiagen, Venlo, The Netherlands). Reads were de-multiplexed, trimmed and assembled to generate a 'Provisional Reference Genome' or PRG (Hird et al. 2011). Each sample's individual reads were mapped onto the PRG and consensus sequences with ambiguity codes were extracted. Results were exported into Geneious R6 (Biomatters, Auckland, New Zealand) for manual selection of loci. All sequences containing a putative locus were checked for significant sequence commonality with non-plant species using BLAST capability of GenBank (NCBI). Only loci suitable for the MassARRAY® iPLEX™ platform (Agena Bioscience™, San Diego, USA) were selected, which requires a single target variable site located within ~100 bp of minimally variable sequence (Oeth et al. 2005; Gabriel et al. 2009). Primers and multiplex assays were designed in Assay Design Suite (ADS) (Agena Bioscience™). Sixty-five individuals from seven different geographic locations (Table 1) were genotyped using iPLEX™ GOLD chemistry (Agena Bioscience™). DNA from four individuals was independently extracted and genotyped twice to determine repeatability. Genotypes were checked manually for duplicate identity; heterozygosity and Wright's fixation index were calculated in GENALEX v6.5 (Peakall and Smouse 2006, 2012). Evidence for linkage disequilibrium and global deviations from Hardy–Weinberg Equilibrium were tested using GENEPOP v4.2 (Rousset 2008). No corrections for multiple tests were applied.

Table 1 Population genetic parameters of markers developed in this study

Locus	H_O	H_E	F_{ST}	HWE
AM1f_1134	0.368	0.413	0.166	Yes
AM1f_1165	0.203	0.381	0.062	No****
AM1f_1228	0.462	0.449	0.091	Yes
AM1f_1283	0.398	0.393	0.141	Yes
AM1f_1351	0.202	0.209	0.167	Yes
AM1f_154	0.032	0.030	0.042	Yes
AM1f_2_2725	0.312	0.270	0.110	Yes
AM1f_2_389	0.248	0.318	0.290	Yes
AM1f_215	0.203	0.233	0.192	Yes
AM1f_2168	0.270	0.308	0.025	Yes
AM1f_2335	0.371	0.403	0.027	Yes
AM1f_287	0.078	0.068	0.087	Yes
AM1f_3270	0.449	0.412	0.150	Yes
AM1f_441	0.124	0.159	0.047	No****
AM1f_453	0.090	0.080	0.078	Yes
AM1f_481	0.229	0.234	0.327	Yes
AM1f_5	0.314	0.318	0.324	Yes
AM1f_5144	0.370	0.378	0.216	Yes
AM1f_524	0.171	0.213	0.077	No*
AM1f_5331	0.402	0.408	0.061	Yes
AM1f_5928	0.205	0.163	0.296	Yes
AM1f_924	0.322	0.302	0.236	Yes
AM1f_9252	0.219	0.268	0.065	No*
AM1f_927	0.343	0.293	0.186	Yes
AM1f_984	0.416	0.390	0.093	Yes
AM2f_176	0.227	0.399	0.099	No***
AM2f_18	0.131	0.133	0.130	Yes
AM2f_2_1164	0.263	0.328	0.341	No**
AM2f_2_123	0.127	0.118	0.191	No*
AM2f_2_292	0.079	0.064	0.155	Yes
AM2f_214	0.370	0.308	0.034	Yes
AM2f_218	0.463	0.472	0.054	Yes
AM2f_234	0.402	0.356	0.042	Yes
AM2f_290	0.462	0.460	0.077	Yes
AM2f_346	0.475	0.404	0.114	Yes
AM2f_49	0.479	0.444	0.068	Yes
AM2f_58	0.381	0.305	0.011	Yes
AM2f_617	0.413	0.370	0.248	Yes
AM2f_629	0.075	0.066	0.075	Yes
AM2f_657	0.478	0.432	0.135	Yes
AM2f_736	0.290	0.329	0.109	Yes
AM2f_9	0.454	0.382	0.143	Yes
Maple_10588	0.608	0.441	0.116	No**
Maple_1086	0.421	0.405	0.179	Yes
Maple_10862	0.262	0.273	0.257	Yes
Maple_1191_e	0.298	0.284	0.168	Yes
Maple_121	0.525	0.419	0.085	Yes
Maple_12182	0.229	0.261	0.107	Yes

Table 1 continued

Locus	H_O	H_E	F_{ST}	HWE
Maple_13	0.298	0.398	0.087	No*
Maple_13_bis	0.106	0.091	0.095	Yes
Maple_1308	0.288	0.283	0.200	Yes
Maple_1322	0.391	0.353	0.144	Yes
Maple_13253	0.184	0.300	0.094	No**
Maple_1389	0.186	0.169	0.116	Yes
Maple_1481	0.187	0.177	0.091	Yes
Maple_1489	0.517	0.429	0.114	Yes
Maple_1557	0.449	0.388	0.022	Yes
Maple_1569	0.095	0.104	0.125	Yes
Maple_1607	0.125	0.108	0.082	Yes
Maple_1643	0.276	0.285	0.058	Yes
Maple_1665	0.221	0.388	0.160	No***
Maple_1699	0.398	0.396	0.132	Yes
Maple_1752	0.292	0.342	0.267	Yes
Maple_1854	0.466	0.431	0.104	Yes
Maple_1856	0.335	0.292	0.083	Yes
Maple_1906	0.324	0.333	0.043	Yes
Maple_20	0.127	0.102	0.140	Yes
Maple_2059	0.108	0.113	0.133	Yes
Maple_2074	0.165	0.162	0.090	Yes
Maple_2076	0.125	0.134	0.072	Yes
Maple_2103	0.336	0.446	0.108	No*
Maple_2109	0.362	0.316	0.104	Yes
Maple_2138	0.416	0.391	0.089	Yes
Maple_2155	0.346	0.344	0.300	Yes
Maple_2394	0.048	0.045	0.033	Yes
Maple_24	0.203	0.319	0.087	No**
Maple_2417	0.241	0.240	0.143	Yes
Maple_2420	0.248	0.341	0.126	No*
Maple_2760	0.195	0.173	0.142	Yes
Maple_2793	0.505	0.410	0.136	Yes
Maple_2828	0.125	0.141	0.359	Yes
Maple_305	0.383	0.432	0.080	Yes
Maple_3075	0.402	0.445	0.090	Yes
Maple_3089	0.476	0.397	0.058	Yes
Maple_3090	0.332	0.309	0.074	Yes
Maple_3120	0.298	0.264	0.102	Yes
Maple_3136	0.400	0.333	0.011	Yes
Maple_3234	0.111	0.128	0.198	No**
Maple_3252	0.365	0.359	0.173	Yes
Maple_3258	0.043	0.036	0.131	Yes
Maple_3748	0.322	0.302	0.236	Yes
Maple_3773	0.203	0.226	0.112	Yes
Maple_3784	0.448	0.407	0.121	Yes
Maple_3814	0.278	0.221	0.273	Yes
Maple_3882	0.398	0.328	0.029	No*
Maple_3918	0.305	0.266	0.051	Yes

Table 1 continued

Locus	H_O	H_E	F_{ST}	HWE
Maple_3941	0.156	0.163	0.040	Yes
Maple_3953	0.111	0.123	0.076	Yes
Maple_3989	0.371	0.424	0.139	Yes
Maple_3999	0.380	0.418	0.105	Yes
Maple_4002	0.408	0.355	0.091	Yes
Maple_4034	0.232	0.193	0.156	Yes
Maple_4044	0.478	0.471	0.026	Yes
Maple_4049	0.395	0.378	0.126	Yes
Maple_4050	0.075	0.069	0.036	Yes
Maple_4074	0.462	0.461	0.071	Yes
Maple_4091	0.292	0.246	0.147	Yes
Maple_4138	0.416	0.406	0.065	Yes
Maple_4144	0.324	0.326	0.180	Yes
Maple_4174	0.184	0.180	0.067	Yes
Maple_4186	0.270	0.302	0.292	Yes
Maple_4218	0.322	0.326	0.060	Yes
Maple_4229	0.489	0.395	0.093	Yes
Maple_4258	0.308	0.408	0.168	No*
Maple_4278	0.336	0.310	0.088	Yes
Maple_4297	0.233	0.277	0.296	Yes
Maple_4308	0.205	0.160	0.130	Yes
Maple_4318	0.205	0.206	0.114	Yes
Maple_4381	0.333	0.373	0.240	Yes
Maple_4385	0.475	0.427	0.106	Yes
Maple_4393	0.481	0.421	0.130	Yes
Maple_4416	0.343	0.303	0.053	Yes
Maple_4438	0.092	0.105	0.095	Yes
Maple_4444	0.313	0.329	0.071	Yes
Maple_4455	0.294	0.302	0.036	Yes
Maple_4456	0.355	0.301	0.181	Yes
Maple_4472	0.168	0.135	0.125	Yes
Maple_4484	0.276	0.292	0.167	Yes
Maple_4512	0.202	0.246	0.332	Yes
Maple_4514	0.416	0.365	0.153	Yes
Maple_4566	0.276	0.340	0.147	No*
Maple_4604	0.459	0.406	0.110	Yes
Maple_4663	0.411	0.448	0.062	Yes
Maple_4665	0.537	0.434	0.120	Yes
Maple_4679	0.221	0.214	0.123	Yes
Maple_4693	0.219	0.199	0.095	Yes
Maple_4696	0.452	0.412	0.079	Yes
Maple_4702	0.827	0.466	0.040	No****
Maple_4704	0.140	0.120	0.076	Yes
Maple_4723	0.220	0.224	0.077	Yes
Maple_4724	0.356	0.383	0.112	Yes
Maple_4731	0.356	0.292	0.059	Yes
Maple_4803	0.494	0.391	0.107	Yes
Maple_4829	0.374	0.389	0.196	Yes

Table 1 continued

Locus	H_O	H_E	F_{ST}	HWE
Maple_4840	0.232	0.237	0.056	Yes
Maple_4847	0.290	0.353	0.100	No*
Maple_4850	0.284	0.254	0.256	Yes
Maple_4896	0.046	0.044	0.032	Yes
Maple_4902	0.076	0.065	0.108	Yes
Maple_4906	0.563	0.464	0.064	Yes
Maple_4920	0.119	0.141	0.151	Yes
Maple_4923	0.436	0.396	0.106	Yes
Maple_499	0.016	0.015	0.048	Yes
Maple_4998	0.478	0.432	0.135	Yes
Maple_5062	0.312	0.405	0.186	No*
Maple_5066	0.184	0.405	0.169	No****
Maple_5092	0.337	0.354	0.243	Yes
Maple_5095	0.349	0.321	0.235	Yes
Maple_5112	0.650	0.462	0.064	No**
Maple_5227	0.552	0.441	0.034	No*
Maple_5231	0.343	0.320	0.065	Yes
Maple_5287	0.470	0.375	0.203	No*
Maple_5345	0.032	0.028	0.097	Yes
Maple_5380	0.344	0.327	0.159	Yes
Maple_5418	0.389	0.421	0.158	Yes
Maple_5421	0.341	0.361	0.147	Yes
Maple_5463	0.449	0.422	0.123	Yes
Maple_5646	0.299	0.370	0.173	No*
Maple_57	0.317	0.338	0.029	Yes
Maple_5761	0.257	0.305	0.159	Yes
Maple_5762	0.468	0.468	0.050	Yes
Maple_5820	0.417	0.395	0.147	Yes
Maple_5940	0.256	0.374	0.195	No**
Maple_6002	0.262	0.357	0.255	No*
Maple_6157	0.259	0.217	0.184	Yes
Maple_6246	0.444	0.413	0.141	Yes
Maple_6317	0.402	0.403	0.061	Yes
Maple_6318	0.322	0.311	0.109	Yes
Maple_6339	0.513	0.436	0.078	Yes
Maple_65	0.511	0.443	0.103	Yes
Maple_6560	0.449	0.412	0.150	Yes
Maple_6578	0.517	0.421	0.112	Yes
Maple_659	0.448	0.330	0.151	Yes
Maple_6626	0.286	0.312	0.104	Yes
Maple_6682	0.430	0.430	0.105	Yes
Maple_6688	0.376	0.319	0.012	Yes
Maple_679	0.345	0.416	0.146	Yes
Maple_75	0.151	0.197	0.163	Yes
Maple_7509	0.462	0.380	0.217	Yes
Maple_7588	0.297	0.312	0.114	Yes
Maple_7702	0.356	0.307	0.063	Yes
Maple_7772	0.484	0.348	0.051	No****
Maple_7856	0.110	0.150	0.072	Yes
Maple_820	0.516	0.414	0.092	Yes
Maple_823	0.356	0.410	0.115	Yes
Maple_8509	0.352	0.339	0.071	Yes
Maple_8688	0.416	0.350	0.244	Yes
Maple_886	0.317	0.283	0.176	Yes
Maple_889	0.192	0.160	0.077	Yes
Maple_9048	0.189	0.148	0.244	Yes
Maple_9291	0.195	0.199	0.200	Yes
Maple_937	0.432	0.450	0.083	Yes
Maple_974	0.386	0.413	0.110	Yes
Maple_9828	0.476	0.440	0.066	Yes

H_O observed heterozygosity, H_E expected heterozygosity, F_{ST} Wright's fixation index, *HWE* Hardy–Weinberg Equilibrium; asterisks refer to significance of deviation from HWE, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$

A total of 598 potential loci were selected. There were no significant matches to non-plant organisms found through BLAST searching. MassARRAY primers were successfully designed for 491 loci and tested on 65 samples from seven populations collected across the native range (Table S2). Of 491 loci, 204 successfully amplified and were polymorphic (Table 1, S3). The duplicate samples showed complete congruence except for one replicate at one locus where the genotype was heterozygous in one sample and homozygous in the other. The overall error rate per locus was 0.08 %. Expected heterozygosity (H_E) ranged from 0.015 to 0.472 and F_{ST} from 0.011 to 0.359. Twenty-nine loci showed evidence of deviation from Hardy–Weinberg Equilibrium (Table 1) and significant linkage ($P < 0.05$) was found among 85 of the markers (Table S3). Our report details the first development of SNP and INDEL markers in the Bigleaf maple, *A. macrophyllum* and will prove useful for future genetic studies in the species.

Acknowledgments This work was supported by the World Resources Institute through Double Helix Tracking Technologies. We would like to thank Anne Minden and Jason O'Brien for their collection efforts, and gratefully acknowledge the help of United State Forest Services staff Shelley Gardner, Valerie Hipkins, Ron Malamphy, Phil Huff and Mark Camisa.

References

Gabriel S, Ziaugra L, Tabbaa D (2009) SNP genotyping using the Sequenom MassARRAY iPLEX platform. Current Protocols in Human Genetics, Chapter 2, Unit 2.12

Hird SM, Brumfield RT, Carstens BC (2011) PRGmatic: an efficient pipeline for collating genome-enriched second-generation sequencing data using a 'provisional-reference genome'. Mol Ecol Resour 11:743–748

Iddrisu MN, Ritland K (2004) Genetic variation, population structure, and mating system in bigleaf maple (*Acer macrophyllum* Pursh). Can J Bot 82:1817–1825

Minore D, Zasada JC (1990) *Acer macrophyllum* Pursh: Bigleaf maple. *Aceraceae*. Maple family. In: Burns RM, Honkala BH (eds) Silvics of forest trees of the United States. USDA forest service agriculture handbook 654, pp 33–40

Oeth P, Beaulieu M, Park C, Kosman D, del Mistro G, van den Boom D, Jurinke C (2005) iPLEX assay: increased plexing efficiency and flexibility for MassArray system through single base primer extension with mass-modified terminators. Sequenom Application Note 8876–8006

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research: an update. Bioinformatics 28:2537–2539

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour 8:103–106

van Orsouw NJ, Hogers R, Janssen A, Yalcin F, Snoeijers S, Verstege E, Schneiders H, van der Poel H, van Oeveren J, Verstegen H, van Eijk MJT (2007) Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. PLoS ONE 2:e1172

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Table S1. Locations and numbers of individuals of *Acer macrophyllum* used for SNP and INDEL marker discovery on the Ion Torrent PGM™ sequencing platform.

Location	# Individuals	Latitude	Longitude
Abbortsford, BC, Canada	2	49.0767	-122.2378
Rockport State Park, WA, USA	2	48.4915	-121.6184
Dosewallips State Park, WA, USA	2	47.6905	-122.9048
Mason County, WA, USA	2	47.2612	-123.4807
Humptuplis, WA, USA	2	47.2477	-123.8899
Grays Harbor County, WA, USA	2	47.1884	-123.5617
Capitol State Forest, Elma, WA, USA	1	46.9981	-123.0428
Mt. Baker-Snoqualmie National Forest, WA, USA	1	46.7615	-121.9581
Gifford Pinchot National Forest, WA, USA	2	46.6927	-121.5472
Gifford Pinchot National Forest, WA, USA	2	46.6367	-121.6245
Gifford Pinchot National Forest, WA, USA	2	46.5831	-121.7266
Lewes County, WA, USA	1	46.4653	-122.1729
Gifford Pinchot National Forest, WA, USA	2	46.4566	-121.7962
Lewes County, WA, USA	2	46.4417	-121.9956
Gifford Pinchot National Forest, WA, USA	1	46.4413	-121.7692
Gifford Pinchot National Forest, WA, USA	2	46.4325	-121.8362
Gifford Pinchot National Forest, WA, USA	1	46.4322	-121.9202

Table S2. Locations and numbers of individuals of *Acer macrophyllum* used for SNP and INDEL marker development on the MassARRAY® iPLEX™ platform (Agena Bioscience™).

Location	# Individuals	Latitude	Longitude
Abbortsford, BC, Canada	9	49.0767	-122.2378
Rockport State Park, WA, USA	10	48.4915	-121.6184
Humptuplis, WA, USA	9	47.2477	-123.8899
Gifford Pinchot National Forest, WA, USA	10	46.4325	-121.8362
Mt. Hood National Forest, OR, USA	9	45.4462	-122.1585
McDonald-Dunn Forest, OR, USA	9	44.6427	-123.3242
Humboldt County, CA, USA	9	40.8962	-123.9247

Table S3. Details of the genetic markers developed in this study.

Locus	Marker type	Alleles	Forward primer sequence	Reverse primer sequence	Linked loci
AM1f_1134	SNP	A/T	TGTGGCGTTTTGGTAGAGAG	AAGTCCCCCTTAGATCTACC	Maple_1557
AM1f_1165	SNP	A/C	CCGGACCTTGTCAAATTCAC	TCCCCAGAGAGTGTGTTC	Maple_4902
AM1f_1228	SNP	A/G	CTTGAGTTTTTCGAAAGTGGG	CTCACAAGTCACACCCAATC	
AM1f_1283	SNP	A/G	ACCTTCCACAAAATTGACAG	AATCTGCATAGCTTCCACTG	AM1f_453, Maple_4514
AM1f_1351	SNP	A/G	GGTGGATAAGGCTAAGTTGG	TTTCCCCTTCTCTAGGAACC	
AM1f_154	SNP	G/T	GATCTCTGCCGTTCAGATAG	GCGCCCAAATTCTGTTAGAC	
AM1f_2_2725	SNP	C/G	CATGGCCTGATATGGACAAC	TTTCTTCAACCTTTTCCTCC	
AM1f_2_389	SNP	C/T	CATGCAATTGATGAAACCCAG	CCTTCGTCTTTGTTGTTTGG	
AM1f_215	SNP	C/T	GTCAGCAGCTTTATGCTTAG	TCCTGAACAAGTAAGTGTAG	
AM1f_2168	SNP	C/T	TAGCAATCAGAGGCCAATCC	CCGGATAACCAAGCAAGAAG	Maple_4731, Maple_7772
AM1f_2335	SNP	C/T	ACCAAGGCCAGAAAAACAGG	TCGTAACCACATCAAATGGG	
AM1f_287	SNP	C/T	CCCATGCATCTCCTCCAAAG	TGTGGTCTAGTTCCGGAATC	
AM1f_3270	SNP	A/G	TATCCCAGCTGACATAACC	TTGCCCTCAGTTTCTTTTTG	Maple_6560, Maple_4229, Maple_7702
AM1f_441	SNP	A/C	GAGATAGCATCCTTTTTCGG	GTTTCTCTGGGAAATGTGGG	
AM1f_453	SNP	A/G	GCCATGTTGATGAGTTTGTC	ACCACAAATGCACCATCCTC	AM1f_1283
AM1f_481	SNP	C/T	TATCTCCATTGTGGGCAACG	GCACAAACCACAGGTGAAAG	
AM1f_5	SNP	A/G	ATACTTGTCTGTCTTGTTG	ACAGCAACACAAATCAACAG	Maple_5231
AM1f_5144	SNP	C/T	TCACTGTCCCACCATTTAGG	GAAGAAGTTCAACCTGAAGC	
AM1f_524	SNP	C/T	CTCATCATGCATGATCACCC	ACAGGGCTATAATAAGAAG	
AM1f_5331	SNP	A/G	TTACTACTTCGTCAACCAGG	GGGTATCCTGCTATCTGCTG	Maple_8509
AM1f_5928	SNP	C/T	AGACTCTGAGATCCACAAGG	GCAACACTCAAAAGCAAGAC	
AM1f_924	SNP	C/G	CATGATCATGTTGTCACTGG	TGGTTTCATCGCAAGCTGAG	Maple_3748, Maple_5112
AM1f_9252	SNP	A/G	TGTAAGAAGCATACCAAC	CGAGAGAGATTGTCCTAGTG	Maple_1665
AM1f_927	SNP	A/G	GCCAGTGCTTTCCTTTTGAG	TCAAGGTGGCTGCCTTTTAC	Maple_6339
AM1f_984	SNP	A/T	AGGCTCAGAAGAAACCAGAG	CAGGATGATCACGAGTGATG	Maple_2138, Maple_1643
AM2f_176	SNP	G/T	AGACCTCTCGATCTTCAATC	GCTGCTAAAGAAGTAGGTTG	Maple_823, Maple_6339, Maple_3075
AM2f_18	SNP	A/G	CAGAGGCAAGAGACACAAAC	GCATCATGCACCCTGTATTG	
AM2f_2_1164	INDEL	DEL/CA	GAGACGTAGCTAGAGAGTTG	TACTGTCTTCATTTGCAACC	
AM2f_2_123	SNP	C/T	GCTCATGTTTGGAATTATAGG	TCTCATCTTGCATCTCCTCG	Maple_3234
AM2f_2_292	SNP	C/T	GAGATCATCAATTGGAGCAC	AACGAGGAGAGAAGCTCAAG	Maple_1489
AM2f_214	SNP	C/T	ACAGGTTCTCTCAACTTGAC	TCGAATTTTGTATCACACC	

Locus	Marker type	Alleles	Forward primer sequence	Reverse primer sequence	Linked loci
AM2f_218	SNP	C/T	ACTGATCCTGCTTTCGAATC	AACGAGTCCAATCATTCCG	
AM2f_234	SNP	A/G	CGGCTCTGGACCTCAAAACA	TCTTGGATCAGGCGAGAATG	Maple_4514
AM2f_290	SNP	A/T	CGATCTGGAGAACAGTACAC	TCCACTGGAGTAATTGAGGG	Maple_4393
AM2f_346	SNP	C/T	GTGCTGAAGTTTCGGATGTC	CAATACTGACCTGGATCCAC	
AM2f_49	SNP	C/T	GGACATACCAGTTTCAAGTC	AAAAGCACCAAAGCCACCAG	
AM2f_58	SNP	A/G	TCTGAACTCCCAAGTGGAGG	TGTGCTAACCACAGTTGT	Maple_6688
AM2f_617	SNP	A/G	GGAGCAACTATATGTACACC	GCCGTAAGTCTAAAGATGTG	Maple_4998, AM2f_657
AM2f_629	SNP	C/G	ATCCAGTAGAACCTTCTCTC	ATGTTGGCGCAGCTGATTTG	Maple_3258
AM2f_657	SNP	A/G	TGGTTAGCAATAACCCAAGG	CTACTTCATCGCCCGTTTTC	Maple_4998, AM2f_617
AM2f_736	SNP	A/C	GGATTTTTACAGTCAAGTC	TACATGCAACACGCAAAGAG	
AM2f_9	SNP	C/T	AGGCAGATGCAACATGACTC	TGCCACACTAACACAAACCG	
Maple_10588	SNP	A/G	GGCTGAGAGTAACATGAATG	TCATCATCAGCAGTTGCATC	Maple_4381
Maple_1086	SNP	G/T	CCAGGTTCTATTTCCATATGA	TTTAGAAGTCCGCAAGGAGG	Maple_5345
Maple_10862	SNP	C/T	ATGGCCGCCACTAAATGTTC	AGGTTAGCTGCTACTAATGC	Maple_4138
Maple_1191_e	SNP	G/T	GCTGCGAATAAAGTATGGAC	TCTCCTTGTCTCCACTTTCC	
Maple_121	SNP	A/G	TCTTGTGCTGATCAGACTCG	CAACTGGCAATGTTGGACTC	Maple_4803
Maple_12182	SNP	A/C	TGTCAATCCGAACTGCTGTG	AATCGAGCCCAGAATGTGAG	
Maple_13	SNP	G/T	TGAAGACATTGAAGACTCCG	GATGTTGATCTTCAGCCACC	
Maple_13_bis	SNP	A/T	AGGCCTTAGCAGCCACAATA	GGGTTTGTAGCATGGAAATG	
Maple_1308	SNP	G/T	CAAGAAATCAAAGAAGACAG	GATTGACTGCTGAGTGGATG	
Maple_1322	SNP	A/G	CCTTACAAGAAGCAATGAGC	GGAAAGAACTTTTGCATTTTC	
Maple_13253	SNP	A/G	CAAGGGTACTATCCTCATCC	ACGACCTTTGTTTGAGGGAG	Maple_2420, Maple_4258
Maple_1389	SNP	C/T	CAGTGAAGCACTTTGATTGG	CAAACTCTCCATCAAATGTC	
Maple_1481	SNP	A/G	GCTGCCTAAAGCTCAACTTG	AAATCCACGATCTTCTCCGC	
Maple_1489	SNP	G/T	CCCCAGAAGTCAGGTCTTTC	GACTCAAAAATTTTGGATCCG	AM2f_2_292
Maple_1557	SNP	C/G	TGTGTCTCCATGTTTGGTGC	AAAAGAACAGTTCCTTCAG	AM1f_1134
Maple_1569	SNP	A/T	ACCATCTCGTCACCACTTAG	GGTGTAAGGTGTCTCATTGC	
Maple_1607	SNP	A/G	GCATTCACGATAGTTTTCC	GTGATTCCAGATGCAGCAAC	
Maple_1643	SNP	G/T	CAGTAGCAGTAGACATCACG	GAATCCCTTCCCGTAATAGC	Maple_2138, AM1f_984
Maple_1665	SNP	A/G	AGTCTCGTTTTGGAATTGCC	CTAGGGATCTTATCAACTGC	AM1f_9252
Maple_1699	SNP	C/T	ACCACCGAAGCTGCAATCAC	CAGACCTCCTGAGTTTGAAG	Maple_305
Maple_1752	SNP	C/T	GGATATGTTGCCTAACCGAC	CTTCAACTTGCCTTCGCATC	Maple_2103

Locus	Marker type	Alleles	Forward primer sequence	Reverse primer sequence	Linked loci
Maple_1854	SNP	A/G	ACGAGTTCATGCATCTGGTG	TCAAAGACCAATGTGCAGAG	
Maple_1856	SNP	C/T	TTGTTTCATCACTACTCCTC	CTAGTGGCGGAGCAAGTAAT	
Maple_1906	SNP	C/T	ATTGGCAGAGCAATGTACCG	TGTTCAGGTCTACAAGGGTC	
Maple_20	SNP	C/T	AAAAGAGAGGACCGAGCTAC	CCCTGTGTAGAGGTTCTAGC	
Maple_2059	SNP	C/T	TCTCTCCAGAAGACTGTCTC	TAACAGCTTCACGAGCATCC	
Maple_2074	SNP	C/T	TCCTGGGCCAGTTCTATAGC	CAATAGCTGATTGTTCTCGG	
Maple_2076	SNP	C/T	GGGACACAAAGAAAGTCTCG	GGAATGAGCTTTGCCGGAAG	
Maple_2103	SNP	G/T	AATAATTCAAGGGGCCTCCG	CTTGCAAGGAAATCCATGTG	Maple_1752
Maple_2109	SNP	C/T	CTTGACATGAGTGACCTGAC	TCAATTCCATCCTTAGCCTC	Maple_5418
Maple_2138	SNP	A/C	CTCACATATGTGCACAGCTC	TAACCAAAAGTCCTCTGACC	AM1f_984, Maple_1643
Maple_2155	SNP	G/T	AATTGGCGCTAGAAGAAGGG	CTTTATCGCTAACGTTGCTC	
Maple_2394	SNP	A/G	CTCTTTGCCAGTGTATAAGC	CTTTCGCCACCAATGTACAG	
Maple_24	SNP	A/G	TTCCTTTGTTGAGAGCCTCC	AAGCATCATATGTGTATCGC	Maple_4850
Maple_2417	SNP	A/G	AGACGGAACAAAATCCTTG	GTTCTTCTGTCAATGCTTGG	Maple_5345
Maple_2420	SNP	A/G	AAATGGTTCGTCGAACACGC	CGTAGAGCACATTTGACCAG	Maple_4258, Maple_13253
Maple_2760	SNP	A/T	GTCCGCAGTTATGTTCAGTC	GGTCTTACCTCTCAAGCATC	
Maple_2793	SNP	C/T	TGGATCATTTAGAAGACCGC	GTAGAAAACAAATCCACATAG	
Maple_2828	SNP	A/G	AATCCTCTGGCCTCCACTAC	GGAGAGGCGAAATTACATTG	
Maple_305	SNP	A/G	GAGGAATCCAATAGCTGG	AATTTCAACACCCTCGACCG	Maple_1699
Maple_3075	SNP	A/G	CAATTTCCCTGGTTCGTTTG	TAGCCTTGGTTGTCTTGTGC	AM2f_176
Maple_3089	SNP	A/G	ATCGCTGAGCGAGTTCAATG	CATGGGCTTGACAGTGTTTG	Maple_4702
Maple_3090	SNP	A/C	GGTTCTTTCTTCGTCTTCTC	AATGGTGAGTCCAAAGCGTC	
Maple_3120	SNP	A/T	CACAATAGTAGTGAGTTTGGC	GCCCTCTAGATTATTCCTCG	
Maple_3136	SNP	A/C	CAAGTCAGGAATGGCATAAG	GGATGTTATACATGACATGC	
Maple_3234	SNP	C/T	CGACATTGCTAGACACCTTC	ATCTTGCATCTCCTCGTATC	AM2f_2_123
Maple_3252	SNP	C/T	TTCATCATCCCAAGGGTCAC	AACTAGAGTGTGTGAGTGAG	
Maple_3258	SNP	A/G	GTGTAGAGCATATTCCACAAC	CCTAAAGAACCATTAGAGGC	Maple_4174, AM2f_629
Maple_3748	SNP	C/G	CATGATCATGTTGTCACTGG	TGGTTTCATCGCAAGCTGAG	Maple_5112, AM1f_924
Maple_3773	SNP	A/G	AAGGATGAGTAGACTGGAAC	ACTGTGCCTACAACAGCTTG	
Maple_3784	SNP	A/G	GGAATTGGCTGGAGATGAAG	CCATCAATTGTAATCGCACC	
Maple_3814	SNP	C/T	TCAGCAAGAAATCCATCAAC	CAACCGCCATCCACATTACC	
Maple_3882	SNP	C/T	AGTATCGGAGATGAAGCCAC	CTGCAAACACCCTCACAAAC	

Locus	Marker type	Alleles	Forward primer sequence	Reverse primer sequence	Linked loci
Maple_3918	SNP	C/T	TACTCCAGCATCGTCCTCTC	GATTACGAGCAAAGACCGAC	Maple_7588
Maple_3941	SNP	C/T	TGAAGGTCTGGTATAGGCAC	TGCATCTCACCCTTTCTAGC	
Maple_3953	SNP	A/G	GTGGCTGCATCAAGAACATA	ATCATTCAGGACCTCAGCAG	
Maple_3989	SNP	C/T	GAGCGACAGTCTGCGTGTTA	CAAGTGTCCTTGCAGAAATG	
Maple_3999	SNP	A/G	ACCCGCTTTGCATCATCCTC	CCACTCTGGAGAATGATGAG	
Maple_4002	SNP	A/G	TCATCCTCCGGTCATATGTC	CTATGTATGTGTGCTGAAGG	
Maple_4034	SNP	C/T	ACACCGTCCGATTATTTGTG	ATCACAGGCTCACAGCTTGG	
Maple_4044	SNP	C/T	GATCTAACAGGAAGGAACCC	GCTTCAACAGAAGGAAGCAC	Maple_4444
Maple_4049	SNP	A/C	TTGTTTTCTTCTTCTCTTC	GTGAAGAGGATGCCATTGAG	
Maple_4050	SNP	C/G	CCTCCTGGTGTTTTTCCTTC	ACCAACAGTCCAACACTGAG	
Maple_4074	INDEL	DEL/C	GGGCCCAAATAAGAAACAAG	CTTCGTGGGTCATGCTTTTG	
Maple_4091	SNP	C/T	GCAAACCAGACCATTTTCAC	GCTGTCTTGGGACTACATTC	
Maple_4138	SNP	A/G	TAGAATCTCCTCTGCATTCC	AATGCAACCGCCGAATGGCT	Maple_10862
Maple_4144	SNP	C/T	GGATGCCGAAAGGTCAAAAG	TCAAGCAGATGCGAAACAGG	
Maple_4174	SNP	C/T	ACAAATAGAGATGCCGTAGC	CTTTAGATGTCAAGCACACC	Maple_4416, Maple_5421, Maple_3258
Maple_4186	SNP	C/T	ACCCACAAAGTCCAAAAAGC	AGCCTTGAGAGTAATCACCC	Maple_659, Maple_4297
Maple_4218	SNP	A/G	TGAAACACTGAAGGCTACTC	CATTCGACCACGAATTGTTG	Maple_823
Maple_4229	SNP	G/T	TCTACACAGTGTGTCTCCTC	GTAGTAGGCACGACGCATA	Maple_6560, AM1f_3270
Maple_4258	SNP	A/G	AGACAGAGGAGGATGTACAG	CATGTTCTCATCATCATCACC	Maple_4829, Maple_2420, Maple_13253
Maple_4278	SNP	C/T	CAACGGTTTTGGGATGAGAG	AACCTTTCTTCTTTGTTGC	Maple_5092
Maple_4297	SNP	A/G	GCCAAAAACCCATAATGGTC	GGTTGAATCAAACAATGAGG	Maple_4186
Maple_4308	SNP	C/T	CCTTCCAGATTCTTATGGC	CGACTGGAACAAACATCGAC	
Maple_4318	SNP	A/G	TGTTCCGGATGCTCTTTCTC	CAGACATCAGCCGGATATTC	Maple_10588
Maple_4381	SNP	C/T	TTTCTCCGATGAAGCGGAAG	ACCAAATCCGGTTCCCACTC	
Maple_4385	SNP	C/G	AGCTAGCAACGCATGATTTC	GGTTAGAGATGACCTTACCT	
Maple_4393	SNP	C/T	TTTGTGGTTATAGTCAGTGC	CACACTAGCTGCTTGCATTG	AM2f_290
Maple_4416	SNP	A/T	ACTTCACATGTGGATTCTCG	TGATGGACAACTTGCATCCG	Maple_4174
Maple_4438	SNP	A/G	GAGCCAAAAATCGACATCGG	AAGATTTTTGGAGGAGGTGG	
Maple_4444	SNP	A/T	TGACTTGATGGAGCTAGCAG	CAAGGTTCAATCGACACCAG	Maple_4044
Maple_4455	SNP	C/T	CACTTCGCTGGGTTATAGAC	TATGCTGCTCGCAAACATTG	
Maple_4456	SNP	A/G	TGAGGCACTATTGCTATGCT	TAGATATTGCGAATCCAAA	Maple_5345
Maple_4472	SNP	C/T	GTCTGGAGCAACTTATTCCC	GGTAGTCGAACATAACGAAC	

Locus	Marker type	Alleles	Forward primer sequence	Reverse primer sequence	Linked loci
Maple_4484	SNP	A/C	AGCATCAGGTTGATCTTTGG	TCGAGATTTGGGAGTAGTGG	
Maple_4512	SNP	C/T	CAAAGAGTTCAATGGTAGCAC	CCGAAATTTCCCAGTTCCAG	
Maple_4514	SNP	C/T	GATGCATTCCTATGGAGCAG	GGAAAGGTGTTGTTGTTGGG	AM2f_234, AM1f_1283
Maple_4566	SNP	C/T	CTTATGTATGGGATTGAGGG	GTTCCCAATTTCCAAGTCAAC	
Maple_4604	SNP	A/G	ACACAGAACCGGAATCTTCG	CTGGATCACCTCCTTTTCAG	
Maple_4663	SNP	C/T	GGTGCTAATAAGAGCCTCAA	TCTGCAACTGAGTCGTCATC	Maple_5287
Maple_4665	SNP	A/G	GTGAATAAGATGGTTCGCAA	GGTCGTCTGAGATAATTTGC	
Maple_4679	SNP	A/G	AAACAACGAGACCACTCCAC	GTTTGGTCCATCCGTAATGC	
Maple_4693	SNP	C/T	CAGATTATTGAATGAAGGGG	CTGCATTTAGTATCTCCACC	
Maple_4696	SNP	A/G	ATGGCAATTTGACCCGCACC	CCTTTATATGGGTTCACGGG	
Maple_4702	SNP	A/G	TCAGCTTCTGAGGATTCCTG	CCCCAGGATGTGAATGAATG	Maple_3089
Maple_4704	SNP	C/T	CAGACCAACAAGTCCAACTG	ACGGAATTTTGTGAGCCACC	
Maple_4723	SNP	C/G	CTGCGAGGTTGCTTTGATTC	AAATAACGCATCCTCTGCGG	
Maple_4724	SNP	C/T	CACTGGACTCTTAGCATCAC	TTGTATTCGTTTTCCGCCCG	
Maple_4731	SNP	C/T	GGTCGATTTTGGAAGGAAAG	AGGTAGCAGTTCTCTCTAGG	Maple_7772, AM1f_2168
Maple_4803	SNP	C/G	CAGACCACGTTCCATACTAC	GGTTGGTTAAGCGCGGTTTC	Maple_121
Maple_4829	INDEL	DEL/AT	ATGGTGACGATTGAGATGGC	CATCACCAATTGGCTTCAAC	Maple_4258
Maple_4840	SNP	G/T	TGAGGATCAAGAGGACTTCG	TCAAACCTCTCAACCCCAAC	
Maple_4847	SNP	C/T	CCCATATAGAGTTCAAAATCC	TGCACATTCATAGAGGCACG	
Maple_4850	SNP	C/G	CGAAAGAGGTACAACCTCAC	ACTCCTACTGGTGTTGGTTC	Maple_24
Maple_4896	SNP	A/G	TTCCATGGGCATCTATGAGG	TGTTCACTCCTCCTGTCAAC	
Maple_4902	SNP	A/G	TCGGTGTGGAGATACTTGAG	GAAGGATGCCAGTTCTCTTG	AM1f_1165
Maple_4906	SNP	A/G	CGCCAATTCATCCAACAATC	GACAACAAAAGTTCTCAACC	
Maple_4920	SNP	C/T	AAGAGCCACAGCTCTGTTTG	AGCAACAACTTGGAACGGTG	
Maple_4923	SNP	C/T	AAGTGATCGTGCTCCGATTC	CATTCCACCAGTGCAAAACC	
Maple_499	SNP	A/G	CAGCAAATAACTGGAAAAA	TTTGGCTTGCCTTACACCTG	
Maple_4998	SNP	A/G	TGGTTAGCAATAACCCAAGG	CTACTTCATCGCCCGTTTTC	AM2f_657, AM2f_617
Maple_5062	SNP	A/G	CTTCGACTTGGAGTCTCGAT	TACAGTTATCCTCCACCACG	
Maple_5066	SNP	A/T	TAGGATACCGCAACAACCTG	CTGCCTGCATTATAGGCAAG	
Maple_5092	SNP	C/T	TTGAAAGGGTGCACCACCTA	CATTTTTGACCCACCTTGTA	Maple_4278
Maple_5095	SNP	G/T	GAAAGAAAGGGAACCAGGAC	TTCTAATGTCCTATGTGACC	
Maple_5112	SNP	A/G	GCACCATCATTCCGACTATC	TTCGGGTATTGGAGCGATTC	Maple_3748, AM1f_924

Locus	Marker type	Alleles	Forward primer sequence	Reverse primer sequence	Linked loci
Maple_5227	SNP	A/C	CTCCTAAAAGAGGGAGACAG	ATTCACGAAGTGATCGATGG	
Maple_5231	SNP	G/T	TTCACCAAGAGCAGGATGAC	GAGCTGTCATGTTTAGAGTTC	Maple_9291, AM1f_5
Maple_5287	SNP	A/G	TCAACGAGTCAGTCCTTACC	AAACGCCCCGCCACTTCCTA	Maple_4663
Maple_5345	SNP	A/G	GAGGGTCTTCTATTCACCAG	GAATGGATGCCACATTCGAG	Maple_4456, Maple_2417, Maple_1086
Maple_5380	SNP	C/T	GTTGATACTACTATGAACTTG	AAAGTGACCGAAGTGAGGAG	
Maple_5418	SNP	C/T	CTTGCGAGATTAGTTATCCC	TACTTAGAACTTGAGCTGCC	Maple_2109
Maple_5421	SNP	C/G	AGGTCTCTAGAGAACATCAC	ATGTTTAGCCAATGTGGGAC	Maple_4174
Maple_5463	SNP	A/G	GAACATTGTCTGCAAACTAC	CCTATCCTACAAGAGCTTGA	
Maple_5646	SNP	C/T	TTACCTTTACGGCTTCGCAC	TGCATTTGCGAGGTGATTAG	
Maple_57	SNP	A/T	ATAGGATTCAGCTACAAGTG	AAATTAGTGGATGCTGGAAG	
Maple_5761	SNP	A/G	AATCCAGAGCTACATCGACC	CATCTTGAAGTTGGTTGCTG	
Maple_5762	SNP	A/G	CAATCCAGAGCGAAACGGAG	TGTGAACTGACTTAGACCCG	
Maple_5820	SNP	A/G	CGTTTCCTCTTGAAGTACCG	TCATCAACTGTGTAGCGCTC	
Maple_5940	SNP	C/T	GGTTTGTAACAAACCTGGAC	GCTGCTGATGGAAATAAGCC	
Maple_6002	SNP	C/T	ACGGTCTTTTGCAGAGGTTG	CATAGCTAATCCTCCCTCAG	
Maple_6157	SNP	A/C	GCAACCATTATGAAGAAGA	CACAAGCTTGTTCACTGGTC	
Maple_6246	SNP	A/G	AGGCTTGAATCCGAGTTTAC	GCTGACAAAAACTGCTTGAC	
Maple_6317	SNP	A/C	TTTGGCTCTCATATGTAAC	TATAGGCATGGCACAAAAGG	
Maple_6318	SNP	A/T	GAAAGTGTTTGGGATTGGGC	TCCATGGAGATTCCTGAGAC	
Maple_6339	SNP	A/G	ACAGCTACAGAGAATTTGGG	TGTGGGTGAGTTTGTATGGC	Maple_823, AM2f_176, AM1f_927
Maple_65	SNP	A/T	TGCTTGCTATAGCCCTCTTG	ATAGTGACCAATCTCGTTTC	
Maple_6560	SNP	A/G	TATCCCAGCTGACATAACC	TTGCCCTCAGTTTCTTTTTG	Maple_7702, AM1f_3270, Maple_4229
Maple_6578	SNP	A/G	GGAAAAGTGCTATCCAATGC	GCATGATGATAAACCTGTTC	
Maple_659	SNP	A/G	CACCAGTTGCTATATTACAG	GTAATGTCAATTGTTATGGC	Maple_4186
Maple_6626	SNP	A/T	CCATTGAGCGAGTTATCGTG	CAACTAACCGGTCCAATTCC	
Maple_6682	SNP	C/G	TCGTTGTTACGAAAAGAGTC	ACAACCACTAAACCAAGTCC	
Maple_6688	SNP	A/G	TAGGTCTGAACTCCCAAGTG	TGTGCTAACCACAGTTGT	AM2f_58
Maple_679	SNP	C/T	GAAGGTTCTCATACCCTAC	AAAACAGAATAGGATTGCCC	
Maple_75	SNP	A/G	CCAGTCCCAGAACAGATTTG	GCACTCCATCGGTAACATTG	
Maple_7509	SNP	A/C	AAGGCTACCGGATTGGTTG	AGATAGCTGCCCCAATGATG	
Maple_7588	SNP	C/T	GTGCAGAGGCCAATAGATTC	GAATCTTCGTGCAGGCTTCC	Maple_3918
Maple_7702	SNP	A/C	TACAACTCTGCTTGTTCCTC	GCTCCAGGAACCATAAGAAG	AM1f_3270, Maple_6560

Locus	Marker type	Alleles	Forward primer sequence	Reverse primer sequence	Linked loci
Maple_7772	SNP	C/T	TCTTGTCTTGGAGGGCATAG	AGAACTTTGACTTGGCTACC	Maple_823, AM1f_2168, Maple_4731
Maple_7856	SNP	C/T	GTTCAGTAACTTCTGCAACC	TCGACTTTGGCATTGGAGAC	
Maple_820	INDEL	DEL/TGA	GTAACTTCCTTGCCCTCTTG	GAAATGTACTTCTGCAGAAGC	
Maple_823	SNP	A/G	GTAATATGGGTCATGACCCG	ACCCGTTGATAGTCCAAACC	AM2f_176, Maple_4218, Maple_6339, Maple_7772
Maple_8509	INDEL	DEL/TTCAGG	CTCTCATTCTGGACCATGAC	TCAGCCTAGGGCTTTGTTTG	AM1f_5331
Maple_8688	SNP	C/T	GAACTTTGTTATTTATAGGC	AATCCTCCCACGATGAAACC	
Maple_886	SNP	A/G	ACCAAATACCCTTTCAGACC	ACCTCAATTCACAGTCCAGC	
Maple_889	SNP	A/T	GCTTGATCTGATGGGCAATG	CAGGGCTAGTTTGATCAAGAG	
Maple_9048	SNP	A/C	CACTGAGCCTTTCCAATCTG	TGCCATCTACCCGAAGAAAC	
Maple_9291	SNP	C/T	AGATCGCGGTCGTAAGTAAC	TCGCGAAAACGCAAGAAGAG	Maple_5231
Maple_937	SNP	A/G	GAGAAAATCCACAACAAGAC	TCGATAACTGGAATGACTGG	
Maple_974	SNP	A/T	AATGCCCTTGATAGGCTTTG	ACCCCATCAACCATGACAAC	
Maple_9828	SNP	G/T	CTGTGATAATATCGGCCAAC	TGTGGAGTCCCCTAGCGCA	

Intentional blank page

Statement of Authorship

Title of Paper	*Development of SNP markers for ayous (Triplochiton scleroxylon K. Schum) an economically important tree species from tropical West and Central Africa*
Publication Status	**Published**
Publication Details	Conservation Genetic Resources (2016) 8: 129-139, DOI: 10.1007/s12686-016-0529-8

Principal Author

Name of Principal Author (Candidate)	Duncan Iain Jardine		
Contribution to the Paper	conducted in-house labwork & data analysis; conducted literature review; wrote manuscript as principal author		
Overall percentage (%)	70%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	12.08.2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

vii. the candidate's stated contribution to the publication is accurate (as detailed above);

viii. permission is granted for the candidate in include the publication in the thesis; and

ix. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	*Celine Blanc-Jolivet*		
Contribution to the Paper	organised sample collections; organised external labwork; edited manuscript 20%		
Signature		Date	16.05

Name of Co-Author	*Rainbo RM Dixon*		
Contribution to the Paper	assisted with sample preparation; edited manuscript		
Signature		Date	30.07.2019

Name of Co-Author	*Eleanor E Dormontt*		
Contribution to the Paper	assisted with sample preparation; edited manuscript		
Signature		Date	05.08.2019

Name of Co-Author	*Bianca Dunker*		
Contribution to the Paper	assisted with sample preparation; edited manuscript		
Signature		Date	12.12.18

Name of Co-Author	*Josephine Gerlach*		
Contribution to the Paper	assisted with sample preparation; edited manuscript		
Signature		Date	06.08.2019

Name of Co-Author	*Birgit Kersten*		
Contribution to the Paper	assisted with data analysis; edited manuscript		
Signature		Date	16.05

Name of Co-Author	*Kor-Jent Van Dijk*		
Contribution to the Paper	assisted with sample preparation; edited manuscript		
Signature		Date	16.07.2019

Name of Co-Author	*Bernd Degen*		
Contribution to the Paper	obtained funding for the research; edited manuscript		
Signature		Date	16.05

Name of Co-Author	*Andrew J Lowe*		
Contribution to the Paper	as principal supervisor; obtained funding for the research; supervised experiments; edited manuscript		
Signature		Date	10.07.2019

Conservation Genet Resour (2016) 8:129–139
DOI 10.1007/s12686-016-0529-8

TECHNICAL NOTE

Development of SNP markers for Ayous (*Triplochiton scleroxylon* K. Schum) an economically important tree species from tropical West and Central Africa

D. I. Jardine[1] · C. Blanc-Jolivet[2] · R. R. M. Dixon[1] · E. E. Dormontt[1] · B. Dunker[1] · J. Gerlach[1] · B. Kersten[2] · K.-J. van Dijk[1] · B. Degen[2] · A. J. Lowe[1]

Received: 17 March 2016 / Accepted: 22 March 2016 / Published online: 30 March 2016

Abstract 182 SNP markers were developed for Ayous (*Triplochiton scleroxylon* K. Schum) by incorporating information from two next generation sequencing approaches (RADseq Floragenex and AFLPseq IonTorrent PGM) into a single genotyping panel for MassARRAY® iPLEX™. This set of markers was successfully used to genotype 753 individuals from 43 populations in five Tropical West and Central African Countries. These loci have an expected heterozygosity range of 0.007–0.501 and F_{ST} from 0 to 0.306.

Keywords Single nucleotide polymorphism · MassARRAY · Obeche · Timber tracking · Cameroon · Congo · Ghana

Introduction

Ayous (*Triplochiton scleroxylon* K. Schum), also known as Obeche, is an important timber species from the tropical West and central African forest, with approximately 38,000 m^3 traded annually (ITTO 2015). The timber from Ayous is most commonly used in sauna panelling, house construction and table tennis rackets. It is found north of the equator in monsoonal equatorial forests, with a discontinuous distribution from Sierra Leone eastwards through to Democratic Republic of the Congo and southwards to Gabon (Hall and Bada 1979; Igboanugo and Iversen 2004). Ayous is a large deciduous tree growing up to 50 m tall with a branchless trunk (bole) of up to 30 m. It is a pioneer species of primary forest, but also commonly found in secondary forests. Unlike other associated forest species, it has large distinctly lobed leaves (5–7 lobes, up to 20 cm across) (Hall and Bada 1979; Bosu and Krampah 2005; Orwa et al. 2009). Ayous is assumed to be self-sterile and outcrossing (Orwa et al. 2009).

Very little molecular marker investigation has been undertaken for natural populations of Ayous (Hardy et al. 2013). A previous study of the species by Akinnagbe (2008) used Amplified Fragment Length Polymorphism's (AFLPs) only focused on Nigerian populations. Considering the importance of this species to the African economy, a detailed inventory of the genetic variation across its geographical range will be a powerful tool for forest monitoring and conservation and thus is considered a high priority. This paper presents a list of Single Nucleotide Polymorphic (SNP) makers suitable for such a purpose in Ayous. The SNPs were applied for the development of reference data to trace the geographic origin of Ayous timber, and thus could be used as a tool to enforce regulations to combat illegal logging (Degen and Henry 2015; Dormontt et al. 2015).

Marker development

SNP markers were developed by a reduced representation approach using either a restriction associated DNA sequencing (RADseq) protocol using two samples (see

Electronic supplementary material The online version of this article (doi:10.1007/s12686-016-0529-8) contains supplementary material, which is available to authorized users.

✉ D. I. Jardine

A. J. Lowe
andrew.lowe@adelaide.edu.au

[1] School of Biological Science, University of Adelaide, Adelaide, SA 5005, Australia

[2] Thünen Institute of Forest Genetics, Sieker Landstraße 2, 22927 Großhansdorf, Germany

Table S1) by Floragenex (Portland, Oregon, USA), hereafter referred to as the Thunen Institute for Forest Genetics (TIFG) approach (all P_loci), or by the protocol of Jardine et al. (2015), using the two samples from the TIFG approach, along with an additional 46 samples (Table S1), hereafter referred to as the University of Adelaide (UA) approach (all A_ loci). DNA from all 48 samples was extracted by TIFG from cambium plugs or dried leaves using the DNA extraction protocol described in Dumolin et al. (1995). The TIFG approach was based on RADseq that combines genome reduction with a high coverage in the genomic regions analysed (Baird et al. 2008; Slavov et al. 2014). Libraries were prepared using the restriction enzyme *Sbf*I, and sequenced on the Illumina HiSeq 2000 platform to create paired end reads of 2 × 100 bp (Floragenex). SNPs were identified in the sequenced individuals using variant call format (VCF) 4.1 (Floragenex). For the UA approach, libraries were developed using the protocol of Jardine et al. (2015). Sequencing was done on the Ion Torrent PGM platform (Life Technologies) using the Ion Torrent PGM Sequencing 400 Kit. Sequencing read analysis was done using both the CLC-Genomic Workbench (Qiagen, Venlo, The Netherlands) and Geneious R6 (Biomatters, Auckland, New Zealand) programs.

Initial marker screening

Although undertaken separately, both marker development approaches used the following procedure for the initial marker screening. Ninety samples, consisting of the original 48 from the marker development stage, along with an additional 48 (Table S1), were used to screen an initial selection of markers. DNA of the extra samples was extracted separately for each approach; with TIFG using the Dumolin et al. (1995) protocol, and UA samples extracted at the Australian Genome Research Facility (AGRF, Adelaide, Australia) using the Nucleospin Plant II Kit (Nachery-Nagel, Düren, Germany). Suitable loci were identified in Assay Design Suite (ADS) (Agena Bioscience) and genotyped on the MassARRAY iPLEX platform (Agena Bioscience), using the iPLEX GOLD chemistry (Agena Bioscience). Genotyping was undertaken by either; INRA Genome Transcriptome Facility (GTF, Bordeaux, France) (for the TIFG approach) or by AGRF (for the UA approach).

Second marker screening

A second panel of markers was then compiled by combining successfully amplified markers from both development approaches into one panel using the ADS to design the primers and multiplex groups. With this second panel, the genotyping of 911 individuals (Table S1) was undertaken, which included replicates of the 90 individuals used in the initial screening. DNA of the 911 samples was extracted at AGRF, from cambium tissue. Using Genodive (Meirmans and Van Tienderen 2004) and Genepop (Rousset 2008), tests for heterozygosity, global deviation from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium were undertaken. Monomorphic (or effectively monomorphic loci) were removed. Individuals and loci that had a <95 % genotyping coverage were also discarded from the final dataset.

Results

A combined total of 1667 variable SNP loci were identified as suitable for testing (see Table 1 for the breakdown per approach), of which 250 were incorporated into multiplex panels and tested for their screening suitability. A total of 238 loci were used to genotype 753 samples. Of these 238 loci, 56 were removed due to failure to amplify (42 loci), low representation (present in <95 % of individuals; 19 loci) and monomorphism (eight loci). A panel of 182 polymorphic loci is hereby presented (Table 2) along with their SNP allele calls, with forward and reverse MassARRAY sequencing primers. Expected heterozygosity (H_E) ranged from 0.007 to 0.501 and F_{ST} from 0 to 0.306. 65 loci were not in HWE and 57 loci were found to be significantly linked ($P < 0.005$) (Table 3). Genotyping runs between platforms showed >99 % concordance in allele calls.

This paper has identified 182 SNP makers that are variable for Ayous (*Triplochiton scleroxylon*) from across its geographical range in Tropical Africa and will allow for further genetic analysis to be undertaken.

Table 1 Comparison and number of loci found from either of the marker discovery approaches

Stage	Approach		Combined total
	TIFG	UA	
Marker development	1538	129	1667
Initial MassARRAY screening			
Used	133	117	250
Suitable	121	96	217
Second MassARRAY screening			
Used[a]	142 (9)	96	238
Suitable	108	74	182

[a] Extra loci from the TIFG approach were included in the second screening, () denotes numbers

Table 2 Details of each SNP allele, with forward and reverse primer sequences

Locus	Alleles	Forward primer sequence	Reverse primer sequence
A_1018	A/T	CTTAGGAGAAAGGTAATGTG	AGTGAACCATGACATGGCTG
A_10244	C/T	GACTAATAATGCTCCAACC	GAACGTGAAATAGTCCCTGC
A_1056	C/T	CTAGGCTTGAAACGACCATC	GGAAGGAATGATACCTCCTG
A_1079	G/T	TATAGCATTTAGGGACCCAC	TCCCAAAGCTACTTATCCTC
A_1099	A/T	GCTGGCATTTGACGTTGTAG	TTTCTTCTTGGAACTGCTTG
A_1109	C/T	CGTGGTCCCAGTATATATGC	GAACCATCGATAATCACATAC
A_1130_1	C/T	AATCAGTGTAAGTAGCTGCC	CCCTAGTTGAAAAAAGCAACC
A_11487	G/T	TTCGTCTTCTCTCTTTCACC	GATCCACTCCATATTGAGGC
A_1167	A/G	TCCTTCTGAAGAAATTTGGG	AGCCAAGATTGAAATGGAAG
A_1194	C/T	TTAGTAAGGGGCTAAGTGGG	TATTGCAAATCAAGTAGCCG
A_1315	C/T	TGATGTGCCTTTTGGAGGTG	GAACAACCAAGGCCAAAGAC
A_1521	A/G	CCATGTAGCAGCTGCATGAA	AATGTTGATTGTGGTGTTGC
A_15414	A/G	ACACCTTTTGGAGCGCTATG	ATCTCCTAGTTAGACACCTC
A_1625	G/T	CAGAGACTTGGACTTCAACC	ATTAGAGGAGTGGGTACAGC
A_168	C/T	TGGTCTTTGCACCTTTTGAG	CAAGACTTCAAGCCATTGAG
A_1684	A/C	ATGCTTTCCTCCCACATCAC	GTTAGGACTCAATGCAATGG
A_172	A/T	TTGGAATTGTGCTTGCATGG	CCAATGCCTTGATGATTGTG
A_1805	G/T	GTATGCCAAGTTTACATCCC	TTGAGGATGGAAAGAAGTGC
A_1900	C/G	CAACTTCAGCAGGATGGATG	GCCGTGTGAGAAAGATGTTG
A_1938	A/T	TTGATCCATAGAGACTAGAC	GCACCATTGCTTGAAACAAC
A_199	G/T	CCTTGTGTCAACTAAAACCC	ACCCTAAAGTGTAGCATCTC
A_208	A/C	CAAAGAACCAAACGTTACGG	AAAGCATGATGTCATGTCTC
A_23178	C/T	ATCAGTACACTTTACCACGC	GGAATGCAAGCACTAGCAAG
A_2440	G/T	AACGACAAGGAGCAAGAGAC	CAAGCAAGAATGGGATCTGG
A_2442	A/G	TTTATGGCCATCCTTCATGC	GTAAACCATAGAACACCACC
A_253	A/G	GAAGATAAAGAATCAGGGTTG	TCATTGTCTGAACTTACACC
A_25893	C/T	ATATGAACATGAGTGGTGGC	TTGTTGAGGTGCCATGCTAC
A_26	A/G	CGGAAAAAGATGATGGAGGG	TGGGAGAGAGTAGTACTAGG
A_2724_2	A/C	GCACTTTGATTTCGGGTGAT	GGTTTGCAAGGACAAAGCTC
A_2753	C/T	TGAAGCCTTAGCCATTTCTC	GTGAGTCTAAAATAAGCGTC
A_2841	G/T	AATTTGCTGGCTGCCATCAC	GAATTCATATCAGACGTTC
A_2942_1	C/T	GAACCAACAAGCCAGCAAAG	GCTACAAAGAACACTCTAAG
A_2942_2	A/G	TCATCGACATAAGACCAGAC	GGCTTGTTGGTTCACAGTTG
A_31129	G/T	ACATCCCATCATTGAAGCCC	GGATGCAAAAGGCATGAAGC
A_3189	A/T	GTATCACCAGAATGACTAGC	CATAAGCTTTGTGCAGCCTC
A_3628_1	G/T	AGACAAATTTTCCACAAAG	TGATGGGTCTATACTATGGC
A_368	C/T	CCAAATGCACAAACTCTGGC	GAAGTTCTTTGCCAAGGCTC
A_387	G/T	CCATTTGCCAGCTTGATCAC	TCCCTAGTCTTGCTACTAAG
A_389	A/C	GTCTTCTGGTTTCACATCCC	GGGCTGTAGAAGCAGAAAAC
A_4_1	C/T	AACTCGGCCGGTACTTCAC	TGACCAACCTCGTAAGTCTG
A_4_2	A/T	ACAAGAGAGTTGGTTGAGGC	GACAGGAAAACTTCCCGTTG
A_4037	C/T	CAGAACAAACTCCATGACAC	AACTAGGCAACATGAAAGGG
A_407	A/G	GGTAATCTTGACCATAGGAG	GCTCGACTTATTGCTAAGGG
A_411	A/G	TGCTAGGACCATACTTGGTG	TTCAAATGTGAAGCAATGGG
A_412_1	A/C	GTATTTGTATTCACATTCTCC	CCTAGCTTATGATGCCAAAG
A_412_2	C/T	TTCTTCTTGTAATCATCAC	ACAAGGGTAAATGTCCTAGC
A_4249	A/C	GGACAGCAAGCATAAAATGG	CTTGACAGCAAAATTGCCAG
A_435	A/C	CTATGTTACATTACATGGC	ATCCAATAGGTCTCAGACAC
A_444	A/T	TCCTCGAATTCAGGAGGAAC	ATGAGGTTGATGGAGGAACG
A_532	C/T	CGATGTTTCTGTTCACACAC	CAGAGTCAGTTTGGTTCAGC

Table 2 continued

Locus	Alleles	Forward primer sequence	Reverse primer sequence
A_5394	A/T	CTCGATAGGACATGAGATGG	ACACGGCACGAATTCAAAGC
A_55	A/C	TGACAATCGAGCATTGCAAG	CTTATGCACAAACGTCTCC
A_55761	A/G	GAAAGAGCTTGTAAGTCAAGT	CTGCTTACATTTTGATGCTG
A_6020	C/T	TTAGGTATCTACACCGCCAG	TATCCACGATGGTGTATGGG
A_626	A/C	TGAAGGTGCCTCTGTTTATG	CCCACTCAAACTCTCTTCAC
A_642_1	A/C	TGATAAAGCTTTGGATCCTC	GGAACGAATTGCTCATCCAG
A_642_2	A/T	TACATTGTTTGGCGATGAGG	GGAGTTGGCTTCCCTTCATC
A_6645_1	G/T	CATTACCTCACCATCCTCAC	GTGCTTCATTGCAAAGGG
A_6645_2	A/G	GTACTTGGTAGGCATCATTTC	AGCAACACGATTATATCCG
A_665	A/G	TTAGACACCTCCACTTGACC	TGTACTCTTCCAAAATGCAC
A_698	C/T	ACAGTAGGTGTAGATGCAAC	GGTGTTACAAATTACATGGG
A_71422	C/T	ATGAGATTGACTAAGCTGCC	GGTTTATTCCACTTGATGGC
A_73919	A/G	AAGCAGGGATCGCAACATTC	TCAGATTTGTGACCAGCGAG
A_77365	A/T	TGTAAAAGGCATTGCCTTAG	GAAGCCAAATCATGAGATGC
A_827	A/T	CTTCAACTTTGTATTGCCTC	CCGTTCTCAAGACCTTCTAC
A_8898	A/G	TGCGATACTCCTAGGGAATC	TCATGAGTCTCGCCTACAAG
A_913	A/C	CCTACTCAACCTATCATCGC	GAAGTATAAGCGTGTTGAGTC
A_929	G/T	CTTGGCAAGCGATCTATGAG	GTCCATCTGCGGTTCCAATC
A_935	C/G	AGCCACTATCTCACCTTTAG	TTGATGGATCTGCTTACGGG
A_9516_1	C/G	TCCTTATCCTCCTTCTGATG	TGGACAGTGGAAAGAAATCG
A_9516_2	C/T	CACTGCAGCTGTCATCTATG	ATCAGAAGGAGGATAAGGAG
A_961	C/G	TTATAGCCTAACGAGGTCGG	TAACATCTTGCCACGTCGTC
A_CS_110	A/G	AAGCAGGGATCGCAACATTC	TCAGATTTGTGACCAGCGAG
A_CS_165_1	C/T	CTTGCCTTCATTTCTCCTGC	GGAAAATATGGGTTTGAAGC
P0065	A/G	GGATGACTTGTTTGATGTGC	AAATCTGGTCCCTCAGCAAG
P0112	A/T	GAGTAACAGAGTGTTGCTCA	TACAATTTGGGAGAATGGAG
P0133	C/T	GGTGGAAAGCAAACAAGGAG	TGCCACCTATAGCAATGCAG
P0182	A/G	ACTGGGTTGACTCCAGATAG	GACAATATCAAGTAGTAGGG
P0245	A/C	TCACCCTGCTCAAGTCATTC	TATGTGTGCGTCCTTTTCGG
P0265	C/T	TCCTATCAGCATTTCCACTC	TGCAGATAAGGTGGCAAAGG
P0380	C/T	TGAAGCAAGCACAAGACAGG	TGCTCCTGCATTTGTTCCTG
P0616	C/T	TACGAAGAATAAATAAGAAG	CCTCTGGATACTTAGCTTCG
P0761	A/T	GTCTTTCCAAGCATTTCTCC	GACAGGAGTCACCATAATTC
P0785	A/G	GATGCGGATATCTGCTCTTG	AAAACCAAGACTGCACACCC
P0809	A/G	GCCACCTTCTTTGCTATCAG	AAACCTTCTTCGAAGCCCTG
P0812	C/T	AGGGAGTAGAGACTAAGAAC	GTGCACACATTTGATTTGCG
P0855	G/T	GGCGAGAATAGAATTAAATG	TCATTACAAGAGCTGGGAGG
P0884	G/T	TTGAAGGAGGCCATTCCTAC	CATATACATCGCGTCTCCTG
P0896	C/G	GGGAGGTTCATGTTGTTCAC	ATTATGATGAGGGTTCATCG
P0917	A/G	GCAAGATGAGGACGATGAAG	TCTTCGTCGTCTTGGATTCG
P0981	C/T	TAGTGTCTTAAGAGGATCAC	GCTTTGGGTTTGAACTATCC
P1064	A/G	CCAGTTTGCACAACACCTTC	CCAACACATACCTTTCATCG
P1094	C/T	AAATGTCTCGAGCTTCAGGC	CAATGCAAGTTCCCATAACC
P1103	C/T	GGGTACTTTACAAAATGAC	TTCTGGCGAATTCTAAGCAC
P1165	A/T	TTCTTTCTGTTGTCTGGGTG	AGAATGACCGCATTCCCTTC
P1265	C/G	ATTCCCGGCAACGGAATTAC	CTTTCCATGTAGCTGGCTTC
P1477	G/T	CCAACCACAGCTTCTATTTC	TTTGCTGACCTTGATCCCAC
P1481	C/G	CGGGCTGATCTATTTCGAAC	ATGATGCAGAACCTATAGAG
P1547	C/T	TCTCCTTCTTTGATGGTGAG	GAGACTGCATCAGTTATGGC

Table 2 continued

Locus	Alleles	Forward primer sequence	Reverse primer sequence
P1559	A/G	AGCTTGAGCATTTGCTAGGG	TTAGATTGCTGATCACTCGG
P1813	A/T	TTAGTGCATATTTGCTCGTC	TTGAGAGAGAAAGAGAGAG
P1835	C/G	CTACTGATAGAAGCCATAGG	TGAGAATGGTTGTGTTACTG
P1860	C/T	CATCATAGACAGCTTGACCC	AACATAGGATTCGGCCCAAC
P1894	C/T	ATATTAGAGGGCTAGGCCAC	TTTGGTTGGAACCTAAGGGC
P1918	C/T	TGTTTCAGATTATTGTGCCC	TCAGCAACAGCAGGAAGTAG
P1919	C/G	AGAGGCCTTCGATAAGATGG	CGTAAGGTTAGTGCTTACAG
P1960	C/G	GGTGCAAATTTCTACCCTCC	AGCCTATATATTGCAGCCAC
P2146	A/C	TTGGACTTTCCTTAGGCTTC	GCCGAAAAGTTGAACTGGAC
P2274	C/T	GGTCATGCTAAGTGTAACTG	CAGGCCATGTCAGTGTAATC
P2290	A/G	CTTCAGAGGATGAGAAACAG	CTACCTTGATGAAGCAAGTC
P2328	C/G	ACATGCATGAAATCATAGGG	GGTGGATGAAGCTTTTCACG
P2367	A/G	GACCAGTGAATCTAATTGGC	CTGTTCTTCTTCAGAGGGAG
P2464	A/G	CACGGTAATCATGGGATTGG	CCCTCTCCTGACATAAATAG
P2496	A/G	AATGCCGATGGAATGGGAAG	ACTTGAGCAGTTAAAACTCC
P2618	A/T	TTCGGTGCAAGGAAATAGGG	ACTTTGCCAACTCCAAGCAG
P2644	A/G	ACATTGCCACTAAACCACCC	GAATAGGACAGCTACACTGG
P2679	C/T	AACGAAAGGGCAAAAGCTCC	TTGGTCCTCCTTTCATCTGC
P2722	A/T	CCTTTTCAGTCTTTCTTACC	TTTCCTTTTTGCAGGTCCGC
P2733	A/T	CCCTATAATTTTCCATTGCC	AAGCACCTAGTTAGCTTTCC
P2749	C/T	ATGAGGAGTGGGTGAGAAAC	GATCTGTCATAATTCCGAGG
P2775	C/T	GAGAATTCCTCCTCTTGACC	GGGTTCAAACCCACACTTAC
P2967	A/T	AGCATGATACTTGTCTACCG	CATCCATGAAGTCAAGTGCG
P3076	C/T	TCAGGACTGGTTGATGAAGG	CGCCCTAGTAGGTCAACCAT
P3093	C/T	CCTGCAAAACGGTGATTTTC	CATCGGTTATTGATGCCTGC
P3137	A/T	TTACCTGCACGATATGTCCC	GTCAGTTTCTCCTGAGTTGG
P3285	A/G	CTTATGTCATTTCCATCGGG	AGGGTCTGATCATGAATGGG
P3303	C/T	CTCTATCTCGTACCTAGAAC	ATCAGCGCCAATGCCAAAAC
P3414	C/G	CCGAGACAGGATAGTCAAAG	TCAAAACCACGGGCAAGAAG
P3480	C/T	TCCAAACAAGGAGGATGCTG	TACAGCTAGTTGGACTCCAG
P3657	A/G	AACAACCATTGCCATGACGG	AGCAGTCATGATGCAAACCC
P3699	A/G	TCAACCTCTTCTTCCTCTTC	AAGGGTTTGGTATCATCATC
P3722	C/T	GGCACGTGCTTTTTGCCTAT	GTGTCCTTTTCCCTCTACTC
P3752	C/T	CCCCTCAAGCTTGATATTCC	ACGGACGAGGGAAGCATTAA
P3788	C/T	TCCTCTGTAAGGTATGTCCC	CTGAAAACGCAGGCGAATAC
P3997	C/G	GTAGTAGGCTTATTAACGAC	TCTTATCCTTCTCAGAGCCC
P4042	G/T	TTGATGAAGTGGTCAGCTTG	CATGCTTACGCAAGTAAACC
P4293	A/C	TGTTGCTCAGCAATTGCAGG	GAAACCTCGAAGGCAATCTC
P4294	C/T	ACTGCACATTCTAGTGGAGG	GCATATGAGATCCACTTTCC
P4492	A/G	TCTCAGCTGCAACAAGCAAC	CCTTTCTCGGCTATTCTGTG
P4617	A/G	AGACCAACCAAAGGAAGCTG	GTGGGTGAATGAGTGGAATG
P4629	A/G	CTCCAATTCTCGAAGTCAAC	AGCATTTGCTGTCCAACTTG
P4638	C/G	ACTATTTGCATGCTGCAGGG	TGAGTTACAACCTCCTCCTC
P4706	A/C	ATGAAGTCCCTGGCCATTTC	GGCATATGGTTTTTGTGCTT
P4772	A/G	CTCTTTGAGCTAATCACAGG	ACCTTTGCACTTAACCGGTC
P4837	A/G	CTGCACACTGGATGCATTAG	CAGTTTGGATCGGCTTCAAG
P4872	C/T	CCACATGCTGGGTTTGATTC	TGATTCCCCATTTTAGCTTG
P4926	C/T	GGTTTTATGCCTAGCTACAC	TGCTCTTTAGTGCTTTTGCC
P4928	A/C	TTGTTCCGTGCTTCGTGATG	CCAAAAGCCCATTTCACTAC

Table 2 continued

Locus	Alleles	Forward primer sequence	Reverse primer sequence
P5178	A/G	CCGTCTATTCATATTCGCTC	AAAATCATGGCACGGATAAG
P5240	C/T	TGGATCGAGCCTGCAATTGG	AGGTGGTTCAGCTTGAAATC
P5264	A/T	TCAGTGTGATTGGTTAGAGG	CATCTTCAGTCAGAGGAAGC
P5404	G/T	CTGAGCTCAAGTTAGAAGCC	TGTGGGAATTCCATGGATAC
P5439	G/T	CTACAGCTAGAATATGACCG	ATTAGTCCGTATGTGACGCC
P5462	C/G	GTGCATCTTTGTGTTGACTC	ATCTCTTCCTTGCTGGAGTC
P5532	A/C	CTGTGTTTCTTTCCCCTTTC	TGCTAGAAAACTACGAAGCC
P5562	C/T	CGGTCAAGGTCATGATCAAG	GAGCCAATTTGGAAAATCCG
P5587	A/G	CTAAATCTCATGAGGACGGG	TTTTGGAGGACTGTAATCTG
P5700	C/T	TGATGCTTCTCACTCTGTTC	CCAGCATCGACATTTGACAG
P5715	A/G	AACAAAGCAAATGACACCTC	ATCGAGTCTGAAAATCTGTC
P5737	C/G	CTTGACGCAACAAAGCGCAC	ATGATCAAATCAATCCCGGC
P5777	G/T	TAAATCGATCGGGTCCTGTC	CTCAAAACACAGCCTGGTTC
P5888	A/G	ACTTGACCTCTTCTGGCTAC	GGGAGTATGGTCTATGTAAG
P5909	G/T	TCCTCCAGAAGGGAAAATGC	CAACCATGAAGCCTACACTG
P5944	A/G	GCAGTAGCAGACCAAGAAAC	GGTTCGGTCATCTTGAAAGG
P5972	A/G	CAAACCAAAAGCTTTCAGCAG	CAGACTATTCTGATGCATGCT
P6137	A/G	ACTACCTAGAGAGATCAAGC	TGAAGTTCATGTCCAACCTG
P6163	A/G	CTCTGTTCCAGGATGAGTTG	GCCCCACAACATATATCAGG
P6225	A/G	CAAGAGACTTACCATAAGCC	GACCCTTTGCTATGAAATGG
P6238	A/C	TTATTTGGGTGCGGGATCGG	TTGTCATCAGCAGCCTCTTC
P6277	A/G	GCAACTCTATACTGATAGAGG	ACAACGCCAAATACACATGC
P6290	C/G	GACCCTAGCTAACATCGAAG	GAGTCAGAACCAAAGGAAGG
P6328	C/G	GTACTTCCCAACCTCTACAC	TGGTCATAGACCATCAGTGC
P6392	A/G	GCATGCATATTTTTCAAAC	CCTAAACAGAGAGGGAAGAC
P6399	A/G	TTGCTTCACTAAACTCCCCC	TAGTCATTGTCAGTTGGGTG
P6402	A/T	AATATCCCGGTTTGAACTCG	CAACTACTCCATTGACAGGG
P6483	G/T	ATGGATTCAACTGAGATGTG	AAGACTGACTGGACAGTGGC
P6527	C/T	TGAGCCACAGTTGATACCAC	AGAAAGCGGCTATCCTAGAC
P6533	C/T	CCTGTCCCAACCATTTTGTC	CCAGTCTCTTTCTCCTTTCC
P6618	A/G	CGGAACAACAATATTGTCTC	ACATCGATCTTGGGAGGTTC
P6704	G/T	GGAGCAGGAGGAATTTAAGG	GGTTCATGTAACAAGGTGGC
P6741	C/G	TGGAGTCCTTGAAACTAGAG	ACATTTCCCATCCAGGAAGC
P6787	A/G	TGCAGGTAATCTGGACTTTC	TTTGCAACAGTCATGGCTGG

Table 3 Population genetic parameters of markers developed in this study

Locus	H_O	H_E	F_{ST}	HWE	Linked loci
A_1018	0.093033	0.091	0.018	Yes	
A_10244	0.517756	0.472	0.017	No*	
A_1056	0.052314	0.052	0.031	Yes	
A_1079	0.027012	0.058	0.017	No****	
A_1099	0.401666	0.383	0.044	No*	A_626
A_1109	0.144287	0.165	0.011	No**	
A_1130_1	0.040127	0.138	0.063	No****	
A_11487	0.113827	0.11	0.132	Yes	A_2841
A_1167	0.063241	0.063	0.007	Yes	A_1684
A_1194	0.03706	0.039	0.011	No*	
A_1315	0.019547	0.033	0.028	No***	
A_1521	0.172077	0.179	0.031	Yes	
A_15414	0.097745	0.344	0.064	No****	
A_1625	0.156959	0.157	0.024	Yes	
A_168	0.048916	0.053	0.035	Yes	
A_1684	0.103366	0.101	0.032	Yes	A_1167
A_172	0.012662	0.015	0.018	No*	
A_1805	0.0283	0.05	0.016	No****	
A_1900	0.203873	0.182	0.068	No*	A_4249
A_1938	0.174097	0.328	0.092	No****	
A_199	0.160011	0.178	0.138	No***	A_6645_1, P1477
A_208	0.055877	0.06	0.027	Yes	
A_23178	0.429399	0.438	0.07	Yes	
A_2440	0.071372	0.074	0.015	Yes	
A_2442	0.488105	0.409	0.051	No***	
A_253	0.355546	0.378	0.073	Yes	
A_25893	0.170731	0.215	0.051	No****	
A_26	0.173093	0.154	0.057	Yes	P3093, P6402
A_2724_2	0.501239	0.466	0.031	Yes	
A_2753	0.330884	0.302	0.165	Yes	A_77365, P5737
A_2841	0.136742	0.275	0.159	No****	A_11487
A_2942_1	0.537573	0.501	0.014	Yes	A_2942_2, P0380
A_2942_2	0.228849	0.336	0.046	No****	A_2942_1
A_31129	0.027852	0.04	0.007	No****	
A_3189	0.283377	0.248	0.059	No*	
A_3628_1	0.028755	0.029	0.01	Yes	
A_368	0.085982	0.085	0.008	Yes	
A_387	0.087368	0.123	0.039	No****	
A_389	0.064079	0.063	0.001	No****	
A_4_1	0.410845	0.426	0.021	Yes	A_4_2, A_532, P0785, P5404
A_4_2	0.149808	0.144	0.01	Yes	A_4_1, P0785
A_4037	0.450415	0.474	0.018	Yes	P3657
A_407	0.489102	0.46	0.082	Yes	
A_411	0.24864	0.24	0.08	Yes	
A_412_1	0.2681	0.232	0.138	No*	A_412_2
A_412_2	0.247073	0.221	0.021	No**	A_412_1, P3137
A_4249	0.2951	0.371	0.096	No****	A_1900
A_435	0.153395	0.152	0.03	Yes	
A_444	0.442758	0.458	0.036	Yes	A_532, A_77365, P0182, P0761, P0917, P4837

Table 3 continued

Locus	H_O	H_E	F_{ST}	HWE	Linked loci
A_532	0.038873	0.031	0.185	No****	A_4_1, A_444, P1103, P3699, P5404
A_5394	0.230543	0.242	0.062	Yes	
A_55	0.152034	0.165	0.022	Yes	
A_55761	0.198766	0.225	0.06	No*	
A_6020	0.144222	0.171	0.018	No**	
A_626	0.155507	0.14	0.141	Yes	A_1099
A_642_1	0.328028	0.326	0.306	Yes	
A_642_2	0.010982	0.011	−0.001	No****	
A_6645_1	0.26864	0.259	0.024	Yes	A_199, A_6645_2, P5264
A_6645_2	0.482689	0.479	0.027	Yes	A_6645_1
A_665	0.238999	0.218	0.032	Yes	P3303
A_698	0.472466	0.431	0.003	Yes	P6483
A_71422	0.020076	0.022	0.008	Yes	
A_73919	0.210979	0.215	0.045	Yes	A_CS_110
A_77365	0.460907	0.458	0.052	Yes	A_2753, A_444, P4837, P5737
A_827	0.397321	0.404	0.031	Yes	
A_8898	0.477334	0.469	0.034	Yes	
A_913	0.083644	0.118	0.028	No****	P6527
A_929	0.107053	0.113	0.036	Yes	
A_935	0.361777	0.364	0.03	Yes	
A_9516_1	0.397059	0.414	0.084	Yes	
A_9516_2	0.176745	0.181	0.031	Yes	
A_961	0.041062	0.041	−0.005	Yes	
A_CS_110	0.209318	0.214	0.046	Yes	A_73919
A_CS_165_1	0.146411	0.151	0.028	Yes	
P0065	0.060671	0.065	0.01	No*	
P0112	0.34828	0.334	0.058	Yes	
P0133	0.345986	0.338	0.091	Yes	
P0182	0.51621	0.496	0.001	Yes	A_444, P2749
P0245	0.0211	0.021	0.008	Yes	
P0265	0.351575	0.372	0.047	Yes	
P0380	0.272244	0.259	0.128	Yes	A_2942_1
P0616	0.249607	0.253	0.048	Yes	
P0761	0.500303	0.48	0.03	Yes	A_444, P4617
P0785	0.43884	0.474	0.022	Yes	A_4_1, A_4_2
P0809	0.166041	0.182	0.025	No*	
P0812	0.378764	0.414	0.101	No*	
P0855	0.019989	0.019	0.059	Yes	
P0884	0.045946	0.048	0.02	No*	
P0896	0.396963	0.393	0.009	Yes	P1094
P0917	0.266164	0.244	0.016	Yes	A_444, P4837
P0981	0.317668	0.35	0.039	No*	
P1064	0.014831	0.016	0.028	Yes	
P1094	0.332687	0.336	0.012	Yes	P0896
P1103	0.492627	0.485	0.023	Yes	A_532
P1165	0.211522	0.215	0.058	Yes	
P1265	0.384077	0.384	0.039	Yes	
P1477	0.206337	0.183	0.083	Yes	A_199
P1481	0.009252	0.009	0.06	No****	

Table 3 continued

Locus	H_O	H_E	F_{ST}	HWE	Linked loci
P1547	0.072262	0.07	0.014	Yes	
P1559	0.346004	0.351	0.202	Yes	
P1813	0.026355	0.025	0.036	No****	
P1835	0.016081	0.015	0.031	No****	
P1860	0.301467	0.3	0.268	Yes	
P1894	0.338365	0.361	0.034	Yes	
P1918	0.08072	0.083	0.049	Yes	
P1919	0.333867	0.455	0.003	No****	
P1960	0.475925	0.446	0.047	Yes	
P2146	0.178643	0.261	0.072	No****	
P2274	0.214375	0.252	0.02	No*	
P2290	0.361529	0.34	0.111	Yes	
P2328	0.091182	0.091	0.039	Yes	
P2367	0.447908	0.425	0.045	Yes	
P2464	0.345378	0.357	0.064	Yes	
P2496	0.099038	0.104	0.045	Yes	
P2618	0.265158	0.41	0.132	No****	
P2644	0.51302	0.49	0.005	Yes	P5737
P2679	0.276638	0.322	0.113	No**	
P2722	0.09604	0.103	0.077	Yes	
P2733	0.279262	0.332	0.037	No***	
P2749	0.441156	0.442	0.105	Yes	P0182
P2775	0.371264	0.381	0.034	Yes	
P2967	0.192119	0.197	0.042	Yes	
P3076	0.446435	0.431	0.027	Yes	P3093
P3093	0.447167	0.442	0.053	Yes	A_26, P3076, P6402
P3137	0.281166	0.264	0.065	Yes	A_412_2
P3285	0.089765	0.091	0.013	Yes	
P3303	0.339193	0.374	0.038	No*	A_665
P3414	0.301193	0.322	0.104	Yes	
P3480	0.074833	0.089	0.052	No**	
P3657	0.448743	0.451	0.048	Yes	A_4037
P3699	0.444768	0.44	0.105	Yes	A_532, P6533
P3722	0.092981	0.111	0.027	No*	
P3752	0.263187	0.311	0.106	No****	
P3788	0.036137	0.035	0.021	No****	
P3997	0.395838	0.393	0.006	Yes	P6483
P4042	0.235358	0.367	0.194	No****	
P4293	0.071199	0.08	0.037	No**	P6328
P4294	0.064661	0.069	0.002	No*	
P4492	0.151633	0.152	0.04	Yes	
P4617	0.234324	0.242	0.028	Yes	P0761
P4629	0.093062	0.093	0.016	Yes	
P4638	0.475152	0.479	0.035	Yes	
P4706	0.519359	0.48	0.027	Yes	
P4772	0.114793	0.137	0.008	No**	
P4837	0.446817	0.45	0.007	Yes	A_444, A_77365, P0917
P4872	0.204229	0.231	0.039	No***	
P4926	0.08712	0.098	0.049	No*	

Table 3 continued

Locus	H_O	H_E	F_{ST}	HWE	Linked loci
P4928	0.014569	0.014	0.02	No****	
P5178	0.004546	0.007	0.007	No*	
P5240	0.294353	0.278	0.06	Yes	
P5264	0.399463	0.391	0.074	Yes	A_6645_1
P5404	0.293849	0.295	0.15	Yes	A_4_1, A_532
P5439	0.018491	0.018	0.033	Yes	
P5462	0.460152	0.478	0.014	Yes	
P5532	0.339573	0.359	0.04	Yes	
P5562	0.373921	0.399	0.09	Yes	
P5587	0.226824	0.236	0.027	Yes	
P5700	0.103751	0.141	0.006	No****	
P5715	0.022219	0.021	0.055	No****	
P5737	0.40955	0.486	0.022	No****	A_2753, A_77365, P2644
P5777	0.409202	0.411	0.117	Yes	
P5888	0.09932	0.097	0.014	Yes	
P5909	0.006722	0.007	0.029	No****	
P5944	0.17922	0.185	0.044	Yes	
P5972	0.169764	0.172	0.032	Yes	
P6137	0.01898	0.018	0.031	No****	
P6163	0.129667	0.124	0.018	Yes	
P6225	0.416036	0.401	0.141	Yes	
P6238	0.337492	0.325	0.156	Yes	P4293
P6277	0.067269	0.07	0.073	Yes	
P6290	0.438493	0.462	0.018	Yes	
P6328	0.007559	0.008	0.016	No****	
P6392	0.33798	0.337	0.251	Yes	
P6399	0.049765	0.071	0.097	No***	
P6402	0.332317	0.459	0.048	No****	A_26, P3093
P6483	0.297177	0.283	0.097	Yes	A_698, P3997
P6527	0.367753	0.342	0.022	Yes	A_913
P6533	0.45612	0.439	0.053	Yes	P3699
P6618	0.477722	0.452	0.072	No**	
P6704	0.183277	0.201	0.041	Yes	
P6741	0.331152	0.318	0.024	Yes	
P6787	0.481235	0.461	0.008	Yes	

H_O observed heterozygosity, H_E expected heterozygosity, F_{ST} infinite allele model, *HWE* Hardy–Weinberg equilibrium

Asterisks refer to significance of deviation from HWE, * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$

Acknowledgments The work was financially supported by the International Tropical Timber Organization (Project PD 620/11 Rev.1) and the Australian Research Council (Project LP120100648). The TIFG approach for loci screening was supported by funding from the Conseil Regional d'Aquitaine (Nos. 20030304002FA and 20040305003FA), European Union, FEDER (No. 2003227) and from Investissements d'avenir, Convention attributive d'aide (No. ANR-10-EQPX-16-01).

References

Akinnagbe A (2008) Genetic variation in *Mansonia altissima* A. Chev. and *Triplochiton scleroxylon* K. Schum under different regimes of human impact in Akure Forest Reserve, Nigeria. Dissertation, University of Göttingen

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376

Bosu PP, Krampah E (2005) *Triplochiton scleroxylon* K. Schum In: Louppe, D Oteng-Amoako AA, Brink M (eds) PROTA. (www.prota4u.org/search.asp). Accessed 12 Dec 2015

Degen B, Henry B (2015) Verifying timber in Africa. In: Carillo R (ed) ITTO tropical forest update 24(1):8–10

Dormontt EE, Boner M, Braun B, Breulmann G, Degen B, Espinoza E, Gardner S, Guillery P, Hermanson JC, Koch G (2015) Forensic timber identification: it's time to integrate disciplines to combat illegal logging. Biol Conserv 191:790–798

Dumolin S, Demesure B, Petit R (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. Theor Appl Genet 91:1253–1256

Hall JB, Bada S (1979) The distribution and ecology of Obeche (*Triplochiton scleroxylon*). J Ecol 67:543–564

Hardy OJ, Born C, Budde K, Daïnou K, Dauby G, Duminil J, Ewédjé E-EB, Gomez C, Heuertz M, Koffi GK (2013) Comparative phylogeography of African rain forest trees: a review of genetic signatures of vegetation history in the Guineo-Congolian region. CR Geosci 345:284–296

Igboanugo ABI, Iversen P (2004) Conservation and use of *Triplochiton scleroxylon* in moist West Africa. In: Iversen P, Sigaud P, France-Lanord M (eds) Forest Genetic Resources 31:50–52

ITTO (2015) Biennial review and assessment of the world timber situation 2013–2014. Appendix 3:125–187

Jardine D, Dormontt E, van Dijk K-J, Dixon R, Dunker B, Lowe A (2015) A set of 204 SNP and INDEL markers for Bigleaf maple (*Acer macrophyllum* Pursch). Conservation Genetics Resources 7:797–801

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes 4:792–794

Orwa C, Mutua A, Kindt R, Jamnadass R, Simons A (2009) Agroforestree database: a tree reference and selection guide. Version 4.0. (www.worldagroforestry.org/af/treedb/). Accessed 12 Dec 2015

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour 8:103–106

Slavov G, Nipper R, Robson P, Farrar K, Allison G, Bosch M, Clifton-Brown J, Donnison I, Jensen E (2014) Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass *Miscanthus sinensis*. New Phytol 201:1227–1239

Supplementary Table 1: Number of *Triplochiton scleroxylon* individuals used in each of the three stages; Marker discovery (from both the TIFG (Floragenex RADseq) and UA (AFLPseq) sequencing approaches), as well as the Initial marker screening and the combined second screenings on the MassARRAY® iPLEX™ platform (Agena Bioscience™).

Country	Nearest location, region	# Individuals[%]			Latitude	Longitude
		MD	IS	SS		
Democratic Republic of Congo	Kole, Orientale			6	1.9744	25.3559
	Yangambi, Orientale	5	5	23	0.7581	24.495
	Opala, Orientale			2	-0.3429	23.8077
	Yahila, Orientale			23	1.8496	23.6129
	Simba, Orientale			11	0.58	22.9747
	Yekana, Orientale			1	0.8627	22.8317
	Boyasegbago, Equator			22	3.0914	20.5514
	Botikpo, Equator			23	3.2054	20.5345
	Boyagonda, Equator			17	3.1128	20.1704
	Gemena, Equator	5	5	39	3.2367	19.8107
	Yembongo, Equator			24	3.1788	19.0064
Republic of the Congo	Ouesso, Sangha			8	1.5833	16.5588
	Ouesso, Sangha		5	20	1.9235	16.4336
	Ouesso, Sangha		5	21	1.5574	16.2544
Cameroon	Yanga, East			30	2.2157	15.5014
	Menziong, East		6	23	3.3796	15.1404
	Adjélu, East			23	2.5683	13.9354
	Djampiel, East			38	3.9125	13.9184
	Letta, East	6	6	22	4.9117	13.6277
	Minta, Central	6	6	21	4.6232	12.8555
	Mbama, East			27	3.8901	12.7639
	Djoum, South	6	6	21	2.7168	12.6631
	Yaounde, Central			20	3.6754	11.4165
	Bafia, Central	6	6	20	4.8567	11.3316
	Esson, South			23	2.6159	11.1966
	Tonga, West		6	21	4.9485	10.7376
	Nyabessan, South	1*	6	19	2.3283	10.4901
Ghana	Oda, Eastern			23	5.9593	-1.0748
	Borobi Forest Reserve, Ashanti	3^	3	8	6.9551	-1.3662
	Nkarabia, Ashanti	5	5	23	6.0397	-1.5621
	Agosa, Brong Ahafo		6	23	7.5503	-2.0153
	Susanho, Brong Ahafo		6	16	7.2447	-2.2016
	Akrodie, Bring/Ahafo			21	6.6972	-2.6156
	Enchi, Western	5	5	23	5.8105	-2.7389
Ivory Coast	Aukope, Agnebi			25	6.4093	-3.9037
	Agboville, Agnebi			25	5.8758	-4.283
	Rubino, Agnebi			26	6.0239	-4.3418
	Garéko, Sud-bandama			25	6.0916	-5.6813
	Gauge, Sud-bandama			25	5.6789	-5.7433
	Gauge, Sud-bandama			25	5.6595	-5.7684
	Issia, Sud-bandama			24	6.4965	-6.5748
	Guiglo, Sud-bandama			24	6.5168	-7.4761
	Logoualé, Sud-bandama			25	7.1423	-7.5308

NB: [%] initials represent the three different stages that samples were used in; MD=Marker Development, IS=Initial Screening, SS= Second Screening. * symbols represent the two samples used in the TIFG RADseq approach, ^only one sample from the Borobi Forest Reserve population was used.

Chapter 4: Population Genetics of *Triplochiton scleroxylon*

The role of refugia and geneflow in defining genetic clusters of a tropical African rainforest tree

DI Jardine[1], EE Dormontt[1], AJ Lowe[1*]

[1]School of Biological Sciences & Environment Institute, University of Adelaide, Adelaide, SA 5005, Australia

*author for communication (andrew.lowe@adelaide.edu.au)

Intention to submit to: Heredity or Molecular Ecology

Statement of Authorship

Title of Paper	The role of refugia and geneflow in defining genetic clusters of a tropical African rainforest tree
Publication Status	Unpublished and unsubmitted work written in manuscript style
Publication Details	Intending to submit to Heredity or Molecular Ecology

Principal Author

Name of Principal Author (Candidate)	Duncan Iain Jardine		
Contribution to the Paper	Designed the study; conducted lab work & data analysis; interpreted data; conducted literature review; wrote manuscript as principal author		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	06.08.2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

x. the candidate's stated contribution to the publication is accurate (as detailed above);

xi. permission is granted for the candidate in include the publication in the thesis; and

xii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Eleanor E Dormontt		
Contribution to the Paper	Assisted with study design; supervised experiments; advised on focus & structure of manuscript; edited manuscript 10%		
Signature		Date	05.08.2019

Name of Co-Author	Andrew J Lowe		
Contribution to the Paper	As principal supervisor; obtained funding for the research; advised on focus & structure of manuscript; edited manuscript 10%		
Signature		Date	10.07.2019

Title: *The role of refugia and geneflow in defining genetic clusters of a tropical African rainforest tree*

Abstract

The forests of the Guineo-Congolian phytochoria spread from Sierra Leone to the Democratic Republic of the Congo, split by a 200 km biogeographical barrier in Benin and Togo, the Dahomey Gap. Despite being Earth's second largest rainforest region, the biogeography of this region remains poorly studied. From available pollen core, phytogeographic and genetic studies, it can be inferred, that during the Last Glacial Maximum (LGM; ~20,000 years before present), the forests of this region contracted to isolated refugia separated by non-forested vegetation. At the end of the LGM, the forest vegetation rapidly expanded to cover a continuous area much larger than exists today, before contracting back to the area similar to today's preclearance state. Along with the life history traits (e.g. mating system and dispersal mechanisms) of individual tree species, these changes to the forest extent have shaped the contemporary genetic structuring of African tropical rainforest species.

In order to progress knowledge about the biogeography of this critically understudied region, we examined the population genetic diversity of *Triplochiton scleroxylon. T. scleroxylon* is an economically important, widely distributed wind dispersed tree. For this study, 753 individuals from 41 populations in five countries (Ivory Coast, Ghana, Cameroon, Republic of Congo and the Democratic Republic of the Congo) from across the distribution of the species were analysed. We found that current populations of *T. scleroxylon* group into three geographically distinct genetic clusters. Our results indicate that contemporary populations of *T. scleroxylon* have had limited geneflow across the Dahomey Gap, with separate genetic clusters forming either side of the gap. This finding is consistent with results from other studies in other species from this region. We also identified a strong genetic boundary between the North-Western and Central regions of the Democratic Republic of the Congo, despite their being no obvious biogeographical barrier. Samples from the North West group with

those from Cameroon, formed the most genetically diverse cluster within the species range. While samples from central populations form their own divergent group with the lowest genetic diversity. Interestingly this clustering boundary is a more significant geneflow barrier than the Dahomey Gap and has not been identified previously in the literature (for plant species). This clustering pattern likely reflects historical refugial population dynamics although the active maintenance of as yet unidentified barriers to geneflow across the divide cannot be ruled out.

Keywords: *Refugia, Geneflow, paleoclimate, population genetic structure, genetic diversity, Guineo-Congolian forests, SNP genotyping.*

Introduction

The Guineo-Congolian rainforests are the second largest expanse of tropical forest in the world (Hardy et al. 2013). The rainforests encompass an almost continuous area from the border between Sierra Leone and Liberia eastwards to the foothills of the Albertine rift on the border between Uganda and the Democratic Republic of the Congo. The forest cover is divided into two blocks by a significant natural break, the Dahomey Gap, a 200 km stretch of savannah in Benin and Togo (see Figure 1). In recent geological times, the African tropical rainforests have undergone significant changes. As recently as the Humid Holocene period (9000-5500 years before present (YBP)), forests covered a much larger area, extending ~500 km further north into the current savannah zone, and including across the Dahomey Gap (Anhuf 2000; Malhi et al. 2013). Prior to the Humid Holocene, the drying effects of the last glacial episode (110,000-12,000 YBP), which peaked during the Last Glacial Maximum (LGM) (26,000-19,000 YBP) had constricted the African rainforests into isolated refugia pockets (Duminil et al. 2013; Hardy et al. 2013). Several theories regarding the number and distribution of refugial locations exist (Anhuf et al. 2006; Maley 1996), however they are broadly congruent (Hardy et al. 2013). These major climatic oscillations have been significant drivers for structuring genetic

variation within species across the Guineo-Congolian rainforest block (Duminil et al. 2013; Iloh et al. 2017). Unfortunately, for the most part, genetic studies have been focused on the Lower Guinea region (centred on Cameroon) only (see Hardy (2013) for summary). The biogeographic influence and genetic structure more widely across the region remains poorly understood, particularly further west to Sierra Leone and east into the Congo Basin (Hardy et al. 2013). The lack of understanding on the genetic structuring of key ecosystem tree species across this vast forested region remains a significant knowledge gap.

Historical climatic conditions are not the only determiners of a species' genetic structure; life history traits also play an important role (Broadhurst et al. 2017; Lowe et al. 2018). In particular, characters that are associated with reproduction (including pollen and seed dispersal, successional stage and range size) directly influence the genetic structure of a species through contemporary geneflow (the transfer of genetic material between individuals and populations of a species) (Dick et al. 2008; Lowe et al. 2018; Petit and Hampe 2006). Tropical tree species especially, are known to exhibit high levels of geneflow. High geneflow is typically one of the mechanisms employed to ensure a species' resilience to environmental changes (Degen and Sebbenn 2016; Dick et al. 2008; Sexton et al. 2015). However, high geneflow results in a uniformity of the genetic signature of the individuals across a species' distribution, reducing the genetic structure within a species (Austerlitz et al. 2000; Degen and Sebbenn 2016; Dick et al. 2008; Hamrick et al. 1992; Newton et al. 1999; Petit and Hampe 2006).

For species where genetic structuring patterns are unknown, they can be postulated by using species with similar life history traits as a proxy to theorise the genetic structuring of the unknown species (Broadhurst et al. 2017; Lowe et al. 2018). Yet, while studies on comparable species can be used to infer the expected genetic arrangement of another species, they may not be accurate. It is always preferable to undertake the analyses directly on a target species if given the opportunity (Degen and Sebbenn 2016; Malhi et al. 2013).

This study considers *Triplochiton scleroxylon* (K. schum) (family Sterculiaceae/Malvaceae), one of the most economically important trade species from tropical Africa. *Triplochiton scleroxylon* can be found throughout the Guineo-Congolian rainforests north of the equator (see fig. 1) (Hall and Bada 1979; Igboanugo and Iversen 2004). It is a deciduous, canopy emergent, primary/pioneer forest species, and, similarly to many other African rainforest tree species, is predominantly outcrossing, self-sterile, and insect pollinated, with wind dispersed seeds (Leakey et al. 1981; Oni 1990).

Previous studies of *T. scleroxylon* have chiefly been in relation to its use as a forestry species (e.g. Bowen et al. 1977; Igboanugo 1989; Ladipo et al. 1991; Leakey 1992; Leakey 2004; Leakey et al. 1982; Leakey and Coutts 1989; Leakey et al. 1981; Leakey and Longman 1986; Leakey et al. 1990; Leakey and Storeton-West 1992; Mayaka 1994; Nketiah et al. 1998; Oni 1990). The health benefits of *T. scleroxylon* (Aranda et al. 2013; Kespohl et al. 2005; Venturini et al. 2004), particularly its anti-diabetic properties have also been of focus (Prohp and Onoagbe 2009; Prohp and Onoagbe 2012; Prohp et al. 2012). However, there have been limited genetic studies into *T. scleroxylon* to date, and none have provided a detailed understanding of its population genetic structure. Prior to this study, only brief genetic diversity analyses had been undertaken and reported. Either as a subset of a study on a close relative species (*Mansonia altisimma*) and focused specifically on a small number of populations in Nigeria (Akinnagbe 2008; Akinnagbe et al. 2010), or as part of a marker development phase for this project (Ch. 3_{II}; Jardine et al. 2016).

Despite the level of genetic isolation and distinction of genetic clusters within *T. scleroxylon* remaining uncertain, assumptions can still be made. Given the geneflow promoting life history traits of *T. scleroxylon* (i.e. outcrossing species with wind dispersed seeds), it is expected that there will be limited population genetic structuring. Nevertheless, considering its impact as a biogeographic barrier in other studied species (Hardy et al. 2013; Malhi et al. 2013; White 1979; White 1983), it is expected that discrete genetic clustering of individuals either side of the Dahomey Gap will occur. Additionally, it is postulated that

across the species range, clustering patterns should align with refugial locations in the Congo Basin and along the Gulf of Guinea coast (Hardy et al. 2013).

Using samples from across the range of *T. scleroxylon*, we aimed to conduct a genetic structuring analysis to identify the key biogeographic patterns and analyse the distribution of genetic diversity within and between identified genetic clusters. A detailed genetic study of *T. scleroxylon* will be an important addition into the understanding of genetic structuring patterns of trees from the African tropical rainforests.

Materials and methods

For this study a total of 911 individuals were sampled from 43 populations in five countries (Ivory Coast (CIV), Ghana (GHA), Cameroon (CMR), Republic of the Congo (COG), Democratic Republic of the Congo (COD) (see map (fig. 1) for sampling locations; and Table 1 for population co-ordinates/locations/ number of individuals). Each sample was genotyped using 182 Single Nucleotide Polymorphism (SNP) markers (see Table 3 from Jardine (2016), (i.e. *Ch. 3$_{II}$)*, for summary of loci and genetic diversity values). The SNP markers were developed using either a restriction associated DNA sequencing (RADseq) protocol by Floragenex (Portland, Oregon, USA), or the double digest RADseq (ddRAD) protocol of Jardine (2015).

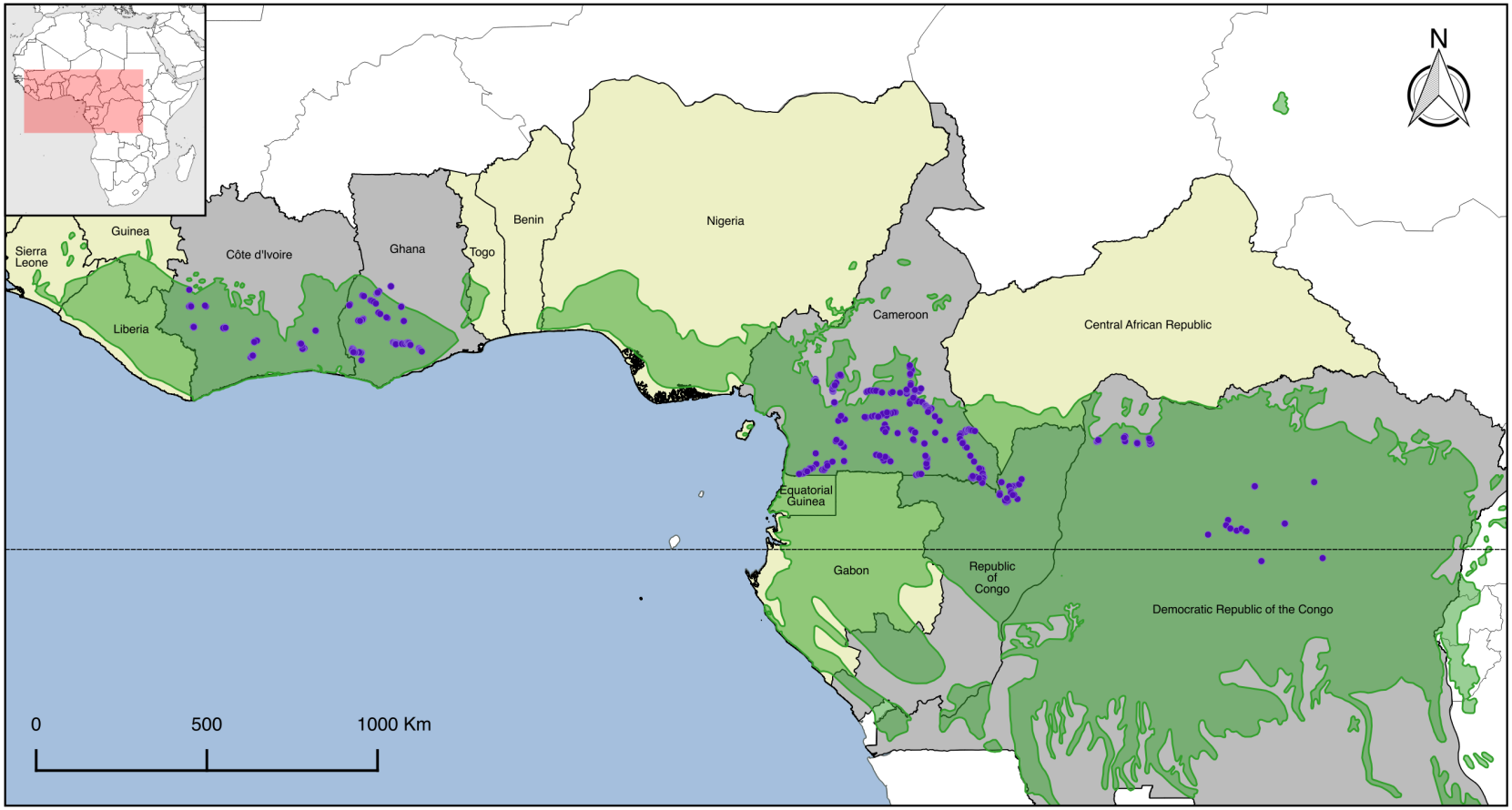


Figure 1: Sampling locations and approximate distribution of *Triplochiton scleroxylon*

NB: Green =African forest vegetation cover (approximate distribution of T. scleroxylon) (above the equator (dotted line) only), Purple dots are locations of samples used in this analysis; Coloured countries where T. scleroxylon can be found, Grey represents countries where sampling has occurred, whist yellow where it has not.

This study utilised a modified version of the genotyping dataset from Jardine et al. (2016). Firstly, SNP loci that were linked (Linkage Disequilibrium (LD)) or out of Hardy Weinberg Equilibrium (HWE)) were removed. A ≥95% genotyping success rate was then re applied to the reduced dataset to remove samples with insufficient coverage (as per Jardine (2016; 2015)). Finally, there was also a minor regrouping, as two populations (DRC_24 (*n=1*) & DRC_26 (*n=2*)) had insufficient numbers of individuals for population level analyses. Subsequently, these samples were incorporated into the nearest population (DRC_25) for this analysis (see Table 1). After the amendments, the final dataset used in this study, consisted of 753 samples, 41 populations and 105 loci.

Firstly, genetic diversity and Analysis of Molecular Variance (AMOVA) were conducted on the dataset. The analyses were performed using GENODIVE (version 2.0b27) (Meirmans and Van Tienderen 2004) (Genetic diversity calculated with the *Genetic Diversity* function), (AMOVA calculated with the *amova* function and the following parameters: standard (not nested) calculation, Infinite Allele Model, 9999 permutations).

Following the genetic assessments, a structuring analysis to determine the most appropriate number of genetic clusters/populations (*k*) in *T. scleroxylon*, was undertaken. This was performed using two clustering programs, DAPC (Jombart et al. 2010) and STRUCTURE (Pritchard et al. 2000). DAPC is a Discriminatory Analysis of Principle Components (DAPC) and is found in the R package *adegenet* (Jombart 2008). The STRUCTURE analysis was completed in two phases, an *initial screen* then a more *focused test*. Due to the size of the dataset, a comprehensive assessment of all genetic cluster options (*k*=1-41) would be too time consuming and most likely uninformative, so an *initial screen* of the dataset was performed to identify a range of suitable *k* values for more detailed testing (the *focused test*). The *initial screen* performed three iterations of each *k* (1 to 41). Each iteration had a burn-in length of 10,000 Markov Chain Monte Carlo replicates (MCMC) followed by a run length of 20,000 MCMC. The *focused test* consisted of ten iterations for each suitable *k*, with each iteration having a burn-in length of 300,000 MCMC, followed by a run length of 700,000 MCMC. For both

tests, default parameters were used, and a random seed was applied to each iteration. To select the *k* values for the *focused test*, the Delta K scores ($^{\Delta}$K) from an Evanno method test (Evanno et al. 2005) (performed with STRUCTURE HARVESTER (Earl and vonHoldt 2012)) on the *initial screen* STRUCTURE results was used. A range of *k* values, including those with the highest $^{\Delta}$K, were then incorporated in the *focused test*.

$^{\Delta}$K scores were also employed to identify the most likely *k* from the *focused test* STRUCTURE results. Additionally, to identify the most applicable *k* for explaining the genetic structuring in *T. scleroxylon*, the results were passed through the CLUMPAK web-based program (using both the *Main Pipeline* and *Best K* functions) (Kopelman et al. 2015).

Once the appropriate *k* had been identified, genetic analyses were then conducted for this new arrangement. In order to undertake these tests, the population formatted dataset was amended to reflect the most applicable group within a *k* that each population assigned to (inferred from the CLUMPAK population Q-matrices (the *ClumppPopFile* output file). Additionally, for comparison, these tests were also performed on a country orientated grouping (based on the relevant country where the sample originates). To amend the original dataset, the GENODIVE *population groupings* option was used. Genetic diversity analyses were performed using the *compare groups* function (GENODIVE), with standard parameters and 9999 permutations. Nested AMOVAs (nesting populations into a group) were then run using the same parameters as previously stated.

To allow for more detailed genetic assessments of the new groupings, the dataset was transformed (using the *transformation* function (GENODIVE)) to dissolve the population level grouping and assign individuals to the relevant country or genetic cluster only. For both new datasets, a standard AMOVA (as previously described) and pairwise F_{ST} (*Pairwise Differentiation* function (GENODIVE)) (using amova method parameters and 9999 permutations) were performed.

Finally, Isolation by Distance (IBD) (Mantel) tests were carried out individually for every genetic cluster or country, by transforming them into independent datasets for testing. A global test was also done, by treating all samples as a single "population". The IBD (mantel) tests were performed using the *mantel.randtest* function (Thioulouse et al. 1997) in the R package *ade4* (Dray and Dufour 2007) with the following parameters: Sample location coordinates were not transformed, and imported as decimal degrees; 99999 replicates.

Results

From the original Jardine et al (2016) dataset (911 individuals, 182 loci), 77 loci were removed due to Linkage Disequilibrium (LD) or problems with Hardy Weinberg Equilibrium (HWE). In addition, 158 individuals did not meet the ≥95% coverage criteria and were also removed. This left a modified dataset of 105 loci and 753 individuals. The subsequent genetic diversity analysis found observed heterozygosity (H_O) was 0.258 (SD±0.016), total heterozygosity (H_T) was 0.272 (SD±0.017), inbreeding coefficient (G_{IS}) was -0.007 (SD±0.004) and fixation index (G_{ST}) was 0.058 (SD±0.007). Population specific H_O ranged from 0.224 (DRC_25) to 0.298 (GH_05), Heterozygocity within a population (H_S) ranged from 0.167 (DRC_25) to 0.298 (GH_05) and G_{IS} ranged from -0.469 (DRC_23) to 0.094 (GH_04). For genetic diversity analysis results of all populations and overall see table 2 and table S1 for all loci results.

The results of the population grouped AMOVA (table S2A), found that most of the variation (~95%) occurred within individuals (F_{IT}) (F-value = 0.052 (SD±0.008)), whilst the least amount of variation (-0.5%) was between individuals (F_{IS}) (F-value = -0.005 (SD±0.004)).

The DAPC clustering analysis found *k*=5 was the most suitable grouping for the dataset (BIC=2001.264) (see fig. 2A). The Evanno test on the *initial screen* STRUCTURE results found that only *k* values 2-4 ($^{\Delta}K$=154.20, 84.40 and 19.79 respectively) were applicable for further testing. Based on the *initial screen* and DAPC results, the *focused test* was run using *k* values 1-7 (2-6 were actually

tested but *k*=1 and *k*=7 needed for the test to work). The results of the *focused test* STRUCTURE analysis were the same as the initial screen, with *k* values 2-4 (ΔK=1311.37, 397.00 and 240.30 respectively) as the most likely. Both the CLUMPAK *main pipeline* and *Best K* tests found *k*=4 as the most probable. For *k*=5, the Evanno test found only minor support (ΔK=16.82), and the CLUMPAK *main pipeline* test could not aggregate all 10 iterations. A comparison of *k*=5 STRUCTURE and DAPC results can be seen in fig. 2. The population assignment proportions and averaged STRUCTURE plots of *k* values 2-4 can be seen in fig. 3.

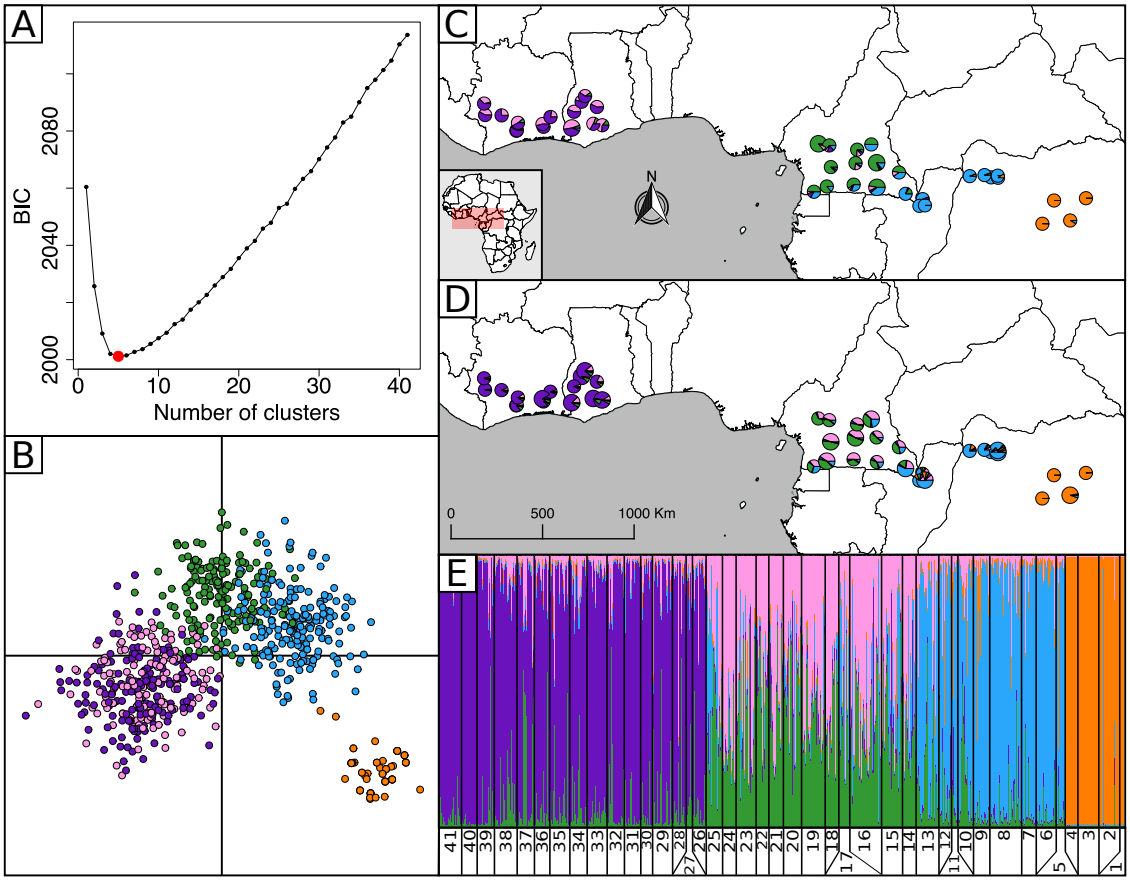


Figure 2: Comparison of DAPC and STRUCTURE *k*=5 clustering patterns
A: Line graph of Baysian Informative Content (BIC) scores for each cluster (k), used to identify the most applicable k (the lowest point) in this case k=5 (represented by the red dot); B: Principal Components scatterplot based on the DAPC output; C: Population pies for DAPC groupings, based on counts of each individual in a population assigning to one of the five groups; D: Population pies for STRUCTURE groupings, based on the CLUMPAK population output file; E: STRUCTURE bargraph based on the individual output file in CLUMPAK, bargraph is orientated left to right with populations organised west to east (see Table 1 for pop number references).

A comparison of the population assignment proportions for *k* values 2-5 identified that clear separation of the genetic clusters could only be found for *k*=3. The absence of well-defined boundaries in *k* values 2, 4 & 5 made them

complex to analyses further: There was an intergrade between the two *k*=2 clusters across the species range; The two central clusters in *k*=4 could not be clearly distinguished apart, and there was reduced confidence in the assignment of individuals within these groups; The two western clusters of the DAPC *k*=5 grouping are equally shared between the Ivory Coast (CIV) and Ghana (GHA) populations.

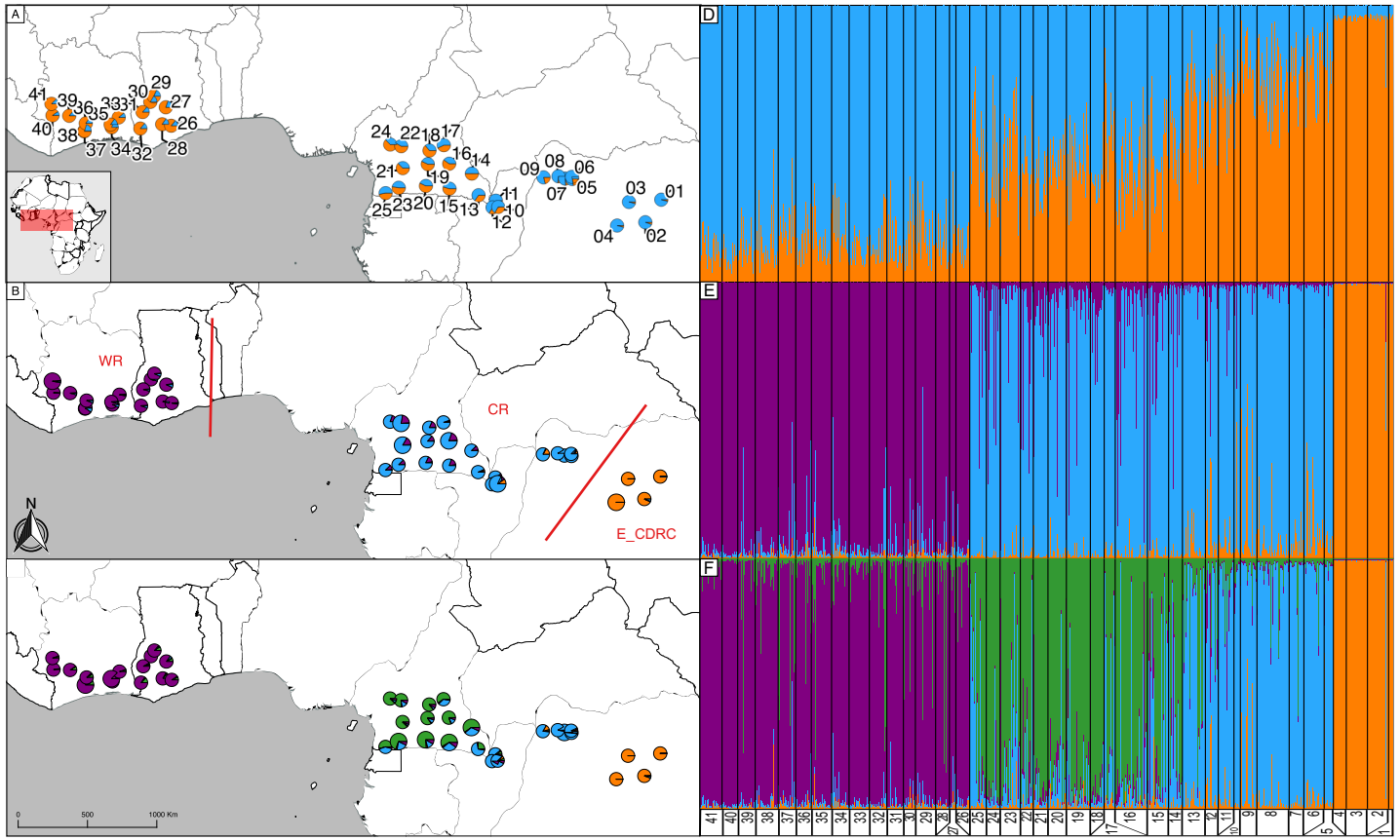


Figure 3: Outputs for most suitable clusters identified from STRUCTURE analysis.
A-C: population pies based on CLUMPAK grouping (A: *k*=2; B: *k*=3; C: *k*=4), D-F: population bar graphs of STRUCTURE analysis (D: *k*=2; E: *k*=3; F: *k*=4).
NB: Numbers in part A & F refer to populations (see table S1 for details); Output pies are the proportion of individuals in each pop assigning to most likely genetic unit, pies generated from CLUMPAK output files; Red lines in B are approximations of genetic unit boundaries; Letter codes in B represent names for the groupings at each k, these are derived from the simple grouping names in Table 1.

The *k*=3 clusters are as follows; an exclusive eastern cluster (E_CDRC), which includes only populations from central Democratic Republic of the Congo (COD) (most easterly sampling locations in this study), a central region cluster (CR), which incorporates all samples from Republic of the Congo (COG) and Cameroon (CMR) as well as the north-western COD population samples, and a western region cluster (WR), that encompasses samples west of the Dahomey Gap (all GHA and CIV samples). Information on which populations are included in each *k*=3 cluster can be seen in table 1 and approximations of breaks between each cluster can be seen in fig. 3B.

The genetic diversity analysis on the *k*=3 grouping (table 2) found H_O ranged from 0.244 (E_CDRC) to 0.261 (CR), H_E ranged from 0.177 (E_CDRC) to 0.266 (CR), G_{IS} ranged from -0.381 (E_CDRC) to 0.022 (WR) and G_{ST} ranged from 0.008 (E_CDRC and WR) to 0.022 (CR). The country genetic diversity analysis (table 2) found H_O ranged from 0.252 (COD) to 0.263 (CMR); H_S ranged from 0.0220 (COD) to 0.271 (CMR); G_{IS} ranged from -0.145 (COD) to 0.033 (CMR); and G_{ST} ranged from 0.001 (GHA) to 0.057 (COD). AMOVAs (standard and nested) of the regrouped datasets found similar results compared to the original population AMOVA. For all tests, most of the variation was found within individual (F_{IT}) whilst, the least amount of variation was occurring between individuals (F_{IS}) (see table S2).

In addition to the distribution wide genetic clustering analysis, a hierarchical clustering approach was also performed in STRUCTURE. This was undertaken for individuals within each of the clusters (from *k*=2 to *k*=5) as well as for each country. We did not identify any additional clustering for any of these novel groups.

The IBD (mantel) tests found that there was a correlation occurring for some tests, but not others (table 2). IBD was significant ($p \leq 0.0001$) for the global test (see fig. 4 for visualisation), and each of the *k*=3 clusters (E_CDRC: $p \leq 0.01$, CR: $p \leq 0.0001$, WR: $p \leq 0.05$). Significant IBD was also found in two Countries (CIV:

$p \leq 0.05$, COD: $p \leq 0.001$). Yet for the other three Countries (GHA, CMR, COG), there was no significant IBD found. See fig. S1 for visualisations of all k=3 genetic clusters and country IBD (mantel) tests.

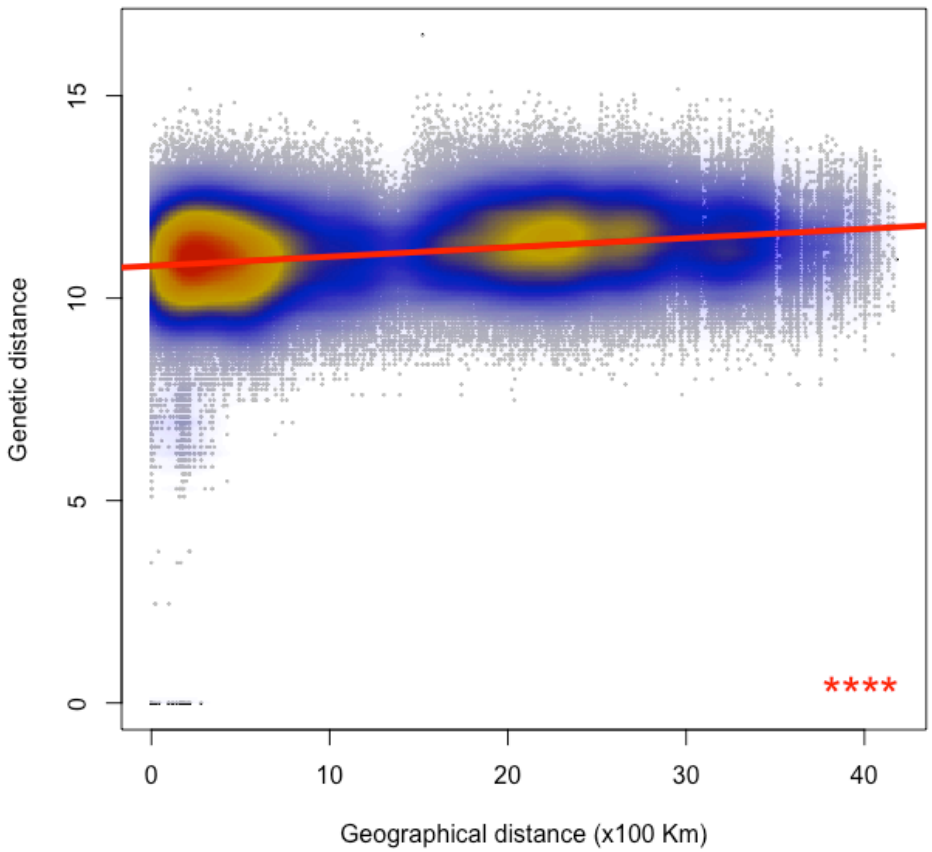


Figure 4: Isolation By Distance (IBD) plot for all samples
*NB: Points are the genetic and geographical distances between two samples; heat map is the density of the comparisons (red is most dense); red line is the line of best fit; red text is the p value of the mantel test results (**** = p≤0.0001)*

Discussion

The breadth of sampling undertaken in this study is extensive and the final dataset incorporates samples from most of the known range of *Triplochiton scleroxylon* (Degen and Sebbenn 2016). Based on the geneflow promoting life history characteristics of the species (e.g. long lived, wind seed dispersal) and the observation of high levels of random mating between individuals (G_{IS} = -0.007) (Table 2), limited genetic structuring was expected. However, the results found there was significant population genetic structuring alongside highly significant IBD correlation ($p \leq 0.0001$) (Table 2, fig. 4)). Yet the highest genetic variation is occurring within individuals (AMOVA test (F_{IT} = ~95%)), which

indicated that the level of structuring may be limited (G_{ST} = 0.058) (Table 2). The clustering analyses indicated that the most likely number of genetic groups across the species' range were *k* values 2-4 (STRUCTURE) (fig. 3) or *k*=5 (DAPC) (fig. 2). Taking into account the effect of random mating, and the lack of clear coherent divisions between genetic clusters, *k*=3 was chosen as the most applicable clustering of *T. scleroxylon* to discuss further. It was decided that the other STRUCTURE groupings provided either limited spatial resolution (*k=2*) or difficulty in defining precise genetic boundaries (*k=4*). The DAPC grouping (*k=5*) was unable to provide any further distinction than STRUCTURE *k=4.* Under the *k*=3 clustering, there was a clear genetic separation across the Dahomey Gap and within the Congo Basin (fig. 3B). While the Dahomey Gap has been observed as an important phylogeographic barrier previously (Hardy et al. 2013), the structure observed within the Congo Basin, which was more pronounced than across Dahomey Gap, has not been previously identified in African tropical rainforest tree species. This novel biogeographical pattern is important in the understanding the biogeography of the Guineo-Congolian phytochoria. These findings also come at a critical time, with the African tropical rainforests at increasing risk from deforestation and the impacts of climate change (Malhi et al. 2013).

In addition to *k*=3 being the most applicable grouping for this study, it is also the most common way that the Guineo-Congolian rainforests are subdivided in the literature. Commonly referred to as the Upper Guinea (UG), Lower Guinea (LG) and Congolia (C) sub-centres (White 1979; White 1983). Originally defined based on local botanical and zoological endemism, they are still used in contemporary publications to describe the structuring patterns for within and between species (Hardy et al. 2013). While not exact, the arrangement of the three sub-centres is similar to the *k*=3 grouping from this study (western region (WR = Upper Guinea; central region (CR) = Lower Guinea; central Democratic Republic of the Congo (E_CDRC) = Congolia). Our results concur that the western and central groups are divided by the Dahomey Gap, the location of the boundary between the central and eastern blocks differs.

Traditionally, the boundary between these two sub-centres was along the Ubangi and Congo Rivers (which is also the Republic of the Congo (COG)/ Democratic Republic of the Congo (COD) border). The rivers had been identified as suitable divides due to their influence as geneflow barriers to plant and animal species (White 1979; White 1983). However, there is little evidence supporting the role of rivers as geneflow barriers in species with long range dispersal mechanisms, such as tree species (Hardy et al. 2013). The only tree species where a genetic barrier that correlates with these historical boundaries was identified, *Milicia excelsa* (Daïnou et al. 2014), was then contradicted in further research (Daïnou et al. 2016). Our results reiterate the notion that rivers are not a geneflow barrier, with samples either side of the Ubangi (COG and North-Western COD samples) or Congo (E_CDRC samples) rivers being genetically similar.

The reason the genetic differentiation of populations within the Congo Basin has remained unknown until now, is because the region has not been widely sampled previously. Most wide-ranging studies, while incorporating samples from the Upper Guinea forests, did not include samples from the Congo Basin (Demenou et al. 2016; Duminil et al. 2013; Duminil et al. 2015; Gomez et al. 2009; Iloh et al. 2017). The only studies on tree species to incorporate samples from the Congo Basin (Daïnou et al. 2016; Daïnou et al. 2014), did not include samples in the centre of the basin, where *T. scleroxylon* samples were collected. A limited number of samples from three locations were also included in a study of *Symphonia globulifera* (Budde et al. 2013), but there were none from north-western COD. For the most part, studies into the genetic structuring patterns of African tropical rainforest species have been limited to the central forests around Cameroon (CMR) (see (Dauby et al. 2014; Hardy et al. 2013) for summary, but also (Bizoux et al. 2009; Born et al. 2011; Born et al. 2008a; Born et al. 2008b; Daïnou et al. 2010; Daïnou et al. 2012; Dauby et al. 2010; Debout et al. 2011; Duminil et al. 2010; Lowe et al. 2010; Muloko-Ntoutoume et al. 2000)). Limited sampling in the Congo Basin is not just restricted to tree species. Studies into other plant (Ley and Hardy 2014; Ley et al. 2014) and animal (Bohoussou et al. 2015; Eaton et al. 2009; Gonder et al. 2011; Leaché and Fujita 2010; Nicolas et

al. 2008; Ntie et al. 2017) groups from the Guineo-Congolia phytocoria have also faced these issues.

While no justification has been provided for this sampling omission by any publication, we postulate that the political instability of the region and the logistical issues in sampling such areas would explain why the genetic patterns within the forests of the Congo Basin have remained unknown until now. Despite this sampling limitation, the genetic patterns from all studies are similar to each other; additionally, they are comparable with the findings of this study. They all identify clear separation of samples either side of the Dahomey Gap, and one or more clusters from the central region, which can be distinct or mixed. The relationship of samples (if any) from eastern locations to the central populations varies. The addition of the findings from this study have enhanced the understanding of genetic structuring patterns in African tropical rainforests, particularly within the Congo Basin.

These structuring patterns are thought to be the result of historical climatic fluctuations during the quaternary period (Hardy et al. 2013; Sexton et al. 2015). These fluctuations have had a significant impact on the genetic structure seen in many African tropical rainforest species including *T. scleroxylon* (Hardy et al. 2013; Sexton et al. 2015). The genetic patterns observed in this study and for many other tree species are likely to be a result of forest contraction/ expansion dynamics that occurred during and since the Last Glacial Maximum (LGM). During the LGM, the African rainforests contracted by up to 84% (Anhuf et al. 2006), reducing the forest cover to isolated refugial pockets scattered across west and central Africa (Anhuf et al. 2006; Maley 1996). The main consequence of this isolation was likely to be a limitation of geneflow between remnant populations.

There are two models recognised describing the layout, size and location of refugia but both have similar patterns (Anhuf et al. 2006; Maley 1996) (see Fig. 1 from Hardy et al. (2013) for an overlay of the two models together). Based on our clustering results, both models have validity in explanting the patterns seen

in *T. scleroxylon*. Both studies present a single refugial population on the border between Ghana (GHA) and the Ivory Coast (CIV), therefore either model can explain the presence a single western (WR) genetic population. However, only Anhuf (2006) identified a single refugia extending along the coast from western Nigeria to southern Gabon which explains the central (CR) genetic population from our study. Yet, the Maley (1996) model would help explain why the *k*=4 (based on the CLUMPAK results) or *k*=5 (based on the DAPC results) clustering results were the most likely grouping of the samples. That model identified multiple small refugial pockets scattered inland from the coast, from western CMR to southern Gabon. But the Maley (1996) proposal for refugia in the Congo Basin would not explain the genetic dissimilarity identified between the north-western and central COD populations. The theoretical extent of the single large crescent polygon encompassing the floodplains of the Congo, and Ubangi rivers, would have included all COD samples. Only the Anhuf (2006) proposal of small refugia within the centre of the Congo Basin would be able to explain why there was no genetic relatedness between the north-western and central COD populations.

Regardless of specific refugial hypotheses, it is accepted that the African rainforests rapidly expanded after the icecap retreat at the end of the LGM. At the peak of Humid Holocene, the forest covered an area larger than seen today and included the forestation of the present-day Dahomey Gap. The loss of this biogeographical barrier to geneflow would have facilitated the transfer of genetic material across a more widespread area and would explain why there was genetic relatedness between the populations either side of the Dahomey Gap. The increased forest expanse would also explain why the north-western COD populations grouped within the central (CR) genetic population, rather than with the eastern cluster (E_CDRC), despite being geographically closer to their compatriots. The refugia that the north-western COD populations originated from could have expanded eastwards from the coast.

The relative uniformity of the African tropical rainforest biome since the end of the LGM has been advantageous for *T. scleroxylon*, as its life history traits are well suited for an ecosystem that is relatively intact. *Triplochiton scleroxylon* is a self-sterile (outcrossing) species, that is wind/insect pollinated and seeds are wind dispersed (Leakey et al. 1981; Oni 1990), which are common life history traits in many tropical rainforest trees (Dick et al. 2008). These characters are capable of facilitating long-range geneflow for a species (Degen and Sebbenn 2016; Dick et al. 2008; Sexton et al. 2015). This potential for extensive geneflow can explain why low overall population differentiation (G_{ST}=0.058 ±0.007) and between the genetic populations (pairwise F_{ST} range 0.044-0.172) was identified. The mating systems employed by *T. scleroxylon* have also been found to be the main determiner of F_{ST} scores (Duminil et al. 2009; Hamrick and Godt 1996). Although there was a relatively low level of population differentiation, there is a considerable amount of overall genetic diversity in *T. scleroxylon* (H_O=0.258 ±0.016, H_T=0.272 ±0.017). A benefit of widespread geneflow is greater overall genetic resilience and adaptability to any potential environmental changes (Degen and Sebbenn 2016; Hamrick et al. 1992). Considering the likelihood of climate change impacting the forest extent in the near future (Malhi et al. 2013), this characteristic of *T. scleroxylon* will likely be beneficial in generations to come. With this in mind, areas that are genetically isolated and have lower genetic diversity than the rest of the species range may be vulnerable; it has been suggested that these should be areas that are targeted for any conservation efforts (Daïnou et al. 2010). Our study identified that the two outer clusters, WR and E_CDRC, had the lowest genetic diversity (G_{ST}=0.008) so conservation efforts in these locations may be worthwhile.

There is scope for future work investigating the genetic structure of *T. scleroxylon* in more depth. Although we have sampled extensively across the range of *T. scleroxylon*, there are still gaps (e.g. parts of COD and COG and completely from Liberia, Nigeria Equatorial Guinea and Gabon). It is possible that without continuous sampling, the genetic separation of our clusters may be overemphasised (Daïnou et al. 2016; Demenou et al. 2016). The inclusion of additional samples would further clarify the genetic structuring patterns in

T. scleroxylon. The benefit of the genotyping approach used in our study is that new samples can be added to the reference dataset without having to repeat the genotyping of the present sample set.

Other potential research areas that would complement the current genetic structuring study, include phylogeographical pattern analyses and investigation of the likely ages of specific linages, as has been utilised for other species (Dauby et al. 2014; Duminil et al. 2015). Unfortunately, the markers used in the current study are not suitable for undertaking any detailed phylogeographic analysis so alternative markers such as barcode sequences would need to be used.

Despite their limited applicability in phylogeographical analysis, the markers developed for this study would be suitable for application in timber tracking. The use of SNP markers to verify and/or support claims of origin for timber importations have been demonstrated and advocated for in recent publications (UNODC 2016; Degen et al. 2017; Dormontt et al. 2015; Lowe et al. 2016). For this to occur a validation of the genetic markers would need to be undertaken within forensic standards and guidelines (SWFS 2015; SWGDAM 2016; SWFS 2018).

Conclusion

The results of this study indicate that the genetic clustering of *T. scleroxylon* is consistent with the key theoretical locations of African Tropical rainforest refugia during the LGM. We identified that contemporary populations of *T. scleroxylon* have limited geneflow across the Dahomey Gap which supports the current understanding of phylogeographic patterns in the area. In addition, we determined there is very strong genetic differentiation between populations in north-western and central COD despite current continuous forest in the area. This divide has not been reported previously. This novel pattern may reflect historical refugial population dynamics or could be actively maintained by as yet unidentified barriers to geneflow across the Congo Basin. This study is the first articulation of three significant biogeographically important clusters across the

African tropical rainforest biome; and identifies the region as an important genetic resource and potential priority for future conservation efforts.

Acknowledgments

The authors are grateful to World Wildlife Fund, Thunen Institute for Forest Genetics, and G2S for the organisation and collection of the test samples for this project. Thank you to the present and past members of the Lowe Lab Group Timber Tracking Team. This work was undertaken with finding from the following: Australian Research Council (linkage), International Tropical timber Organisation (ITTO).

Chapter 4 References

[UNODC] United Nations Office on Drugs and Crime (2016) Best Practice Guide for Forensic Timber Identication. United Nations, United Nations Office at Vienna, Austria

Akinnagbe A (2008) Genetic Variation in *Mansonia Altissima* A. Chev. and *Triplochiton Scleroxylon* K. Schum Under Different Regimes of Human Impact in Akure Forest Reserve, Nigeria. Cuvillier Verlag

Akinnagbe A, Gailing O, Finkeldey R (2010) Genetic diversity of *Mansonia altissima* A. Chev. under different regimes of human impact in the Akure Forest Reserve, Nigeria. Forestry Studies in China 12:193-200 doi:10.1007/s11632-010-0407-5

Anhuf D (2000) Vegetation History and Climate Changes in Africa North and South of the Equator (10° S to 10° N) During the Last Glacial Maximum. In: Smolka P, Volkheimer W (eds) Southern Hemisphere Paleo- and Neoclimates: Key Sites, Methods, Data and Models. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 225-248. doi:10.1007/978-3-642-59694-0_15

Anhuf D et al. (2006) Paleo-environmental change in Amazonian and African rainforest during the LGM. Palaeogeography, Palaeoclimatology, Palaeoecology 239:510-527 doi:10.1016/j.palaeo.2006.01.017

Aranda A et al. (2013) Antigenic proteins involved in occupational rhinitis and asthma caused by obeche wood (*Triplochiton scleroxylon*). PloS ONE 8:e53926 doi:10.1371/journal.pone.0053926

Austerlitz F, Mariette S, Machon N, Gouyon P-H, Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. Genetics 154:1309-1321

Bizoux JP, Daïnou K, Bourland N, Hardy OJ, Heuertz M, Mahy G, Doucet JL (2009) Spatial genetic structure in Milicia excelsa (Moraceae) indicates extensive gene dispersal in a low density wind pollinated tropical tree. Molecular Ecology 18:4398-4408 doi:10.1111/j.1365-294X.2009.04365.x

Bohoussou KH et al. (2015) The phylogeography of the rodent genus *Malacomys* suggests multiple Afrotropical Pleistocene lowland forest refugia. Journal of Biogeography 42:2049-2061 doi:10.1111/jbi.12570

Born C, Alvarez N, McKEY D, Ossari S, Wickings EJ, HOSSAERT‐McKEY M, CHEVALLIER MH (2011) Insights into the biogeographical history of the Lower Guinea Forest Domain: evidence for the role of refugia in the intraspecific differentiation of Aucoumea klaineana. Molecular Ecology 20:131-142 doi:10.1111/j.1365-294X.2010.04919.x

Born C, Hardy OJ, Chevallier MH, Ossari S, Atteke C, Wickings E, Hossaert‐Mckey M (2008a) Small scale spatial genetic structure in the Central African rainforest tree species Aucoumea klaineana: a stepwise approach to infer the impact of limited gene dispersal, population history and habitat fragmentation. Molecular Ecology 17:2041-2050 doi:10.1111/j.1365-294X.2007.03685.x

Born C et al. (2008b) Colonization processes and the maintenance of genetic diversity: insights from a pioneer rainforest tree, Aucoumea klaineana. Proceedings of the Royal Society of London B: Biological Sciences 275:2171-2179 doi:10.1098/rspb.2008.0446

Bowen MR, Howland P, Last FT, Leakey RB (1977) *Triplochiton scleroxylon:* its conservation and future improvement. FAO Forest Genetic Resources Information 6:38-47
Broadhurst L et al. (2017) Genetic diversity and structure of the Australian flora. Diversity and Distributions 23:41-52 doi:10.1111/ddi.12505
Budde K, González-Martínez S, Hardy OJ, Heuertz M (2013) The ancient tropical rainforest tree *Symphonia globulifera* L. f.(Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa. Heredity 111:66-76 doi:10.1038/hdy.2013.21
Daïnou K, Bizoux JP, Doucet JL, Mahy G, Hardy OJ, Heuertz M (2010) Forest refugia revisited: nSSRs and cpDNA sequences support historical isolation in a widespread African tree with high colonization capacity, *Milicia excelsa* (Moraceae). Molecular Ecology 19:4462-4477 doi:10.1111/j.1365-294X.2010.04831.x
Daïnou K et al. (2016) Revealing hidden species diversity in closely related species using nuclear SNPs, SSRs and DNA sequences–a case study in the tree genus *Milicia*. BMC Evolutionary Biology 16:259 doi:10.1186/s12862-016-0831-9
Daïnou K, Laurenty E, Mahy G, Hardy OJ, Brostaux Y, Tagg N, Doucet J-L (2012) Phenological patterns in a natural population of a tropical timber tree species, Milicia excelsa (Moraceae): Evidence of isolation by time and its interaction with feeding strategies of dispersers. American journal of botany 99:1453-1463 doi:10.3732/ajb.1200147
Daïnou K et al. (2014) Speciation slowing down in widespread and long-living tree taxa: insights from the tropical timber tree genus *Milicia* (Moraceae). Heredity 113:74-85 doi:10.1038/hdy.2014.5
Dauby G, Duminil J, Heuertz M, Hardy OJ (2010) Chloroplast DNA polymorphism and phylogeography of a Central African tree species widespread in mature rainforests: Greenwayodendron suaveolens (Annonaceae). Tropical Plant Biology 3:4-13 doi:10.1007/s12042-010-9041-6
Dauby G, Duminil J, Heuertz M, Koffi G, Stevart T, Hardy OJ (2014) Congruent phylogeographical patterns of eight tree species in Atlantic Central Africa provide insights into the past dynamics of forest cover. Molecular Ecology 23:2299-2312 doi:10.1111/mec.12724
Debout GDG, Doucet JL, Hardy OJ (2011) Population history and gene dispersal inferred from spatial genetic structure of a Central African timber tree, *Distemonanthus benthamianus* (Caesalpinioideae). Heredity 106:88-99 doi:10.1038/hdy.2010.35
Degen B, Blanc-Jolivet C, Stierand K, Gillet E (2017) A nearest neighbour approach by genetic distance to the assignment of individual trees to geographic origin. Forensic Science International: Genetics 27:132-141 doi:10.1016/j.fsigen.2016.12.011
Degen B, Sebbenn AM (2016) Genetics and Tropical Forests. In: L P, Köhl M (eds) Tropical Forestry Handbook. Springer-Verlag, Berlin Heidelberg, pp 885-920. doi:10.1007/978-3-642-54601-3_75
Demenou BB, Piñeiro R, Hardy OJ (2016) Origin and history of the Dahomey Gap separating West and Central African rain forests: insights from the phylogeography of the legume tree *Distemonanthus benthamianus*. Journal of Biogeography doi:10.1111/jbi.12688

Dick CW, Hardy OJ, Jones FA, Petit RJ (2008) Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. Tropical Plant Biology 1:20-33 doi:10.1007/s12042-007-9006-6
Dormontt EE et al. (2015) Forensic timber identification: It's time to integrate disciplines to combat illegal logging. Biological Conservation 191:790-798 doi:10.1111/1755-0998.12387
Dray S, Dufour A-B (2007) The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software 22:1-20 doi:10.18637/jss.v022.i04
Duminil J, Brown RP, Ewédjè E-EB, Mardulyn P, Doucet J-L, Hardy OJ (2013) Large-scale pattern of genetic differentiation within African rainforest trees: insights on the roles of ecological gradients and past climate changes on the evolution of *Erythrophleum* spp (Fabaceae). BMC evolutionary biology 13:195 doi:10.1186/1471-2148-13-195
Duminil J, Hardy OJ, Petit RJ (2009) Plant traits correlated with generation time directly affect inbreeding depression and mating system and indirectly genetic structure. BMC Evolutionary Biology 9:177 doi:10.1186/1471-2148-9-177
Duminil J et al. (2010) CpDNA‐based species identification and phylogeography: application to African tropical tree species. Molecular Ecology 19:5469-5483 doi:10.1111/j.1365-294X.2010.04917.x
Duminil J, Mona S, Mardulyn P, Doumenge C, Walmacq F, Doucet JL, Hardy OJ (2015) Late Pleistocene molecular dating of past population fragmentation and demographic changes in African rain forest tree species supports the forest refuge hypothesis. Journal of Biogeography doi:10.1111/jbi.12510
Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4:359-361 doi:10.1007/s12686-011-9548-7
Eaton MJ, Martin A, Thorbjarnarson J, Amato G (2009) Species-level diversification of African dwarf crocodiles (Genus *Osteolaemus*): a geographic and phylogenetic perspective. Molecular Phylogenetics and Evolution 50:496-506 doi:10.1016/j.ympev.2008.11.009
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology 14:2611-2620 doi:10.1111/j.1365-294X.2005.02553.x
Gomez C, Dussert S, Hamon P, Hamon S, de Kochko A, Poncet V (2009) Current genetic differentiation of *Coffea canephora* Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. BMC Evolutionary Biology 9:167 doi:10.1186/1471-2148-9-167
Gonder MK et al. (2011) Evidence from Cameroon reveals differences in the genetic structure and histories of chimpanzee populations. Proceedings of the National Academy of Sciences 108:4766-4771 doi:10.1073/pnas.1015422108
Hall JB, Bada SO (1979) The distribution and ecology of Obeche (*Triplochiton scleroxylon*). Journal of Ecology:543-564

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. Philosophical Transactions of the Royal Society of London B: Biological Sciences 351:1291-1298 doi:10.1098/rstb.1996.0112
Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. Population Genetics of Forest Trees:95-124 doi:10.1007/978-94-011-2815-5_7
Hardy OJ et al. (2013) Comparative phylogeography of African rain forest trees: a review of genetic signatures of vegetation history in the Guineo-Congolian region. Comptes rendus Geoscience 345:284-296 doi:10.1016/j.crte.2013.05.001
Igboanugo ABI (1989) Wood production and structure in laboratory-grown cuttings of *Triplochiton scleroxylon* with changes in light intensity. Forest Ecology and Management 26:179-188 doi:10.1016/0378-1127(89)90119-9
Igboanugo ABI, Iversen P (2004) Conservation and use of *Triplochiton scleroxylon* in moist West Africa. Forest Genetic Resources 31:50-52
Iloh AC, Schmidt M, Muellner-Riehl AN, Ogundipe OT, Paule J (2017) Pleistocene refugia and genetic diversity patterns in West Africa: Insights from the liana *Chasmanthera* dependens (Menispermaceae). PloS ONE 12:e0170511 doi:10.1371/journal.pone.0170511
Jardine DI et al. (2016) Development of SNP markers for ayous (*Triplochiton scleroxylon* K. Schum) an economically important tree species from tropical West and Central Africa. Conservation Genetics Resources:1-11 doi:10.1007/s12686-016-0529-8
Jardine DI, Dormontt EE, van Dijk K-J, Dixon RRM, Dunker B, Lowe AJ (2015) A set of 204 SNP and INDEL markers for bigleaf maple (*Acer macrophyllum* Pursch). Conservation Genetics Resources 7:797-801 doi:10.1007/s12686-015-0486-7
Jombart T (2008) *adegenet*: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405 doi:10.1093/bioinformatics/btn129
Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11:94 doi:10.1186/1471-2156-11-94
Kespohl S et al. (2005) Identification of an obeche (*Triplochiton scleroxylon*) wood allergen as a class I chitinase. Allergy 60:808-814 doi:10.1111/j.1398-9995.2005.00794.x
Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources doi:10.1111/1755-0998.12387
Ladipo D, Leakey R, Grace J (1991) Clonal variation in a four year old plantation of *Triplochiton scleroxylon* K. Schum. and its relation to the predictive test for branching habit. Silvae Genetica 40:130-135
Leaché AD, Fujita MK (2010) Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proceedings of the Royal Society of London B: Biological Sciences:rspb20100662 doi:10.1098/rspb.2010.0662

Leakey RRB (1992) Enhancement of rooting ability in *Triplochiton scleroxylon* by injecting stockplants with auxins. Forest Ecology and Management 54:305-313 doi:10.1016/0378-1127(92)90019-6
Leakey RRB (2004) Clonal approaches to hardwood forestry in the tropics. Paper presented at the Prospects for high-value hardwood timber plantations in the 'dry' tropics of northern Australia, Mareeba, 19-21 October
Leakey RRB, Chapman VR, Longman KA (1982) Physiological studies for tropical tree improvement and conservation. Factors affecting root initiation in cuttings of *Triplochiton scleroxylon* K. Schum. Forest Ecology and Management 4:53-66 doi:10.1016/0378-1127(82)90028-7
Leakey RRB, Coutts MP (1989) The dynamics of rooting in *Triplochiton scleroxylon* cuttings: their relation to leaf area, node position, dry weight accumulation, leaf water potential and carbohydrate composition. Tree Physiology 5:135-146 doi:10.1093/treephys/5.1.135
Leakey RRB, Ferguson NR, Longman KA (1981) Precocious flowering and reproductive biology of *Triplochiton scleroxylon* K. Schum. The Commonwealth Forestry Review:117-126
Leakey RRB, Longman KA (1986) Physiological, environmental and genetic variation in apical dominance as determined by decapitation in *Triplochiton scleroxylon.* Tree Physiology 1:193-207 doi:10.1093/treephys/1.2.193
Leakey RRB et al. (1990) Low-technology techniques for the vegetative propagation of tropical trees. The Commonwealth Forestry Review:247-257
Leakey RRB, Storeton-West R (1992) The rooting ability of *Triplochiton scleroxylon* cuttings: the interactions between stockplant irradiance, light quality and nutrients. Forest Ecology and Management 49:133-150 doi:10.1016/0378-1127(92)90166-7
Ley A, Hardy OJ (2014) Contrasting patterns of gene flow between sister plant species in the understorey of African moist forests–the case of sympatric and parapatric Marantaceae species. Molecular Phylogenetics and Evolution 77:264-274 doi:10.1016/j.ympev.2014.04.026
Ley AC, Dauby G, Köhler J, Wypior C, Röser M, Hardy OJ (2014) Comparative phylogeography of eight herbs and lianas (Marantaceae) in central African rainforests. Frontiers in Genetics 5 doi:10.3389/fgene.2014.00403
Lowe AJ et al. (2018) Standardized genetic diversity life history correlates for improved genetic resource management of Neotropical trees. Diversity and Distributions 24:730-741 doi:10.1111/ddi.12716
Lowe AJ et al. (2016) Opportunities for Improved Transparency in the Timber Trade through Scientific Verification. BioScience 66:990-998 doi:10.1093/biosci/biw129
Lowe AJ, Wong KN, Tiong YS, Iyerh S, Chew FT (2010) A DNA Method to Verify the Integrity of Timber Supply Chains; Confirming the Legal Sourcing of Merbau Timber From Logging Concession to Sawmill. Silvae Genetica 59:263
Maley J (1996) The African rain forest–main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary.

Proceedings of the Royal Society of Edinburgh, Section B: Biological Sciences 104:31-73 doi:10.1017/S0269727000006114
Malhi Y, Adu-Bredu S, Asare RA, Lewis SL, Mayaux P (2013) African rainforests: past, present and future. Phil Trans R Soc B 368:20120312 doi:10.1098/rstb.2012.0312
Mayaka TB (1994) A family of segmented polynomial functions for modelling the border effect on the diameter growth of ayous (*Triplochiton scleroxylon* K. Schum). Forest Ecology and Management 70:275-283 doi:10.1016/0378-1127(94)90093-0
Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes 4:792-794 doi:10.1111/j.1471-8286.2004.00770.x
Muloko-Ntoutoume N, Petit RJ, White L, Abernethy K (2000) Chloroplast DNA variation in a rainforest tree (Aucoumea klaineana, Burseraceae) in Gabon. Molecular Ecology 9:359-363 doi:10.1046/j.1365-294X.2000.00859.x
Newton A, Allnutt T, Gillies A, Lowe A, Ennos R (1999) Molecular phylogeography, intraspecific variation and the conservation of tree species. Trends in Ecology & Evolution 14:140-145 doi:10.1016/S0169-5347(98)01555-9
Nicolas V et al. (2008) Phylogeographic structure and regional history of *Lemniscomys striatus* (Rodentia: Muridae) in tropical Africa. Journal of Biogeography 35:2074-2089 doi:10.1111/j.1365-2699.2008.01950.x
Nketiah T, Newton A, Leakey R (1998) Vegetative propagation of *Triplochiton scleroxylon* K. Schum in Ghana. Forest Ecology and Management 105:99-105 doi:10.1016/S0378-1127(97)00274-0
Ntie S et al. (2017) Evaluating the role of Pleistocene refugia, rivers and environmental variation in the diversification of central African duikers (genera *Cephalophus* and *Philantomba*). BMC Evolutionary Biology 17:212 doi:10.1186/s12862-017-1054-4
Oni O (1990) Between-tree and floral variations in pollen viability and pollen tube growth in obeche (*Triplochiton scleroxylon*). Forest Ecology and Management 37:259-265 doi:10.1016/0378-1127(90)90095-S
Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. Annual Review of Ecology, Evolution, and Systematics:187-214 doi:10.1146/annurev.ecolsys.37.091305.110215
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959
Prohp TP, Onoagbe IO (2009) Anti-diabetic properties and toxicological studies of *Triplochiton scleroxylon* on the heart enzymes in normal and streptozotocin-induced diabetic rabbits. Pakistan Journal of Nutrition 8:1025-1029
Prohp TP, Onoagbe IO (2012) Effects of aqueous extract of *Triplochiton scleroxylon* on some haematological parameters and blood glucose concentrations in non-diabetic rabbits. African Journal of Biotechnology 11:198-202 doi:10.5897/AJB11.2819
Prohp TP, Onoagbe IO, Joel P (2012) Plasma glucose concentration and lipid profile in streptozotocin-induced diabetic rats treated with extracts of

Triplochiton scleroxylon K. Schum International. Journal of Analytical, Pharmaceutical and Biomedical Sciences 1:22-29
Sexton GJ et al. (2015) Influence of putative forest refugia and biogeographic barriers on the level and distribution of genetic variation in an African savannah tree, *Khaya senegalensis* (Desr.) A. Juss. Tree Genetics & Genomes 11:103 doi:10.1007/s11295-015-0933-3
Thioulouse J, Chessel D, Dole S, Olivier J-M (1997) ADE-4: a multivariate analysis and graphical display software. Statistics and Computing 7:75-83 doi:10.1023/A:1018513530268
Venturini M et al. (2004) Cross‐reactivity between obeche wood (*Triplochiton scleroxylon*) and natural rubber latex. Allergy 59:225-228 doi:10.1046/j.13989995.2003.00137.x
White F (1979) The Guineo-Congolian Region and its relationships to other phytochoria. Bulletin du Jardin botanique national de Belgique/Bulletin van de Nationale Plantentuin van Belgie:11-55
White F (1983) The vegetation of Africa: a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa by F White. Natural Resources Research Report XX, UNESCO, Paris, France

Supplementary information for Chapter 4

Supplementary Table 1: Description of Population, Country and genetic cluster groupings of *T. scleroxylon* samples used in this study

Country	Nearest location, region	Pop. code	Pop. #	# Individuals	Latitude	Longitude	k=3 cluster
COD	Kole, Orientale	DRC_22	01	5	1.9744	25.3559	E_CDRC
	Yangambi, Orientale	DRC_02	02	23	0.7581	24.4950	
	Yahila, Orientale	DRC_23	03	23	1.8496	23.6129	
	Simba, Orientale	DRC_25*	04	14	0.5800	22.9747	
	Boyasegbago, Equator	DRC_16	05	10	3.0914	20.5514	CR
	Botikpo, Equator	DRC_34	06	22	3.2054	20.5345	
	Boyagonda, Equator	DRC_11	07	16	3.1128	20.1704	
	Gemena, Equator	DRC_30	08	35	3.2367	19.8107	
	Yembongo, Equator	DRC_17	09	18	3.1788	19.0064	
COG	Ouesso, Sangha	CB_07	10	7	1.5833	16.5588	
	Ouesso, Sangha	CB_03	11	17	1.9235	16.4336	
	Ouesso, Sangha	CB_06	12	14	1.5574	16.2544	
CMR	Yanga, East	C_05	13	25	2.2157	15.5014	
	Menziong, East	C_06	14	15	3.3796	15.1404	
	Adjélu, East	C_04	15	23	2.5683	13.9354	
	Djampiel, East	C_07	16	35	3.9125	13.9184	
	Letta, East	C_15	17	12	4.9117	13.6277	
	Minta, Central	C_14	18	15	4.6232	12.8555	
	Mbama, East	C_08	19	26	3.8901	12.7639	
	Djoum, South	C_03	20	20	2.7168	12.6631	
	Yaounde, Central	C_09	21	16	3.6754	11.4165	
	Bafia, Central	C_13	22	14	4.8567	11.3316	
	Esson, South	C_02	23	22	2.6159	11.1966	
	Tonga, West	C_12	24	15	4.9485	10.7376	
	Nyabessan, South	C_01	25	18	2.3283	10.4901	
GHA	Oda, Eastern	GH_04	26	15	5.9593	-1.0748	WR
	Borobi Forest Reserve, Ashanti	GH_05	27	7	6.9551	-1.3662	
	Nkarabia, Ashanti	GH_03	28	15	6.0397	-1.5621	
	Agosa, Brong Ahafo	GH_08	29	22	7.5503	-2.0153	
	Susanho, Brong Ahafo	GH_06	30	13	7.2447	-2.2016	
	Akrodie, Bring/Ahafo	GH_07	31	18	6.6972	-2.6156	
	Enchi, Western	GH_02	32	19	5.8105	-2.7389	
CIV	Aukope, Agnebi	CIV_06	33	22	6.4093	-3.9037	
	Agboville, Agnebi	CIV_04	34	19	5.8758	-4.2830	
	Rubino, Agnebi	CIV_07	35	22	6.0239	-4.3418	
	Garéko, Sud-bandama	CIV_08	36	17	6.0916	-5.6813	
	Gauge, Sud-bandama	CIV_03	37	19	5.6789	-5.7433	
	Gauge, Sud-bandama	CIV_02	38	25	5.6595	-5.7684	
	Issia, Sud-bandama	CIV_09	39	19	6.4965	-6.5748	
	Guiglo, Sud-bandama	CIV_10	40	17	6.5168	-7.4761	
	Logoualé, Sud-bandama	CIV_12	41	24	7.1423	-7.5308	

*NB: *population DRC_25 includes three additional individuals from neighbouring populations (see methods section for more information); Country codes (ISO designated): COD = Democratic Republic of the Congo, COG = Republic of the Congo, CMR = Cameroon, GHA = Ghana, CIV = Ivory Coast/Côte d'Ivorie; see fig. 3B for approximate boundaries of genetic clusters.*

Supplementary Table 2: Genetic diversity analyses for Population, Country and Genetic groups

Country	Population	#	Num	Eff_num	H_O	H_S	H_t	H'_t	G_{IS}	G_{ST}	G'_{ST} (Nei)	G'_{ST} (Hed)	G''_{ST}	D_est	Mantel Test (p value)
COD															
	DRC_22		1.419	1.298	0.250	0.179	0.179		-0.396						
	DRC_02		1.533	1.313	0.243	0.184	0.184		-0.321						
	DRC_23		1.419	1.314	0.261	0.178	0.178		-0.469						
	DRC_25		1.419	1.284	0.224	0.167	0.167		-0.338						
	DRC_16		1.752	1.388	0.243	0.245	0.245		0.008						
	DRC_34		1.771	1.408	0.254	0.248	0.248		-0.025						
	DRC_11		1.771	1.438	0.264	0.264	0.264		0.001						
	DRC_30		1.838	1.440	0.274	0.263	0.263		-0.044						
	DRC_17		1.790	1.414	0.258	0.251	0.251		-0.029						
Summary COD		*166*			0.252	0.220			-0.145	0.057	0.063	0.075	0.081	0.019	***
COG															
	CB_07		1.724	1.391	0.251	0.253	0.253		0.006						
	CB_03		1.800	1.411	0.251	0.251	0.251		0.000						
	CB_06		1.771	1.450	0.278	0.274	0.274		-0.014						
Summary COG		*35*			0.260	0.259			-0.003	0.006	0.009	0.009	0.012	0.003	NS
CMR															
	C_05		1.876	1.436	0.267	0.265	0.265		-0.006						
	C_06		1.829	1.425	0.262	0.262	0.262		0.000						
	C_04		1.886	1.439	0.266	0.269	0.269		0.008						
	C_07		1.895	1.450	0.261	0.272	0.272		0.042						
	C_15		1.819	1.443	0.283	0.272	0.272		-0.044						
	C_14		1.857	1.478	0.271	0.290	0.290		0.067						
	C_08		1.895	1.447	0.251	0.271	0.271		0.074						
	C_03		1.829	1.455	0.267	0.275	0.275		0.027						
	C_09		1.857	1.439	0.254	0.271	0.271		0.062						
	C_13		1.867	1.438	0.271	0.275	0.275		0.011						
	C_02		1.876	1.444	0.253	0.272	0.272		0.068						
	C_12		1.790	1.412	0.243	0.255	0.255		0.046						
	C_01		1.886	1.461	0.263	0.280	0.280		0.062						
Summary CMR		*256*			0.263	0.271			0.033	0.007	0.008	0.01	0.011	0.003	NS
GHA															
	GH_04		1.829	1.428	0.242	0.267	0.267		0.094						
	GH_05		1.829	1.463	0.298	0.298	0.298		0.001						
	GH_03		1.829	1.438	0.252	0.268	0.268		0.062						
	GH_08		1.800	1.425	0.258	0.260	0.260		0.006						
	GH_06		1.790	1.414	0.259	0.258	0.258		-0.003						
	GH_07		1.829	1.418	0.255	0.256	0.256		0.005						
	GH_02		1.829	1.447	0.263	0.269	0.269		0.023						
Summary GHA		*109*			0.261	0.268			0.026	0.001	0.001	0.001	0.002	0	NS
CIV															
	CIV_06		1.800	1.449	0.268	0.267	0.267		-0.002						
	CIV_04		1.800	1.424	0.253	0.256	0.256		0.012						
	CIV_07		1.838	1.435	0.253	0.262	0.262		0.035						
	CIV_08		1.752	1.407	0.241	0.247	0.247		0.026						
	CIV_03		1.800	1.436	0.254	0.262	0.262		0.030						
	CIV_02		1.848	1.432	0.262	0.263	0.263		0.005						
	CIV_09		1.819	1.422	0.251	0.257	0.257		0.022						
	CIV_10		1.781	1.435	0.250	0.259	0.259		0.036						
	CIV_12		1.829	1.456	0.268	0.271	0.271		0.008						
Summary CIV		*184*			0.256	0.261			0.019	0.008	0.009	0.011	0.012	0.003	*
	OSX statistic				0.019	0.092			0.333	0.103	0.114	0.135	0.146	0.034	
	p value				NS	*			*	NS	NS	NS	NS	NS	
k=3															
Summary E_DRC		65			0.244	0.177			-0.381	0.008	0.011	0.011	0.013	0.002	**
Summary CR		395			0.261	0.266			0.016	0.022	0.023	0.030	0.031	0.009	****
Summary WR		293			0.258	0.264			0.022	0.008	0.008	0.011	0.011	0.003	*
	OSX statistic				0.022	0.124			0.566	0.020	0.019	0.028	0.027	0.008	
	p value				*	****			****	NS	NS	NS	NS	NS	
Overall (±SD)			2.000	1.393 (0.030)	0.258 (0.016)	0.256 (0.016)	0.272 (0.017)	0.273 (0.017)	-0.007 (0.004)	0.058 (0.007)	0.059 (0.008)	0.078 (0.010)	0.079 (0.010)	0.022 (0.003)	****

*NB: See Table 1 for population codes. # = number of samples in each country/genetic cluster. Num = Number of alleles, Eff_num = Effective number of alleles in a population, H_O = Observed Heterozygosity, H_S = Heterozygosity Within Populations, H_t = Total Heterozygosity, H'_t = Corrected total Heterozygosity, G_{IS} = Inbreeding Coefficient, G_{ST} = Fixation Index, G'_{ST} (Nei) = Nei, Corrected Fixation Index, G'_{ST} (Hed) = Hedrick, Standardised Fixation Index, G''_{ST} = Corrected Standardised Fixation Index, D_est = Jost, Population Differentiation, OSX statistic = Goudet (1995) test statistic, p values: NS = p ≥0.05, * p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001, **** p ≤ 0.0001. Mantel test results taken from Adegenet mantel test (Jombart 2008). Remainder of results are from either compare groups (OSX statistics and p values only) or Genetic Diversity tests in GENODIVE.*

Supplementary Table 3: Results of Genetic Diversity Analysis for Loci

Locus	Num	Eff_num	H_o	H_s	H_t	H'_t	G_{IS}	G_{ST}
A_1018	2	1.100	0.098	0.094	0.095	0.096	-0.040	0.017
A_1056	2	1.054	0.055	0.053	0.055	0.055	-0.032	0.028
A_1521	2	1.217	0.180	0.184	0.190	0.190	0.021	0.031
A_1625	2	1.187	0.165	0.162	0.166	0.166	-0.014	0.023
A_168	2	1.055	0.051	0.054	0.056	0.056	0.054	0.032
A_1684	2	1.112	0.108	0.104	0.107	0.107	-0.047	0.035
A_208	2	1.063	0.059	0.062	0.063	0.063	0.050	0.018
A_23178	2	1.777	0.450	0.452	0.486	0.487	0.003	0.071
A_2440	2	1.079	0.075	0.076	0.077	0.077	0.016	0.015
A_253	2	1.606	0.373	0.390	0.424	0.425	0.045	0.081
A_2724_2	2	1.789	0.477	0.455	0.469	0.469	-0.049	0.03
A_2942_1	2	1.907	0.517	0.490	0.496	0.496	-0.055	0.012
A_315	2	1.013	0.013	0.013	0.013	0.013	-0.002	-0.004
A_3628_1	2	1.029	0.030	0.029	0.030	0.030	-0.024	0.009
A_368	2	1.092	0.090	0.087	0.088	0.088	-0.033	0.007
A_4_2	2	1.167	0.154	0.147	0.149	0.149	-0.044	0.01
A_4037	2	1.892	0.470	0.488	0.495	0.496	0.038	0.015
A_407	2	1.767	0.464	0.448	0.498	0.499	-0.036	0.099
A_411	2	1.316	0.261	0.248	0.269	0.270	-0.054	0.08
A_435	2	1.180	0.161	0.157	0.161	0.161	-0.024	0.027
A_5394	2	1.319	0.242	0.250	0.265	0.265	0.034	0.056
A_55	2	1.198	0.159	0.171	0.174	0.175	0.066	0.021
A_626	2	1.146	0.139	0.131	0.155	0.155	-0.060	0.149
A_642_1	2	1.483	0.344	0.336	0.482	0.486	-0.023	0.303
A_642_2	2	1.011	0.012	0.011	0.011	0.011	-0.003	-0.002
A_6645_2	2	1.830	0.457	0.469	0.484	0.484	0.024	0.032
A_665	2	1.258	0.224	0.211	0.217	0.217	-0.059	0.024
A_698	2	1.686	0.449	0.419	0.421	0.421	-0.070	0.003
A_71422	2	1.022	0.021	0.023	0.023	0.023	0.075	0.007
A_77365	2	1.769	0.437	0.449	0.473	0.474	0.026	0.051
A_827	2	1.645	0.393	0.405	0.417	0.417	0.030	0.027
A_8898	2	1.836	0.475	0.470	0.490	0.490	-0.009	0.04
A_929	2	1.127	0.112	0.116	0.121	0.121	0.034	0.041
A_935	2	1.567	0.377	0.374	0.384	0.384	-0.008	0.026
A_9516_1	2	1.705	0.416	0.427	0.468	0.469	0.026	0.086
A_9516_2	2	1.220	0.185	0.186	0.191	0.192	0.006	0.029
A_961	2	1.043	0.043	0.042	0.042	0.042	-0.015	-0.006
A_CS_110	2	1.271	0.220	0.220	0.229	0.229	0.004	0.039
A_CS_165_1	2	1.177	0.154	0.156	0.160	0.160	0.013	0.027
P0112	2	1.497	0.363	0.342	0.362	0.362	-0.062	0.054
P0133	2	1.483	0.338	0.336	0.368	0.369	-0.004	0.087
P0182	2	1.932	0.514	0.498	0.500	0.500	-0.034	0.006
P0245	2	1.022	0.022	0.022	0.022	0.022	-0.018	0.008
P0265	2	1.591	0.369	0.384	0.406	0.406	0.040	0.053
P0616	2	1.339	0.262	0.261	0.273	0.274	-0.002	0.044
P0761	2	1.876	0.499	0.482	0.496	0.496	-0.036	0.029
P0785	2	1.899	0.460	0.490	0.499	0.499	0.062	0.019
P0855	2	1.020	0.021	0.020	0.021	0.021	-0.062	0.049
P0896	2	1.614	0.392	0.393	0.397	0.397	0.003	0.01
P0917	2	1.300	0.254	0.238	0.242	0.242	-0.067	0.017
P1064	2	1.017	0.016	0.017	0.017	0.017	0.078	0.025
P1103	2	1.891	0.490	0.487	0.496	0.496	-0.007	0.019
P1165	2	1.273	0.222	0.222	0.236	0.237	-0.001	0.063
P1265	2	1.621	0.403	0.396	0.413	0.414	-0.018	0.042
P1477	2	1.204	0.191	0.175	0.188	0.188	-0.092	0.071

Locus	Num	Eff_num	H_o	H_s	H_t	H'_t	G_{IS}	G_{ST}
P1481	2	1.009	0.010	0.009	0.010	0.010	-0.092	0.081
P1547	2	1.075	0.076	0.072	0.073	0.073	-0.054	0.015
P1559	2	1.539	0.363	0.362	0.458	0.460	-0.003	0.21
P1835	2	1.016	0.017	0.016	0.017	0.017	-0.057	0.047
P1894	2	1.563	0.355	0.372	0.385	0.385	0.047	0.032
P1918	2	1.091	0.085	0.086	0.090	0.091	0.015	0.05
P1960	2	1.727	0.450	0.434	0.453	0.454	-0.037	0.041
P2290	2	1.488	0.356	0.338	0.384	0.385	-0.054	0.12
P2328	2	1.100	0.096	0.094	0.098	0.098	-0.015	0.042
P2496	2	1.116	0.104	0.107	0.112	0.113	0.031	0.047
P2644	2	1.912	0.512	0.492	0.495	0.495	-0.040	0.006
P2722	2	1.114	0.101	0.106	0.114	0.114	0.050	0.071
P2775	2	1.611	0.388	0.392	0.403	0.404	0.008	0.029
P2967	2	1.244	0.201	0.203	0.213	0.213	0.007	0.047
P3076	2	1.686	0.422	0.420	0.431	0.431	-0.004	0.025
P3093	2	1.749	0.444	0.443	0.467	0.468	-0.003	0.053
P3137	2	1.334	0.268	0.258	0.274	0.274	-0.039	0.056
P3285	2	1.100	0.094	0.094	0.095	0.096	-0.004	0.018
P3414	2	1.474	0.316	0.333	0.370	0.371	0.050	0.102
P3880	2	1.006	0.007	0.006	0.007	0.007	-0.023	0.02
P3997	2	1.612	0.388	0.392	0.394	0.394	0.010	0.004
P4492	2	1.179	0.159	0.157	0.163	0.163	-0.013	0.039
P4629	2	1.103	0.098	0.096	0.098	0.098	-0.014	0.017
P4638	2	1.870	0.474	0.481	0.497	0.497	0.013	0.032
P4706	2	1.876	0.521	0.481	0.499	0.500	-0.083	0.037
P5240	2	1.361	0.285	0.274	0.294	0.295	-0.039	0.069
P5264	2	1.608	0.394	0.390	0.418	0.419	-0.010	0.067
P5404	2	1.417	0.305	0.304	0.361	0.362	-0.005	0.158
P5439	2	1.019	0.019	0.019	0.019	0.019	-0.033	0.023
P5462	2	1.907	0.480	0.492	0.499	0.499	0.024	0.014
P5532	2	1.559	0.356	0.371	0.386	0.387	0.039	0.04
P5562	2	1.667	0.390	0.414	0.452	0.453	0.058	0.084
P5574	2	1.004	0.005	0.004	0.005	0.005	-0.029	0.027
P5587	2	1.308	0.238	0.243	0.250	0.250	0.023	0.025
P5777	2	1.696	0.429	0.424	0.484	0.486	-0.012	0.125
P5888	2	1.107	0.104	0.100	0.102	0.102	-0.042	0.018
P5944	2	1.227	0.188	0.191	0.201	0.201	0.017	0.047
P5972	2	1.207	0.178	0.177	0.184	0.184	-0.004	0.035
P6163	2	1.142	0.136	0.128	0.131	0.131	-0.063	0.024
P6225	2	1.669	0.436	0.413	0.488	0.490	-0.056	0.154
P6238	2	1.478	0.351	0.333	0.405	0.407	-0.053	0.178
P6290	2	1.816	0.436	0.465	0.475	0.476	0.062	0.022
P6328	2	1.008	0.008	0.008	0.008	0.008	-0.019	0.015
P6392	2	1.506	0.354	0.347	0.465	0.468	-0.022	0.253
P6527	2	1.492	0.362	0.340	0.347	0.347	-0.066	0.02
P6533	2	1.741	0.456	0.439	0.466	0.467	-0.037	0.057
P6704	2	1.251	0.192	0.208	0.218	0.218	0.075	0.045
P6715	2	1.003	0.003	0.003	0.003	0.003	-0.044	0.041
P6741	2	1.444	0.325	0.317	0.323	0.323	-0.025	0.018
P6787	2	1.779	0.459	0.452	0.455	0.455	-0.015	0.007
Overall	*2*	*1.393*	*0.258*	*0.256*	*0.272*	*0.273*	*-0.007*	0.058
(± SD)		*(0.030)*	*(0.016)*	*(0.016)*	*(0.017)*	*(0.017)*	*(0.004)*	*(0.007)*

NB: genetic diversity values were analysed using Genetic Diversity tests in GENODIVE: Num = Number of alleles, Eff_num = Effective number of alleles in a population, H_O = Observed Heterozygosity, H_S = Heterozygosity Within Populations, H_t = Total Heterozygosity, H'_t = Corrected total Heterozygosity, G_{IS} = Inbreeding Coefficient, G_{ST} = Fixation Index.

Supplementary Table 2: AMOVA tests. Standard (A) and nested (B) Analysis of Molecular Variance (AMOVA) tests on population, country or genetic cluster orientated datasets

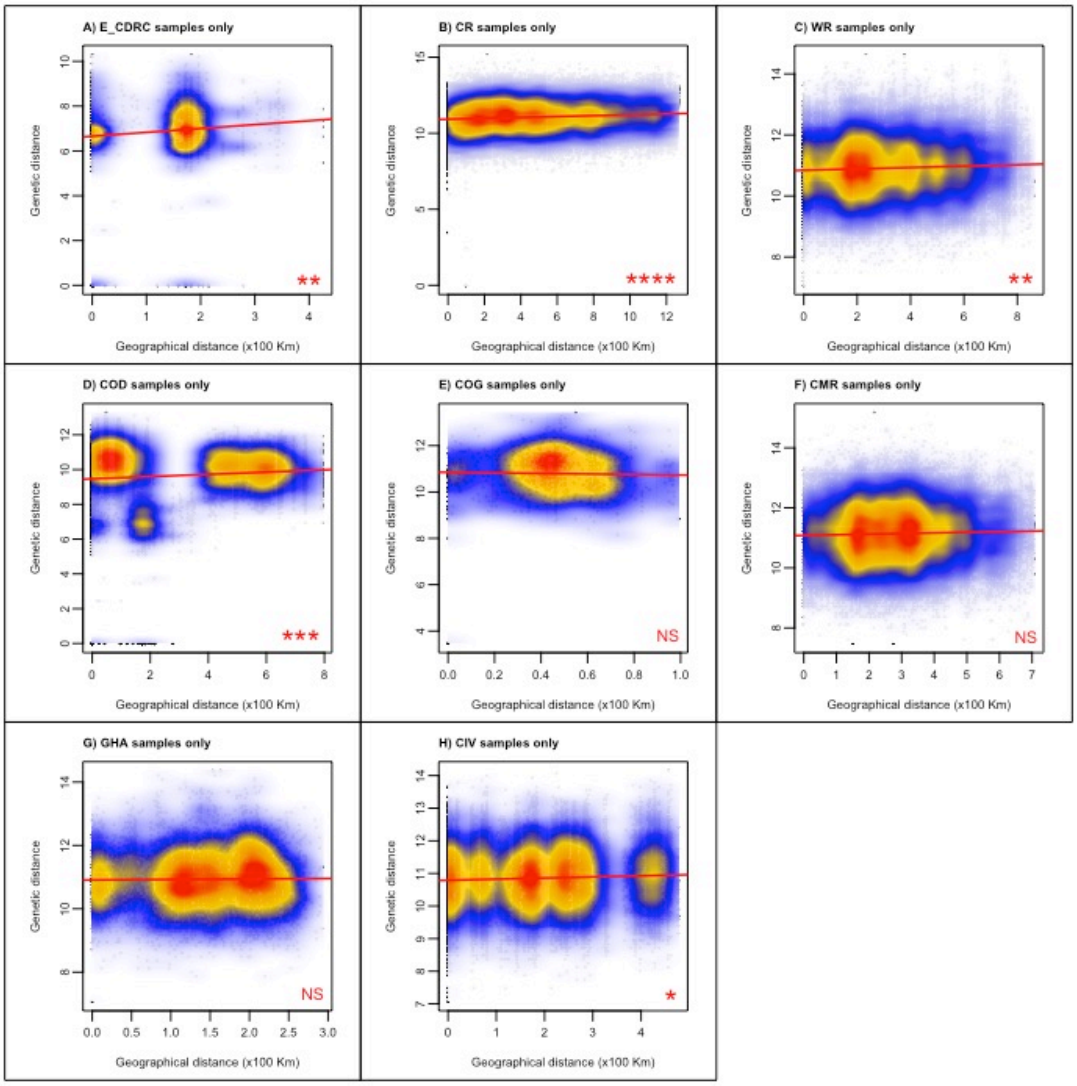
A) Standard

Grouping	Source of Variation	%var	F-stat	F-value	SD (±)	CI 2.5%	CI 97.5%	F'-value
Population	Within Individual	94.8	F_{IT}	0.052	0.008	0.038	0.068	
	Among Individual	-0.5	F_{IS}	-0.005	0.004	-0.012	0.002	
	Among Population	5.7	F_{ST}	0.057	0.007	0.044	0.072	0.077
k=3	Within Individual	92.0	F_{IT}	0.080	0.012	0.059	0.104	
	Among Individual	0.9	F_{IS}	0.010	0.004	0.002	0.019	
	Among Population	7.0	F_{ST}	0.070	0.011	0.052	0.093	0.095
Country	Within Individual	93.7	F_{IT}	0.063	0.009	0.046	0.082	
	Among Individual	1.1	F_{IS}	0.011	0.004	0.003	0.019	
	Among Population	5.2	F_{ST}	0.052	0.008	0.038	0.069	0.071

B) Nested

Grouping	Source of Variation	Nested in	%var	F-stat	F-value	SD (±)	CI 2.5%	CI 97.5%	F'-value
k=3	Within Individual	--	92.0	F_{IT}	0.080	0.012	0.059	0.104	
	Among Individual	Population	-0.4	F_{IS}	-0.005	0.004	-0.012	0.003	
	Among Population	k=3	1.5	F_{SC}	0.016	0.002	0.013	0.020	0.022
	Among k=3	--	6.9	F_{CT}	0.069	0.011	0.050	0.091	0.093
Country	Within Individual	--	93.7	F_{IT}	0.063	0.009	0.046	0.083	
	Among Individual	Population	-0.5	F_{IS}	-0.005	0.004	-0.012	0.002	
	Among Population	Country	1.8	F_{SC}	0.019	0.002	0.015	0.023	0.025
	Among Country	--	5.0	F_{CT}	0.050	0.008	0.035	0.067	0.068

NB: AMOVA Analyses performed in GENODIVE using the AMOVA test. p values removed from results as per Meirmans (2015) suggestion. %var = percentage of the variance in the data explained by the F-stat used, F-stat = F statistic used in analysis, F-value = the value of the F statistic, SD = Standard Deviation, CI = 2.5 and 97.5% confidence intervals.



Supplementary Figure 1: Isolation By Distance (IBD) plots.
Plots of genetic clusters (A-C) and countries (D-H)
*NB: Points are the genetic and geographical distances between two samples; heat map is the density of the comparisons (red is most dense) heat map made using MASS; red line is the line of best fit; red text is the p value of the mantel test results, p values: NS = p ≥0.05, * p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001, **** p ≤ 0.0001. Genetic distances are representative of the samples being analysed hence why they do not all start at 0.*

Supplementary Table 5: Pairwise F_{ST} scores at Country and Genetic cluster groupings

Group	Pairing		F_{ST}
k=3	E_CDRC	CR	0.105
	E_CDRC	WR	0.172
	CR	WR	0.044
Country	COD	COG	0.023
	COD	CMR	0.05
	COD	GHA	0.103
	COD	CIV	0.102
	COG	CMR	0.024
	COG	GHA	0.068
	COG	CIV	0.062
	CMR	GHA	0.038
	CMR	CIV	0.039
	GHA	CIV	0.006

NB: Pairwise F_{ST} scores for all individuals reassigned at k=3 genetic clusters and country groupings. Analysis was done using the GENODIVE Pairwise Differentiation test. See Figure 2 and Table S1 for group inclusions.

References

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. Journal of heredity 86:485-486

Jombart T (2008) *adegenet*: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405 DOI:10.1093/bioinformatics/btn129

Meirmans PG (2015) Seven common mistakes in population genetics and how to avoid them. Molecular ecology 24:3223-3231 DOI:10.1111/mec.13243

Chapter 5: Assignment testing in ayous

Using genetic assignment to identify the geographical region of origin of timber from the African tropical rainforest tree species, ayous (Triplochiton scleroxylon K. schum)

DI Jardine[1], EE Dormontt[1], AJ Lowe[1*]

[1]School of Biological Sciences & Environment Institute, University of Adelaide, Adelaide, SA 5005, Australia

*Author for communication (Andrew.lowe@adelaide.edu.au)

intention to submit to: Forensic Sciences International: Genetics

Statement of Authorship

Title of Paper	Using genetic assignment to identify the geographical region of origin of timber from the African tropical rainforest tree species, ayous (Triplochiton scleroxylon K. schum)
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Intending to submit to Forensic Sciences International: Genetics

Principal Author

Name of Principal Author (Candidate)	Duncan Iain Jardine		
Contribution to the Paper	designed the study; conducted lab work & data analysis; interpreted data; conducted literature review; wrote manuscript as principal author		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	13.08.2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

xiii. the candidate's stated contribution to the publication is accurate (as detailed above);

xiv. permission is granted for the candidate in include the publication in the thesis; and

xv. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Eleanor E Dormontt		
Contribution to the Paper	assisted with study design; supervised experiments; advised on focus & structure of manuscript; edited manuscript 10%		
Signature		Date	05.08.2019

Name of Co-Author	Andrew J Lowe		
Contribution to the Paper	as principal supervisor; obtained funding for the research; advised on focus & structure of manuscript; edited manuscript 10%		
Signature		Date	10.07.2019

Title: *Using genetic assignment to identify the geographic origin of timber from the African tropical rainforest tree species, ayous (Triplochiton scleroxylon K. schum*)

Abstract

Assignment testing is the allocation of an individual to a putative population of origin with which it has greatest similarity, most often based on genetic profiles. This approach has been proposed for validating the claimed origin of legally traded species, including timber. A significant proportion of illegally logged timber is entering legitimate supply chains, so methods capable of verifying the origin of products, such as genetic assignment testing, are vitally important to support sustainable timber trade.

Genetic assignment testing can also fit within legislation that has been implemented by major timber importing nations to reduce illegitimate harvesting. However, the natural population genetic structure of many timber species rarely reflect legal boundaries or meaningful jurisdictions; and can influence the reliability of tests seeking to assign to these artificial geopolitical groups. Whilst many publications have alluded to this situation, there has been little exploration of its effect within a specific legal scenario.

To better understand the potential role of assignment testing in verifying timber legality claims, origin assessments were performed on blind timber samples using a genetic reference database of ayous, a commonly traded tropical African timber species. Results showed that claims of test samples were most successfully verified when assigning to geo-genetic groups. There was a lack of geopolitical/geo-genetic cohesion present in ayous, which reduced the capacity to assign to country and concession groups. Nevertheless, assignments to legally informative country groups were successfully conducted by employing a likelihood measure to support decision-making.

These results demonstrate that the use of assignment testing to verify country of origin claims in timber samples shows sound potential, as long as appropriate assessment approaches are utilised. Additionally, blind testing samples from timber material should be used where possible to verify confidence in genetic assignment methods.

Key Words: *forensic botany, genetic assignment testing, illegal wildlife trade, origin assessment, timber tracking, traceability, wildlife forensics*

Introduction

Assignment testing is a statistical approach used to allocate an individual to the population with which it shares most similarity (Ng et al. 2017). It can be used with any technique where a stable measured distinction can be identified at various geographical locations, such as genetic profiles (Ng et al. 2017). Assignment tests have been proposed as a mechanism to verify the claimed origin of samples from traded species such as fish (Hansen et al. 2001; Nielsen 2016) and timber (Dormontt et al. 2015; Lowe and Cross 2011). Additionally, the use of assignment testing in a wildlife forensic context is well supported (Johnson et al. 2014; Ogden and Linacre 2015). In timber for example, assignment testing has been identified as the only dependable evidence-based technique for verifying the origin of samples (Dormontt et al. 2015; Finch et al. 2017). Considering that an estimated 10-50% of all timber traded globally is from illegal sources (Jianbang et al. 2016; Nellemann and INTERPOL 2012), any method that can verify the origin and identify potential mislabelling in timber importations would be well received (Finch et al. 2017).

Assignment testing can support existing legislation implemented by major timber importing and exporting nations to control the trade of illegal timber (Dormontt et al. 2015; Lowe et al. 2016; Norman and Saunders 2017). Under these laws, for timber importations to be approved several statements are required. This includes a declaration of the origin of the timber and presentation of evidence regarding the certainty of the claim (Finch et al. 2017; Ng et al. 2017). Results from assignment tests have the potential to provide supporting evidence to verify the legitimacy (or lack thereof) of the claimed origin of consignments (Ogden et al. 2009). Therefore, if results from assignment testing are to be utilised within this legal framework, they need to be conclusive and beyond reasonable doubt (i.e. meet international legal standards) (Johnson et al. 2014).

Assignment tests have the greatest reliability when used to determine the natural population of origin of a sample. However, the arrangement of these biological groups may not align with those required for legal enforcement (e.g. nation-states/country or concession/geographic population) (Ogden et al. 2009). Consequently, the accuracy of assignment testing may be reduced when data is restructured to something that is more legally meaningful, i.e. strictly assigning to a country (Degen et al. 2017). In essence, it can be a "square peg round hole" scenario. This effect was highlighted in recent publications on Malaysian timber species *Gonystylus bancanus* (Ng et al. 2016) and *Shorea platyclados* (Ng et al. 2017), where an obvious difference in the self-assignment results for samples between concessions and genetic regions was found.

Without an accurate understanding of the statistical certainty surrounding an assignment result, there is also a greater potential for incorrect results to occur when verifying a samples origin (Lowe et al. 2010). An erroneous assignment result could have significant judicial and ethical consequences and could negatively affect the credibility of the method, technology and/or the practitioner/laboratory providing the results. However, despite these critical issues, there is a lack of assessment of these errors when assigning to various spatial scales (e.g. to concession, country or genetic population) in the wildlife forensics literature. Without such reporting, it is difficult to apply assignment testing within a legal context (Johnson et al. 2014; Ogden et al. 2009). Research that is focused specifically on the direct application of assignment testing in a legal context will increase our understanding of how assignment testing in timber species can be utilised to help reduce the trade of illegally sourced timber.

This study was conducted using ayous (*Triplochiton scleroxylon* K. schum), a widely distributed, wind dispersed tree species from the African Guineo-Congolian rainforest (Hall and Bada 1979; Igboanugo and Iversen 2004). ayous is one of the most economically important tree species in the region and is at presently listed as "of least concern" on the IUCN redlist (1998). ayous was selected due to its ubiquitous use and the absence of current restrictions that can hinder research efforts (such as the requirements for CITES permits). Further, the development of tools to support the sustainable exploitation of timber resources are best developed and applied before

extensive negative impacts on wild populations are realised. Our approach was to conduct assignment tests on blind timber samples (i.e. those for which the origin was known but was kept hidden from the investigator until after an assessment of the claimed origin had been made) to evaluate the effectiveness of the assignment and the capacity to detect falsified claims.

Ayous is a species whose biological (geo-genetic) population structuring is not concordant to the country and concession (geo-political) boundaries where it is naturally found (Ch. 4). The lack of geo-genetic/geopolitical alignment will allow assessment of rates of errors (especially Type I) when assigning to geopolitical groups, compared to genetic population assignments. A greater understanding of the strengths and limitations of genetic assignment approaches in ayous will help facilitate the application of reliable scientific verification tools in timber supply chains.

Materials/Methods

To assess the effectiveness of assignment testing in ayous, we utilised a two-step process. Firstly, a reference dataset was developed by clustering the samples into the three grouping arrangements (hereafter grouping levels) used previously (Ch. 4) (concession (i.e. geographical population), country, genetic population); each level representing a spatial scale that could be utilised in assignment testing. Then assignment tests were performed on these reference datasets to determine the suitability of each for assignment testing. In the second stage of this study, blind samples were assigned to their most likely group within the reference datasets. This study used pre-existing Single Nucleotide Polymorphic (SNP) loci and genotyping results (Ch. 4). Test samples were genotyped using all 105 loci from that study. The genotyping results for 753 individuals formed the reference datasets.

Firstly, the data was grouped according to the 41 geographical populations of collection (hereafter concession groups). This high-resolution verification of sample claims is a requirement (when applicable) of several timber laws (e.g. Australia and the EU) (Lowe et al. 2016), so an understanding of the capacity of assignment testing at this level is of

interest. Secondly, the data was arranged by country of collection (hereafter country groups), which consisted of the Democratic Republic of the Congo (COD), the Republic of the Congo (COG), Cameroon (CMR), Ghana (GHA) and the Ivory Coast (CIV). These groups are used for verifying country of origin claims, which is needed for all timber trading laws (Lowe et al. 2016). For the third level (hereafter genetic groups), the data were organised into the three genetic populations identified in Ch. 4. Each group covered a range of geopolitical areas (E_CDRC: a sub national Eastern group, with samples from a Central COD region; CR: a multinational Central group, consisting of samples from the North Western regions of COD, as well as all samples from COG and CMR; and WR: a dual national Western group consisting of all samples from GHA and CIV). Information on the arrangement of reference samples into the three grouping levels can be found in Figure 1 and Table S1.

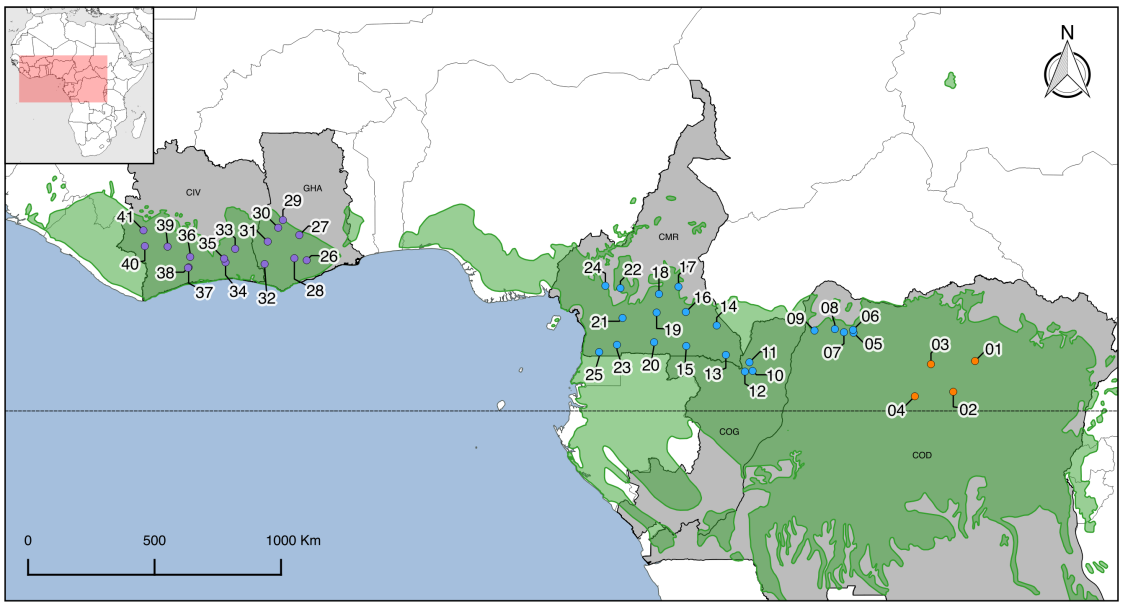


Figure 1: Location of ayous reference populations used in the assignment testing.
NB: numbers on dots correlate with their population names in Table S1, Colours for each group represent the three genetic population clusters (Purple = WR, Blue = CR and Orange = E_CDRC (see Table S1 for more information)), Green represents forest cover of the Guineo-Congolian rainforest and the approximate distribution of ayous (NB: above the equator (dotted line) only), grey shaded countries are those that have reference samples.

Assignment tests were undertaken on all the reference samples at each of the three grouping levels to understand the inherent reliability through self-assignment. The more reference samples from a given group self-assign, the greater the confidence in the assignment outcome of test samples. For reference groups with low numbers of self-assigning samples, there is limited reliability. In GENODIVE (Meirmans and Van

Tienderen 2004); the program used for this study, the proportion of self-assigning reference samples for each group (at each grouping level) is expressed as a likelihood ratio. The lower the ratio, the more self-assigning reference samples in a group. These likelihood ratios were then used to evaluate the significance of a test samples assignment result (see further in methods for details).

All assignment tests were performed using the "*Population assignment*" function from GENODIVE (NB: GENODIVE uses a Paetkau (1995) frequency-based model and a "leave one out" approach to calculate the most similar group to assign to). Standard parameters were used for all tests (i.e. Allele frequencies: Calculate from marker data; Zero frequencies: Replace with frequency = 0.005; Test statistic: Likelihood ratio (l_{home}/l_{max}); Significance: Threshold (alpha -level) = 0.002, Apply threshold separately to every population; Permutations: Number of datasets = 10000, Self-fertilization rate = 0).

To demonstrate the usability of assignment testing in ayous, we used a blind experiment approach to assess the claimed origin of 25 timber (test) samples. For each test sample, a set of coordinates claiming to be the origin were provided (by either WWF or ITTO) some of which had been falsified, while other locations were legitimate. After the genuine coordinates of the test samples were disclosed; it was found that five of the samples were previously unsampled trees from existing reference populations whose legitimate origin was not altered (*the VTI samples*). These were authentication samples, the results of which should be comparable to those in the corresponding reference group and highlight how well assignment testing works for samples from within an area included in the reference dataset. The legitimate origin of the remaining 20 samples had been altered. The genuine origin of ten of these samples (*the WWF samples*) was known with a high degree of precision (to concession group level), whilst for the other ten samples (*the G2S samples*) only the country was known. These were Investigation samples, as the samples were from locations external to the reference dataset; and were included to assess how well the assignment test works for samples from novel locations. For further information regarding the claimed and actual (legitimate) origin of the test samples see Table 1.

Table 1: Test sample information. *Information on the extractions and locations (both claim and genuine origin) of the test samples used in this analysis*

			Claimed Origin							Legitimate origin				
								Assigned Reference Group[3]					Nearest Reference Group[4]	
Supplier	Sample Number	Tested[1]	Name	Pooled[2]	Latitude	Longitude	Country	Concession	Genetic	Latitude	Longitude	Country	Concession	Genetic
VTI	C_03_TRI_05	Yes	a001	Yes	2.75759	12.82677	CMR	20	CR			*Not altered*		
	CIV_04_TRI_15	Yes	a006	Yes	5.88983	-4.23523	CIV	34	WR			*Not altered*		
	CIV_02_TRI_14	Yes	a007	Yes	6.39746	-3.89501	CIV	38	WR			*Not altered*		
	CIV_07_TRI_15	Yes	a008	Yes	5.97563	-4.30329	CIV	35	WR			*Not altered*		
	GH_04_TRI_20	Yes	a016	Yes	5.84042	0.82384	GHA	26	WR			*Not altered*		
WWF	BT_2014_533	Yes	a025	Yes	6.3	-0.022	GHA	26	WR	5.57883	-2.25545	GHA	32	WR
	BT_2014_543	Yes	a097	No	3.15017	13.61635	CMR	16	CR	1.630097	12.0897	GAB	-	-
	BT_2014_547	No	-	-	-5	16	COD			6.41525	-1.20916	GHA	27	WR
	BT_2014_551	Yes	a096	Yes	5.9	-3.75	CIV	34	WR	5.35566	-2.26029	GHA	32	WR
	BT_2014_563	Yes	a024	Yes	6	-5	CIV	36	WR	6.49586	-2.47783	GHA	31	WR
	BT_2014_567	Yes	a019	Yes	-2	19	COD	4	E_CDRC	6.01586	-2.04149	GHA	28	WR
	BT_2014_568	Yes	a020	Yes	3.518	13.44178	CMR	16	CR	2.54425	11.92747	CMR	33	CR
	BT_2014_578	Yes	a092	Yes	5	10	CMR	24	CR	6.58086	-2.38226	GHA	31	WR
	BT_2014_594	Yes	a099	No	7.75	-0.988	GHA	29	WR	6.41922	-2.37705	GHA	31	WR
	BT_2014_598	Yes	a091	Yes	6.4	-1.2	GHA	27	WR	6.40983	-1.20743	GHA	27	WR
G2S	G2S_O_T1	No	-	-	6.5	-1.52	GHA			-	-	GHA	-	WR
	G2S_O_T3	Yes	a098	No	6.5	-1.52	GHA	27	WR	-	-	GHA	-	WR
	G2S_O_T6	No	-	-	2.1	11.6	CMR			-	-	COG	-	CR
	G2S_O_T7	No	-	-	2.43	17.25	COG			-	-	COG	-	CR
	G2S_O_T9	Yes	a100	Yes	0.94	20.89	COD	4	E_CDRC	-	-	CMR	-	CR
	G2S_O_T11	Yes	a084	No	3.2	14.28	CMR	16	CR	-	-	CMR	-	CR
	G2S_O_T13	Yes	a081	Yes	2.1	11.6	CMR	23	CR	-	-	CMR	-	CR
	G2S_O_T15	No	-	-	2.1	11.6	CMR			-	-	CMR	-	CR
	G2S_O_T16	No	-	-	2.1	11.6	CMR			-	-	CMR	-	CR
	G2S_O_T18	Yes	a090	Yes	3.2	14.28	CMR	16	CR	-	-	CMR	-	CR

NB: [1]the samples claim was either tested or not (samples not tested either due to genotyping failure or because the level of detail regarding its legitimate origin was insufficient to assess the claim, [2] duplicate extractions completed and pooled for a sample, [3]this is the nearest reference group for the concession or genetic groups based on the provided coordinates, [4]this is the closest reference group to the legitimate coordinates provided. For reference concession, country (except GAB = Gabon) and genetic cluster codes see Table S1.

Timber material for all test samples were collected externally (by the suppliers (either VTI, WWF or G2S)) and sent to our facility (the Advanced DNA, Identification and Forensics Facility (ADIFF), University of Adelaide, Australia) for extraction and analysis. For every test sample, duplicate DNA extractions were carried out and remained independent. DNA extractions were undertaken with either an in-house extraction protocol (IPN# WO 2015/070279 A1) (Lowe et al. 2015)), or a modified INNUPREP Plant DNA Kit extraction (Analytik Jena AG, Jena, Germany), see Sup Material 1 for modified protocol. The test sample extractions were then genotyped using the same procedure as Jardine (2016), on the MassARRAY® iPLEX™ platform (Agena Bioscience™, San Diego, USA), using the iPLEX™ GOLD chemistry (Agena Bioscience™) at the Australian Genome Research Facility (AGRF, Brisbane, Australia).

Replicate genotypes of each test sample were compared and combined if matching ≥95%. or treated singularly if not. A final screen to remove test samples (either single or paired) that did not meet the genotyping cut-off (≥95% as per Ch. 4) was then carried out before analysis. Test samples with usable genotypes were then allocated to the relevant group (the claim group) for each grouping level (for concession, to nearest reference group based on the provided coordinates; for country, the claimed country of origin; and genetic, the most relevant). Test samples were added to the reference data collectively, but examined individually, using the include/exclude function of GENODIVE, and run using the same parameters as previously. Each test sample assigned to the reference group with the most similar profile (the inferred group), which could be either the originating claim group, or a different group. For test samples that assigned to a different group, the likelihood ratio of the assignment to the inferred group was recorded.

Counter to an approach previously proposed by Ogden (2009), the most likely origin of the test samples was not determined. Instead, we assessed the probability that a sample did not originate from whence it was claimed (as recommended by Lowe & Cross (2011)), a technique more in line with the requirements of existing timber legislation. The decision to accept or reject the

claimed origin of a test sample was determined based on the inferred group that it assigned to, with the legitimate origins of each sample then used to evaluate whether the decision was correct or incorrect. For test samples where the claim and inferred group were the same, the claim was accepted and not assessed. Whereas when this was not the case and samples assigned to different inferred groups (i.e. where inferred ≠ claim group), then the assignment was evaluated to determine whether to reject or accept the claim. This procedure was implemented because the function of assignment testing in GENODIVE is set up to detect migrants (likelihood of different group assignment), rather than identify support for an individual being in the correct group.

For test samples that assigned to different inferred groups, three strategies were employed to evaluate the allocation. The objective of this approach was to further understand the influence that decision-making plays in assignment testing. This was done by designing the strategies to emphasise how statistical support for an assignment can be used as a decision-making tool and how it impacts the outcome when not used correctly. For a summary of the strategies and how they were implemented see Table 2. For the first strategy (hereafter Strategy A), the inferred group was considered the legitimate origin of a test sample and the claim was rejected. This strategy was used to understand the significance (and effect) of not evaluating the assignment, focusing particularly on the prevalence of problematic Type I errors (i.e. incorrect rejection of a legitimate claim) that indicate a miscarriage of justice. In contrast to this, the other two strategies utilised an assessment measure to inform the decision-making process. This was done by using likelihood ratio calculations as a threshold measure to identify the support of a test sample's assignment to the inferred group. If the likelihood ratio of the test sample to the inferred group was larger than that of the claim reference group (from the reference data self-assignment test), it indicated that the sample was a migrant and not from the claim group. Reciprocally, a smaller number implies that the sample may still come from the claim group. For both assessment strategies, if the assignment ratio was larger, then there was sufficient support to reject the claim. However, the strategies differed in the treatment of test samples with smaller likelihood

ratios. For one approach (hereafter Strategy B), the claim was accepted if there was insufficient support to reject. Whereas the other (hereafter Strategy C), allowed for a "No Decision" verdict, regarding the claim's legitimacy. These two strategies would highlight the impact of making a poorly supported decision (i.e. Strategy B) that could lead to the formation of Type II errors (i.e. incorrect acceptance of a falsified claim) and result in an error of impunity, over not providing a decision (Strategy C) when there is limited support in the assignment, which would reduce of the number of results generated.

In addition to these three strategies, a recreation of what currently exists for most timber imports (i.e. assignment testing not used to evaluate the legitimacy of a consignments claim), and the claimed origins of all test samples were accepted. For this approach (hereafter Strategy D) there are only two possible outcomes, either the claims were legitimate or false, with a falsified claim being a type II error.

Table 2: Summary of the decision-making strategies used for test samples that assign to inferred groups that are different to the claim group

			errors expected			Assessment technique	
Strategy	**About**	**Reason**	**Type I**	**Type II**	**Decision if Inferred ≠ Claim**	**LR <**	**LR >**
A	No assessment of LR (always reject)	Impact of not assessing outcome	Yes	No	Reject	No assessment	
B	Only reject if LR is larger, else accept	Impact of making a judgement on poorly supported results	No	Yes	Assess significance	Accept	Reject
C	Do not comment on results with smaller LR and reject claims for those with a larger LR	Impact of not making a judgment leading to a reduction in results	No	No	Assess significance	No Decision	Reject
D	No Test (accept all claims as legitimate)	Status quo (for DNA testing)	No	Yes			

NB: LR = Likelihood ratio

Results

Reference sample self-assignments

For the reference samples, there was a significant difference in the proportion that assigned to the correct groups (the self-assignment proportion) between each of the three grouping levels (i.e. genetic, country or concession groups). Overall, the genetic group assignments had the highest self-assignment proportion (97% (*n=732*)); followed by the country groups (76% (*n=569*)); with the concession groups having the lowest (19% (*n=140*)). This was also reflected in the overall likelihood ratios for each dataset, where it was smallest (best) at

the country (LR=13.763) and genetic group (LR=9.292) levels, and largest (worst) at the concession grouping level (LR=37.043).

Breaking down the reference sample assignments to each grouping level, the self-assignment proportions at the genetic group level were even between the three groups 96-98% (see Table S4 for all genetic group assignments). At the country level, for all groups the majority of reference samples self-assigned, with CMR (85% (*n=217*)) the highest and GHA (65% (*n=71*)) having the lowest proportions. At this grouping level, there were clear patterns to the samples that did not self-assign, the largest (31% (*n=34*)) was for GHA samples assigning to CIV (see Table S3 for all country group assignments). For the Concession group assignments, there were five groups that did not have any self-assigning samples; and only four groups had more than half of their samples correctly self-assigned. For samples that did not self-assign, there was no apparent pattern to the allocations to particular concession groups (see Table S2 for all concession group assignments).

The results for all grouping levels are mirrored in the likelihood ratios of each group, with larger scores found in groups with few self-assigning samples, and smaller scores in groups with high self-assigning proportions. The likelihood ratios for all reference groups at the three grouping levels including the overall likelihood ratios can be seen in Table 3.

Table 3: Likelihood ratios for the assignment testing of the reference data at each of the three grouping levels

Grouping level	Group name	likelihood ratio
Concession	DRC_22	72.218
	DRC_02	28.386
	DRC_23	17.206
	DRC_25	29.579
	DRC_16	44.218
	DRC_34	22.639
	DRC_11	29.642
	DRC_30	19.392
	DRC_17	30.63
	CB_07	66.947
	CB_03	30.994
	CB_06	31.153
	C_05	25.397
	C_06	33.249
	C_04	26.797
	C_07	22.343
	C_15	41.94
	C_14	35.173
	C_08	25.295
	C_03	27.495
	C_09	34.575
	C_13	38.913
	C_02	27.844
	C_12	31.54
	C_01	31.888
	GH_04	37.765
	GH_05	76.326
	GH_03	36.854
	GH_08	26.383
	GH_06	39.542
	GH_07	28.926
	GH_02	28.707
	CIV_06	25.071
	CIV_04	30.423
	CIV_07	28.187
	CIV_08	28.894
	CIV_03	28.181
	CIV_02	24.328
	CIV_09	32.164
	CIV_10	30.099
	CIV_12	23.951
	Overall for dataset	*37.043*
Country	COD	16.952
	COG	16.077
	CMR	12.448
	GHA	10.524
	CIV	9.827
	Overall for dataset	*13.763*
Genetic	E_CDRC	20.561
	CR	9.37
	WR	6.641
	Overall for dataset	*9.292*

Test samples

Duplicate DNA extractions were successfully carried out on all 25 test samples, with all 50 extracts being genotyped. Of the 50 genotypes produced (two replicates for each test sample), 30 could be paired (15 test samples), four were unusable but the paired genotypes were successful (four test samples), and 12 failed altogether (six test samples). A total of 19 test sample genotypes (with ≥95% assay coverage) were then used in the assignment tests. See Table 1 for all test sample extraction and genotyping information.

Table 4: Results of the test sample assignments at the three grouping levels

Grouping	Sample	Claim group	Inferred group (LR)	Actual group	Strategy SA Call	SA Decision	SB Call	SB Decision	SC Call	SC Decision	SD Call	SD Decision
Concession	a001	20	25 (9.233)	20	R	I	A	C		NDM	A	C
	a006	34	38 (23.351)	34	R	I	A	C		NDM	A	C
	a007	33	41 (22.009)	33	R	I	A	C		NDM	A	C
	a008	35	35	35	A	C		-		-	A	C
	a016	26	37 (15.817)	26	R	I	A	C		NDM	A	C
	a019	4	36 (154.842)	28	R	C	R	C	R	C	A	I
	a020	16	18 (0.338)	23	R	C	A	I	NDM		A	I
	a024	36	41 (15.588)	31	R	C	A	I	NDM		A	I
	a025	26	35 (14.069)	32	R	C	A	I	NDM		A	I
	a081	23	15 (8.231)	CMR	Sample not assessed							
	a084	16	14 (12.28)	CMR	Sample not assessed							
	a090	16	22 (3.68)	CMR	Sample not assessed							
	a091	27	28 (7.082)	27	R	I	A	C		NDM	A	C
	a092	24	39 (37.82)	31	R	C	R	C	R	C	A	I
	a096	34	31 (22.734)	32	R	C	A	I		NDM	A	I
	a097	16	24 (1.963)	GAB	Sample not assessed							
	a098	27	31 (18.826)	GHA	Sample not assessed							
	a099	29	36 (4.436)	31	R	C	A	I		NDM	A	I
	a100	4	22 (124.174)	CMR	R	C	R	C	R	C	A	I

					Strategy							
					SA		SB		SC		SD	
Grouping	Sample	Claim group	Inferred group (LR)	Actual group	Call	Decision	Call	Decision	Call	Decision	Call	Decision
Country	a001	CMR	CMR	CMR	A	C		-		-	A	C
	a006	CIV	CIV	CIV	A	C		-		-	A	C
	a007	CIV	CIV	CIV	A	C		-		-	A	C
	a008	CIV	CIV	CIV	A	C		-		-	A	C
	a016	GHA	GHA	GHA	A	C		-		-	A	C
	a019	COD	GHA (39.419)	GHA	R	C	R	C	R	C	A	I
	a020	CMR	CMR	CMR	A	C		-		-	A	C
	a024	CIV	CIV	GHA	A	I		-		-	A	I
	a025	GHA	CIV (0.346)	GHA	R	I	A	C		NDM	A	C
	a081	CMR	CMR	CMR	A	C		-		-	A	C
	a084	CMR	COG (1.691)	CMR	R	I	A	C		NDM	A	C
	a090	CMR	CMR	CMR	A	C		-		-	A	C
	a091	GHA	GHA	GHA	A	C		-		-	A	C
	a092	CMR	CIV (33.625)	GHA	R	C	R	C	R	C	A	I
	a096	CIV	GHA (6.378)	GHA	R	C	A	I		NDM	A	I
	a097	CMR	CMR	GAB				Sample not assessed				
	a098	GHA	CIV (4.555)	GHA	R	I	A	C		NDM	A	C
	a099	GHA	GHA	GHA	A	C		-		-	A	C
	a100	COD	CMR (25.840)	CMR	R	C	R	C	R	C	A	I

Strategy

SA SB SC SD

Grouping level	Sample	Claim group	Inferred group (LR)	Actual group	Call	Decision	Call	Decision	Call	Decision	Call	Decision
Genetic	a001	CR	CR	CR	A	C		-		-	A	C
	a006	WR	WR	WR	A	C		-		-	A	C
	a007	WR	WR	WR	A	C		-		-	A	C
	a008	WR	WR	WR	A	C		-		-	A	C
	a016	WR	WR	WR	A	C		-		-	A	C
	a019	E_CDRC	WR (124.843)	WR	R	C	R	C	R	C	A	I
	a020	CR	CR	CR	A	C		-		-	A	C
	a024	WR	WR	WR	A	C		-		-	A	C
	a025	WR	WR	WR	A	C		-		-	A	C
	a081	CR	CR	CR	A	C		-		-	A	C
	a084	CR	CR	CR	A	C		-		-	A	C
	a090	CR	CR	CR	A	C		-		-	A	C
	a091	WR	WR	WR	A	C		-		-	A	C
	a092	CR	WR (42.086)	WR	R	C	R	C	R	C	A	I
	a096	WR	WR	WR	A	C		-		-	A	C
	a097	CR	CR	GAB				Sample not assessed				
	a098	WR	WR	WR	A	C		-		-	A	C
	a099	WR	WR	WR	A	C		-		-	A	C
	a100	E_CDRC	CR (96.468)	CR	R	C	R	C	R	C	A	I

NB: assignments performed using GENODIVE Population assignment test function. NDM=No Decision Made (as per strategy 3), LR = Likelihood ratio, samples that couldn't be assessed were because insufficient information regarding their legitimate origin was provided or (as for a097) was outside the reference dataset. A = Accept claim call made, R = reject claim call made, (call made by self-assignment result or LR score (> than home group), C = Correct call made, I = Incorrect call made (decision based on legitimate origin of sample) Group codes can be found in Sup Table A (except GAB = Gabon)

All 19 test samples were able to be assigned for the three grouping levels. Results of test samples, including likelihood ratios for those assigning to different groups to the claim can be seen in Table 4; additionally, visualisations of all test sample assignments can be seen in Figure S1.

As previously stated, the test samples were categorised into two types (authentication (*the VTI test samples*) or investigative (*the WWF and G2S test samples*)), and this was revealed after the true origin of the samples had been disclosed (i.e. unblinded). All five authentication test samples (i.e. new samples from existing concessions) were successfully used in the assessment. For both the country and genetic grouping level assignments, all five authentication test samples self-assigned to the correct group. At the concession level, only one self-assigned to the claim group, with the other four erroneously assigning to different groups. In all four instances the Likelihood Ratio of these assignments to the inferred group was less than that of the claim reference groups (the implications of this will be reported further in results).

The remaining 14 test samples were investigative (i.e. individuals from unsampled locations). The legitimate origin of all of these samples had been altered to some extent. Only one investigative test sample (a091) remained in the same concession group (location altered by 10 km). A higher proportion of the investigative test samples had legitimate country (62%) or genetic (77%) group claims. See Table 1 for all test sample legitimate origins.

For some investigative test samples, the level of detail regarding their legitimate origin, limited how/when/where they could be used. There were five samples that had to be removed from the analysis (either completely or parts of), when this was disclosed. One individual (a097) was removed entirely because it came from Gabon, an unsampled country. A further four samples (a081, a084, a090 and a098) were excluded from only the concession of origin tests, because of a lack of spatial resolution regarding the legitimate origin (only country of origin known), and inability to assess the result (claim and inferred concession groups in same country).

The highest proportion of investigative test samples to self-assign was when assigning to the genetic groups (77% (10/13), and only three test samples (a019, a092, a100) assigned to different groups than claimed. See figure 3 for a visualisation of the assignment of investigative test sample a092 at the three grouping levels.

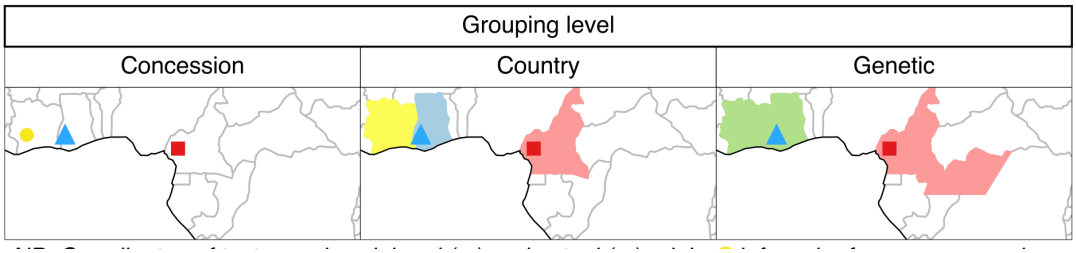


Figure 3: Assignment results of test sample a092.
NB: Coordinates of test samples claimed (■) and actual (▲) origin, inferred reference concession (●) group assignment. Country/Genetic group shading for: claim group (▰), inferred group (▰), actual group (▰) or if inferred & actual group the same (▰).

None of the investigative test samples self-assigned to the claim group at the concession level. Nevertheless, despite successfully assigning to different groups than the claim, the likelihood ratio of the assignments was only higher than the claim reference group in three instances (a019, a092, a100). Additionally, these were the only investigative test samples that assigned with significance (i.e. Likelihood Ratio to inferred group > Likelihood Ratio of claim reference group) at the Country and Genetic Grouping level as well. Three additional investigative test samples assigned to different country groups than claimed, yet the assignments were unsupported (i.e. Likelihood ratio to inferred group < likelihood ratio of claim reference group). The remaining seven investigative test samples self-assigned at the Country grouping level. Six of these self-assignments were correct, with only one (a024), incorrectly assigning back to the claimed country group (see Figure 2 for visualisation of this).

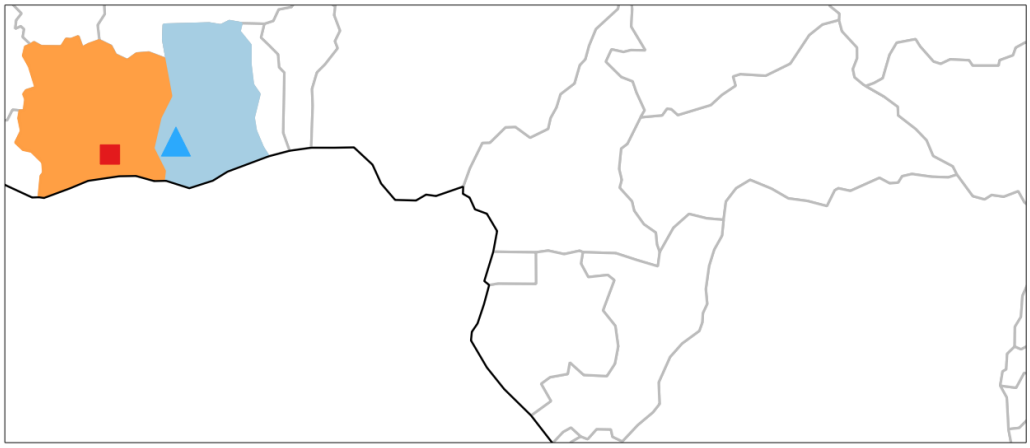


Figure 2: test sample a024 incorrect self-assignment to country group
NB: Coordinates of test samples claimed (■) and actual (▲) origin. Shading for actual country (▰) and claim & inferred group (▰) as the same country

For all of the test samples that assigned to different groups, there were noticeable differences between the three strategies employed for evaluating the results. Type I errors (incorrect rejection of a legitimate claim) were only generated when the inferred group was considered the legitimate origin of the sample (as per Strategy A). When using this strategy, they were generated in both the concession (*n = 5*) and country (*n = 3*) group tests. Four of the concession level Type I errors occurred in the authentication samples, whose origins had not been falsified. If instead the significance of the assignments were assessed (as per Strategies B & C), only the claims of the three samples with well supported assignments to their inferred groups (a019, a092 and a100) could be correctly rejected regardless of the grouping level. Yet, incorrect results (i.e. incorrect rejection of samples with legitimate claims) only occurred when attempting to adjudicate in tests samples whose assignments were insufficient (as per Strategy B). These Type II errors occurred in both concession (*n = 5*) and country (*n = 1*) level assessments. When the strictest decision-making strategy was used, and decisions were only formed if an assignment was well supported (i.e. Strategy C) then no incorrect results occurred. Unfortunately, it limited the number of samples that could be assessed (*n=3* for all grouping levels). For a summary of the assignments using the various strategies at the three grouping levels see Tables 5-7.

Table 5: Summary of assessment strategies of test samples at the concession grouping level.

	Strategy											
	A			**B**			**C**			**D**		
	Total	**ATS**	**ITS**	**Total**	**ATS**	**ITS**	**Total**	**ATS**	**ITS**	**Total**	**ATS**	**ITS**
count of all results	**14**	5	9	**13**	4	9	**3**	0	3	**14**	5	9
correct acceptance	**1**	1	0	**5**	4	1	**-**	-	-	**6**	5	1
correct rejection	**8**	0	8	**3**	0	3	**3**	0	3	**-**	-	-
total correct results	**9**	1	8	**8**	4	4	**3**	0	3	**6**	5	1
percentage correct results	***64***	*20*	*89*	***62***	*100*	*44*	***100***	*-*	*100*	***43***	*100*	*11*
incorrect acceptance [II]	**0**	0	0	**5**	0	5	**-**	-	-	**8**	0	8
incorrect rejection [I]	**5**	4	1	**0**	0	0	**0**	0	0	**-**	-	-
Total incorrect results	**5**	4	1	**5**	0	5	**0**	0	0	**8**	0	8
percentage incorrect results	***36***	*80*	*11*	***38***	*0*	*56*	***0***	*-*	*0*	***57***	*0*	*89*

NB: [I] Incorrect rejection equates to a type I error (i.e. incorrect rejection of a sample with a legitimate claim), [II] Incorrect acceptance equates to a type II error (i.e. incorrect acceptance of a sample with a falsified claim), ATS = Authentication Test Samples, ITS = Investigative Test Samples

Table 6: Summary of assessment strategies of test samples at the country grouping level.

	Strategy											
	A			**B**			**C**			**D**		
	Total	**ATS**	**ITS**	**Total**	**ATS**	**ITS**	**Total**	**ATS**	**ITS**	**Total**	**ATS**	**ITS**
count of all results	**18**	5	13	**7**	0	7	**3**	0	3	**18**	5	13
correct acceptance	**10**	5	5	**3**	0	3	**-**	-	-	**13**	5	8
correct rejection	**4**	0	4	**3**	0	3	**3**	0	3	**-**	-	-
Total correct results	**14**	5	9	**6**	0	6	**3**	0	3	**13**	5	8
percentage correct results	***78***	*100*	*69*	***86***	*-*	*86*	***100***	*-*	*100*	***72***	*100*	*62*
incorrect acceptance [II]	**1**	0	1	**1**	0	1	**-**	-	-	**5**	0	5
incorrect rejection [I]	**3**	0	3	**0**	0	0	**0**	0	0	**-**	-	-
Total incorrect results	**4**	0	4	**1**	0	1	**0**	0	0	**5**	0	5
percentage incorrect results	***22***	*0*	*31*	***14***	*-*	*14*	***0***	*0*	*0*	***28***	*0*	*38*

NB: [I] Incorrect rejection equates to a type I error (i.e. incorrect rejection of a sample with a legitimate claim), [II] Incorrect acceptance equates to a type II error (i.e. incorrect acceptance of a sample with a falsified claim), ATS = Authentication Test Samples, ITS = Investigative Test Samples

Table 7: Summary of assessment strategies of test samples at the Genetic grouping level.

	Strategy											
	A			**B**			**C**			**D**		
	Total	**ATS**	**ITS**	**Total**	**ATS**	**ITS**	**Total**	**ATS**	**ITS**	**Total**	**ATS**	**ITS**
count of all results	**18**	5	13	**3**	0	3	**3**	0	3	**18**	5	13
correct acceptance	**15**	5	10	**0**	0	0	**-**	-	-	**15**	5	10
correct rejection	**3**	0	3	**3**	0	3	**3**	0	3	**-**	-	-
Total correct results	**18**	5	13	**3**	0	3	**3**	0	3	**15**	5	10
percentage correct results	***100***	*100*	*100*	***100***	*100*	*100*	***100***	*-*	*100*	***83***	*100*	*77*
incorrect acceptance [II]	**0**	0	0	**0**	0	0	**-**	-	-	**3**	0	3
incorrect rejection [I]	**0**	0	0	**0**	0	0	**0**	0	0	**-**	-	-
Total incorrect results	**0**	0	0	**0**	0	0	**0**	0	0	**3**	0	3
percentage incorrect results	***0***	*0*	*0*	***0***	*0*	*0*	***0***	*-*	*0*	***17***	*0*	*23*

NB: [I] Incorrect rejection equates to a type I error (i.e. incorrect rejection of a sample with a legitimate claim), [II] Incorrect acceptance equates to a type II error (i.e. incorrect acceptance of a sample with a falsified claim), ATS = Authentication Test Samples, ITS = Investigative Test Samples

Discussion

Assignment testing has been proposed as a suitable technique for verifying the claimed origin of timber samples (Dormontt et al. 2015). This study demonstrates its application and was used to assess the claimed origin of ayous (*Triplochiton scleroxylon*) timber (test) samples. This study examined claims of test samples with legitimate origins already incorporated in the reference dataset (the authentication test samples) as well as those from locations external to the data (the investigation test samples).Test samples were able to be assigned to a reference population, regardless of their type (investigation or authentication). For those that assigned to a different group than the one claimed we used an evidence-based approach to determine if this alteration was indicating a falsification to the claimed origin of a test sample. In the three instances where this could be achieved it was found to be correct. Additionally, this falsification was identified regardless of the grouping level of the samples.

The emphasis of this study was to detect false origin claims, an approach that is in line with the aims of government agencies that regulate timber trading. Falsified claims were detected through assessment of the likelihood ratio associated with test sample assignments that were inconsistent with the original claim. Assignment testing was most successfully applied when used to detect long-range falsification claims (i.e. where the claim and genuine origin were very distant). In the three instances where this occurred it was easily detectable, regardless of the grouping level used. Considering the movement of timber between many African countries, there is also the potential for these types of forgeries to occur in reality (EC 2008; ITTO 2010; Nellemann and INTERPOL 2012). However, it is probable that falsifications would also be occurring at a much more localised level (i.e. between neighbouring countries or concessions within a country) (Degen et al. 2013; Nellemann and INTERPOL 2012). Despite the majority of the investigative test samples having falsified concession claims, assessing the claims at this level was challenging, due to the low proportion of self-assigning reference samples. While concession scale tests may be warranted in support of timber-trading laws in the EU and Australia where concession of

origin declarations are required (where applicable) (Lowe et al. 2016), country of origin tests are likely to be the most useful, as all timber-trading laws require this scale of identification (Lowe et al. 2016).

There were limitations to assignment testing at both the geopolitical grouping levels (country and concession), primarily because they are artificial constructs (Ball et al. 2011; Ogden and Linacre 2015). Genetic group assignments on the other hand, were the most reliable as they accurately reflect the population genetic dynamics of the species. Nevertheless, while there may not be a geopolitical signal in the data, assignment tests can still be performed at these grouping levels, by using genetic populations as a proxy; meaning that the suitability of assignment testing as a technique is largely determined by the extent to which geo-genetic and geopolitical boundaries co-occur. Additionally, the greater the level of genetic structuring within a species, the more detailed and precise a geopolitical assignment can be undertaken (Lowe and Cross 2011).

For ayous there are only three genetic populations, spread across the range (Ch. 4), two of which encompass multiple countries (Central Region (CR): consisting of samples from Republic of the Congo (COG) and Cameroon (CMR), as well as samples from the North Western regions of Democratic Republic of the Congo (COD); and Western Region (WR) : all samples from Ghana (GHA) and Ivory Coast (CIV)) (see Fig 1). The impact of not having a discrete genetic signal for each country within those two groups was clearly evident in the reference samples that did not self-assign. The most noticeable example being for the country level assignments, where 31% (GHA) and 23% (CIV) of reference samples assigned to the other country of the WR genetic group. Additionally, the one test sample (a024) to incorrectly assign back to the claim country (CIV) was actually from GHA. Moreover, when the results of both the test and reference sample assignments were assessed for self-assignment to the appropriate genetic group then every test sample and almost all reference samples (99% (concession) and ~98% (country)) correctly assigned. The lack of certainty regarding the reference sample country grouping assignments meant that a supported assignment could be not determined for four test samples (i.e. the

likelihood ratio was lower than that of the reference group). Only one of which (a096) had a falsified country claim, with the other three test samples incorrectly assigning to other countries within the true genetic population.

The genetic structure of ayous, like many other tree species from the African tropical rainforests is limited to a small number of genetic clusters (Ch. 4; Degen and Sebbenn 2016). This is due to the influence of geneflow, the transfer of genetic material between individuals (from one generation to another) throughout the species range (Schwartz and Karl 2008). The level of geneflow for a species is driven predominantly by the biological and reproductive strategies (life history traits) employed to control geneflow to benefit the individual as well as the species (Lowe et al. 2014). While being a biological function to ensure genetic resilience to changing conditions (Austerlitz et al. 2000; Newton et al. 1999; Petit and Hampe 2006), it does reduce the spatial resolution that assignment tests reliably detect (Lowe and Cross 2011).

The expansion and contraction of the Guineo-Congolian rainforest cover in recent geological history have also influenced geneflow patterns of most plant and animal species in that region (Ch. 4; Hardy et al. 2013). The genetic dissimilarity of the NW and Central COD populations, despite being a continuous forest, is evidence of this (see Ch. 4) for more information on this topic).

For this study, three decision making strategies were utilised for assignment testing assessment (see Table 2 for summary). As expected, incorrect evaluations were still made when inappropriate mechanisms were used (i.e. Strategy A & Strategy B). Both of these strategies were framed around adjudicating on test samples with little or no evidence (either assuming the group that a test sample assigns to is the true origin (i.e. Strategy A) or defaulting to accept a claim if insufficient proof to reject (i.e. Strategy B)). By using an evidence-based approach, the likelihood ratios, no erroneous decisions were made when using the final strategy (i.e. Strategy C). However, it did mean that there was a reduction on the number of samples where supported decisions could be made (*n=3* for all grouping levels). Our results show that assignment testing is

appropriate for working within a legal framework, as long as an appropriate evidence-based framework for decision-making, such as Strategy C, is employed (Nielsen 2016; Ogden et al. 2009).

This study highlights the reliability of using an appropriate assessment strategy for test samples that assign to different groups than claimed (i.e. identification of false origin claims). We believe that a similar strategy to evaluate the probability of samples that assign back to claim groups could be employed (and this has also been looked at in other papers (see Ogden & Linacre (2015)). While this strategy was not undertaken in this study, due the limitations of the assignment program used (GENODIVE), we suggest that further work explores how these analysis functions can be incorporated into assignment testing programs for use in timber.

To increase the significance of assignment testing results, alternative methods may be useful. One approach incorporates the testing of multiple samples from a single shipment, with the density and patterns of the assignments being used as a multiplication factor to improve support (or lack thereof) for results (Degen et al. 2013; Tnah et al. 2009). Considering that in most instances there is only one origin declaration for the entire shipment, this approach has also been suggested as being more relevant for actual casework (Degen et al. 2013). Another approach is the use of data from multiple techniques to produce a better result (see (Dormontt et al. 2015; Ogden et al. 2009) for more information). For example, this study was undertaken in conjunction with other facilities/laboratories using alternative techniques (wood anatomy and stable isotopes), whilst the results of each method remained separate for comparison (Degen et al. 2015), there is potential to explore how results can be effectively combined in future work.

An alternative for attempting to boost support for assignments, is to repurpose the data and use in individualisation tests instead. Instead of assigning an unknown individual to the most likely population, these tests can assess the match between two or more specimens to determine if they are the same

individual. Individualisation tests (also known as "DNA profiling") have also been proposed as a more cost-effective approach to verifying claims, capable of working within the timber tracking legislation framework (Lowe et al. 2010; Ogden and Linacre 2015). Additionally, datasets used for individualisation testing are also more easily forensically validated, unlike for assignment testing; but this has not been completed for a plant species yet. While not necessarily a requirement for use in court cases (in Australia at least), it does afford greater confidence than un-validated methods (Ogden and Linacre 2015). Work to validate testing should be conducted in the future to forensic standards and guidelines (SWFS 2015; SWGDAM 2016; SWFS 2018).

Conclusion

Here we demonstrate that assignment testing is suitable for verifying the origin of timber samples when likelihood ratios incorporated into the decision-making process. While our approach resulted in a reduction in the number of test samples successfully assigned, it was effective at nullifying both miscarriages of justice (Type I error - incorrect rejection of a legitimate claim) and errors of impunity (Type II error - incorrect acceptance of a falsified claim). We conclude that assignment testing has a place as a legal tool for monitoring and controlling the trade of timber species when used within an appropriate decision-making framework.

Box 1: Limitation of results

There was a lack of sample site diversity to the test samples, with similar patterns (claim and actual groups) used repeatedly. Of the blinded test samples, most had legitimate country (62% (8/14)) or genetic group (77% (10/14)) claims. Yet the representation was restricted to two countries (four each GHA and CMR). Similarly, only three countries (CIV, GHA, CMR) were represented in the non-blinded test samples. In regard to falsified claims, of the possible alterations, only a limited number were used in the country (4/20) or genetic group (3/6) tests. The lack of diversity in the test samples may have restricted the understanding of the capabilities of assignment testing

in ayous. It would be beneficial to conduct further testing with additional samples. However, subsequent testing may be hampered by difficulty obtaining additional samples. To overcome this obstacle, the use of a simulated species is an option worth considering and has been used previously in assignment tests (e.g. (Degen et al. 2017)). This would create a potentially limitless number of test samples that would be available to further investigate improving assignment accuracies.

Acknowledgements

The authors would like to thank the present and past members of the Lowe Lab Group Timber Tracking Team. This work was undertaken with funding from the following: Australian Research Council (linkage), International Tropical Timber Organisation (ITTO), World Wildlife Fund (WWF). Appreciations also go to WWF, G2S and the Thünen Institute for Forest Genetics for the collection and organisation of the blind samples.

References

[EC] European Commission (2008) Cross-border Trade of Timber and Wood Products in West Africa. European Commission, Brussels, Belgium https://www.illegal-logging.info/sites/files/chlogging/uploads/CrossBorderTimberFlows.pdf

[ITTO] International Tropical Timber Organisation (2010) GOOD NEIGHBOURS: Promoting intra-African markets for timber and timber products. ITTO Technical Series #35

[SWFS] Society for Wildlife Forensic Sciences (2015) Standards and guidelines for forensic botany identification https://wwwwildlifeforensicscienceorg/wp-content/uploads/2016/07/annex-18-standards-and-guidelines-for-forensic-botany-identification-2pdf

[SWGDAM] Scientific Working Group for DNA Analysis Methods (2016) SWGDAM Validation Guidelines for DNA Analysis Methods https://docswixstaticcom/ugd/4344b0_813b241e8944497e99b9c45b163b76bdpdf

[SWFS] Society for Wildllife Forensic Sciences Technical Working Group (2018) Standards and Guidelines for Wildlife Forensic Analysis, Version 3. Ed. Lucy M.I. Webster. Published by the Society for Wildlife Forensic Science, 19th November 2018, pp.21. https://www.wildlifeforensicscience.org/wp-content/uploads/2018/11/SWFS-Standards-and-Guidelines_Version-3_19-11-18.pdf

African Regional Workshop (Conservation & Sustainable Management of Trees Z, July, 1996) (1998) *Triplochiton scleroxylon.* The IUCN Red List of Threatened Species 1998: e.T32174A9684953. http://dx.doi.org/10.2305/IUCN.UK.1998.RLTS.T32174A9684953.en. Accessed 16 January 2017

Austerlitz F, Mariette S, Machon N, Gouyon P-H, Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. Genetics 154:1309-1321

Ball MC, Finnegan LA, Nette T, Broders HG, Wilson PJ (2011) Wildlife forensics:"Supervised" assignment testing can complicate the association of suspect cases to source populations. Forensic Science International: Genetics 5:50-56 doi:10.1016/j.fsigen.2010.02.002

Degen B, Blanc-Jolivet C, Stierand K, Gillet E (2017) A nearest neighbour approach by genetic distance to the assignment of individual trees to geographic origin. Forensic Science International: Genetics 27:132-141 doi:10.1016/j.fsigen.2016.12.011

Degen B, Sebbenn AM (2016) Genetics and Tropical Forests. In: L P, Köhl M (eds) Tropical Forestry Handbook. Springer-Verlag, Berlin Heidelberg, pp 885-920. doi:10.1007/978-3-642-54601-3_75

Degen B, Bouda ZH-N, Blanc-Jolivet C (2015) Completion report of the ITTO project PD 620/11 Rev.1 (M): Development & implementation of a species identification and timber tracking system with DNA fingerprints and s isotope in Africa. ITTO

https://www.itto.int/files/itto_project_db_input/2996/Competition/PD-620-11-R1-M-Completion%20Report.pdf
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 Science International: Genetics 7:55-62 doi:10.1016/j.fsigen.2012.06.003
Dormontt EE et al. (2015) Forensic timber identification: It's time to integrate disciplines to combat illegal logging. Biological Conservation 191:790-798 doi:10.1111/1755-0998.12387
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. Applications in Plant Sciences 5:1600158 doi:10.3732/apps.1600158
Hall JB, Bada SO (1979) The distribution and ecology of Obeche (*Triplochiton scleroxylon*). Journal of Ecology:543-564
Hansen MM, Kenchington E, Nielsen EE (2001) Assigning individual fish to populations using microsatellite DNA markers. Fish and Fisheries 2:93-112 doi:10.1046/j.1467-2960.2001.00043.x
Hardy OJ et al. (2013) Comparative phylogeography of African rain forest trees: a review of genetic signatures of vegetation history in the Guineo-Congolian region. Comptes rendus Geoscience 345:284-296 doi:10.1016/j.crte.2013.05.001
Igboanugo ABI, Iversen P (2004) Conservation and use of *Triplochiton scleroxylon* in moist West Africa. Forest Genetic Resources 31:50-52
Jardine DI et al. (2016) Development of SNP markers for ayous (*Triplochiton scleroxylon* K. Schum) an economically important tree species from tropical West and Central Africa. Conservation Genetics Resources:1-11 doi:10.1007/s12686-016-0529-8
Jardine DI, Dormontt EE, Lowe AJ (in-prep) The role of refugia and geneflow in defining genetic clusters of a tropical African rainforest tree. (*NB: this is chapter 4*)
Jianbang G, Cerutti P, Masiero M, Pettenella D, Andrighetto N, Dawson T (2016) Quantifying Illegal Logging and Related Timber Trade. International Union of Forest Research Organizations (IUFRO), Vienna, Austria,
Johnson RN, Wilson-Wilde L, Linacre A (2014) Current and future directions of DNA in wildlife forensic science. Forensic Science International: Genetics 10:1-11 doi:10.1016/j.fsigen.2013.12.007
Lowe AJ et al. (2014) Standardized genetic diversity-life history correlates for improved genetic resource management of Neotropical trees. Diversity and Distributions doi:10.1111/ddi.12716
Lowe AJ, Cross HB (2011) The application of DNA methods to timber tracking and origin verification. IAWA Journal 32:251 doi:10.1163/22941932-90000055
Lowe AJ et al. (2016) Opportunities for Improved Transparency in the Timber Trade through Scientific Verification. BioScience 66:990-998 doi:10.1093/biosci/biw129
Lowe AJ, Jardine DI, Cross HB, Degen B, Schindler L, Höltken AM (2015) A method of extracting plant nucleic acids from lignified plant tissue. WO2015070279,

Lowe AJ, Wong KN, Tiong YS, Iyerh S, Chew FT (2010) A DNA Method to Verify the Integrity of Timber Supply Chains; Confirming the Legal Sourcing of Merbau Timber From Logging Concession to Sawmill. Silvae Genetica 59:263

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes 4:792-794 doi:10.1111/j.1471-8286.2004.00770.x

Nellemann C, INTERPOL (2012) Green carbon, black trade: illegal logging, tax fraud and laundering in the world's tropical forests. A Rapid Response Assessment. United Nations Environment Programme, GRID-Arendal,

Newton A, Allnutt T, Gillies A, Lowe A, Ennos R (1999) Molecular phylogeography, intraspecific variation and the conservation of tree species. Trends in Ecology & Evolution 14:140-145 doi:10.1016/S0169-5347(98)01555-9

Ng CH, Lee SL, Tnah LH, Ng KKS, Lee CT, Diway B, Khoo E (2017) Geographic origin and individual assignment of *Shorea platyclados* (Dipterocarpaceae) for forensic identification. PloS ONE 12:e0176158 doi:10.1371/ journal.pone.0176158

Ng KKS et al. (2016) Forensic timber identification: a case study of a CITES listed species, *Gonystylus bancanus* (Thymelaeaceae). Forensic Science International: Genetics 23:197-209 doi:10.1016/j.fsigen.2016.05.002

Nielsen EE (2016) Population or Point-of-Origin Identification. In: Naaum A, Hanner R (eds) Seafood Authenticity and Traceability: A DNA-based Pespective. pp 149-169. doi:10.1016/B978-0-12-801592-6.00008-5

Norman M, Saunders J (2017) Regulating the Trade in Illegal Timber: Asian Approaches Compared – State of Play June 2017. Forest Trends Report Series. Forest Policy, Trade, and Finance. Forest Trends, Washington DC, USA,

Ogden R, Dawnay N, McEwing R (2009) Wildlife DNA forensics—bridging the gap between conservation genetics and law enforcement. Endangered Species Research 9:179-195 doi:10.3354/esr00144

Ogden R, Linacre A (2015) Wildlife forensic science: A review of genetic geographic origin assignment. Forensic Science International: Genetics 18:152-159 doi:10.3354/esr00144

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. Annual Review of Ecology, Evolution, and Systematics:187-214 doi:10.1146/annurev.ecolsys.37.091305.110215

Schwartz TS, Karl SA (2008) Population genetic assignment of confiscated gopher tortoises. The Journal of Wildlife Management 72:254-259 doi:10.1007/s10592-005-9078-5

Tnah LH, Lee SL, Ng KK, Tani N, Bhassu S, Othman RY (2009) Geographical traceability of an important tropical timber (*Neobalanocarpus heimii*) inferred from chloroplast DNA. Forest Ecology and Management 258:1918-1923 doi:10.1016/j.foreco.2009.07.029

Supplementary information for Chapter 5

Supplementary Table 1: Reference sample population locations as well as their respective allocation to Country and Genetic groups.

Country[1]	Pop code	Pop #[2]	Latitude	Longitude	# Inds[3]	Genetic Cluster[4]
COD	DRC_22	1	1.9744	25.3559	5	E_CDRC
	DRC_02	2	0.7581	24.4950	23	
	DRC_23	3	1.8496	23.6129	23	
	DRC_25	4	0.5800	22.9747	14	
	DRC_16	5	3.0914	20.5514	10	CR
	DRC_34	6	3.2054	20.5345	22	
	DRC_11	7	3.1128	20.1704	16	
	DRC_30	8	3.2367	19.8107	35	
	DRC_17	9	3.1788	19.0064	18	
COG	CB_07	10	1.5833	16.5588	7	
	CB_03	11	1.9235	16.4336	17	
	CB_06	12	1.5574	16.2544	14	
CMR	C_05	13	2.2157	15.5014	25	
	C_06	14	3.3796	15.1404	15	
	C_04	15	2.5683	13.9354	23	
	C_07	16	3.9125	13.9184	35	
	C_15	17	4.9117	13.6277	12	
	C_14	18	4.6232	12.8555	15	
	C_08	19	3.8901	12.7639	26	
	C_03	20	2.7168	12.6631	20	
	C_09	21	3.6754	11.4165	16	
	C_13	22	4.8567	11.3316	14	
	C_02	23	2.6159	11.1966	22	
	C_12	24	4.9485	10.7376	15	
	C_01	25	2.3283	10.4901	18	
GHA	GH_04	26	5.9593	-1.0748	15	WR
	GH_05	27	6.9551	-1.3662	7	
	GH_03	28	6.0397	-1.5621	15	
	GH_08	29	7.5503	-2.0153	22	
	GH_06	30	7.2447	-2.2016	13	
	GH_07	31	6.6972	-2.6156	18	
	GH_02	32	5.8105	-2.7389	19	
CIV	CIV_06	33	6.4093	-3.9037	22	
	CIV_04	34	5.8758	-4.2830	19	
	CIV_07	35	6.0239	-4.3418	22	
	CIV_08	36	6.0916	-5.6813	17	
	CIV_03	37	5.6789	-5.7433	19	
	CIV_02	38	5.6595	-5.7684	25	
	CIV_09	39	6.4965	-6.5748	19	
	CIV_10	40	6.5168	-7.4761	17	
	CIV_12	41	7.1423	-7.5308	24	

NB: Table adapted from chapter, [1]Country codes (ISO designated): COD = Democratic Republic of the Congo, COG = Republic of the Congo, CMR = Cameroon, GHA = Ghana, CIV = Ivory Coast/Côte d'Ivorie; [2]population number is used in text; [3]refers to the number of individuals in the reference population; [4]Allocation to particular genetic cluster (see Figure 1 for approximate boundaries of genetic clusters

Supplementary Material 1: Modified innuPREP Plant DNA kit extraction protocol.

The following is a modification of the standard procedure for *Protocol 2: gDNA isolation from plant material using Lysis Solution OPT*.

The following is a modification of the standard procedure for *Protocol 2: gDNA isolation from plant material using Lysis Solution OPT*. Steps in regular font are retained from manufacturer. **Those in bold have been modified from manufacturers protocol**. Underlined parts are additions to the original protocol

Step 1 (tissue): 50-100 mg starting tissue (wood shavings).
Step 2 (homogenisation): homogenisation of tissue in bead Ruptor (Omni international) with ceramic beads and liquid nitrogen.
Step 3 (Lysis): add 800µl of lysis Solution OPT to the homogenized sample, mix well (vortex 5 secs). Incubate for 120 mins @ 65°C and if using a shaking heat block @ 500 rpm.
Step 3a (separate coarse woody debris): briefly centrifuge samples (30 secs @ 17500g) and transfer 400µl of the buffer to a prepared 1.5ml tube.
Step 4 (precipitation): add 100µl of Precipitation Buffer P, vortex for 5 secs. Incubate at Room Temperature (RT) for 5 mins. Centrifuge at maximum speed for 5 mins.
Step 5 (pre-filtration): add supernatant to the Prefilter in the Receiver Tube. Centrifuge for 1 min @ 10,000g.
Step 6: This step is left out.
Step 7 (Bind DNA): Retain Receiver Tube but remove (discard) Prefilter. Add 200µl of SBS to the filtrate (Receiver Tube) and mix via pipetting. Transfer mixture to a new Spin Filter/Receiver Tube. Centrifuge for 2 mins @ 10,000g.
Step 8 (washing): Discard Receiver Tube but retain Spin Filter (add to a new Receiver Tube). Add 650µl of MS to the Spin Filter, centrifuge for 1 min @ 10,000g. discard filtrate, and repeat step (Add 650µl of MS to the Spin Filter, centrifuge for 1 min @ 10,000g).
Step 9 (remove Ethanol): Discard Filtrate but retain both the Spin Filter and Receiver Tube. Centrifuge at maximum speed for 2 mins.
Step 10 (elution): Discard Receiver Tube but retain the Spin Filter (add to an Elution Tube). Add 20µl of Elution Buffer directly onto the filter, incubate @ RT for 3 mins. Centrifuge for 1 min @ 6000g. Repeat step (add 20µl of Elution Buffer directly onto the filter, incubate @ RT for 3 mins. Centrifuge for 1 min @ 6000g), discard Spin Filter.

Supplementary Table 2: assignment results for reference samples assigning to concession groups

Inferred Concession Group \ Current Concession Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	*# inds to assign to group*
1		5	4	3																																							12
2	2	4	6	1																																							13
3	3	8	13	4																																							28
4		4		6																																							10
5							1	3																																			4
6						10	4	4	1	1	2	3	1									1																					27
7					2		3	4					1																														10
8		1			3	3	5	12	6		2	2	1				1					1																					37
9		1			2	4	1	4	4	1		2	2												2																		23
10								1	1																																		2
11						1			1	1	8	1	2		1		1																										16
12					1	2	1	2	2	1		3	1	1																													14
13					2	1		2	1	2	2	3	5	2	1	1	1			1	1		2		1																		28
14								2					1	1	2	4			3	1			2		1																		17
15							1						2		2	6	1	2	3	1	2		1		1				1							1							24
16													1	3	5	5	2	1	4	3	1	1	2	4																			32
17						1					1		1	1	1	1						1			2																		9
18															2	2	1	1	3	3		2	1	1						1												17
19											1		1		1	6			4	1	5	2		1	1																		23
20													2		1		2	2	1	3	2	3	1	2	1							1					1		1				23
21								1						2	1	2		1	2	2	1	1	1	1	2			1								1							19
22															1		1	3	2	2			2	1							1											13
23									1				3	3	1	4	1			2	1	1	6		5																		28
24														1		1	1	2	2			1	1	3				1														13
25									1		1		1		2	1		1			3		1	1	1																		13
26																										1	1	2	2	3	2	1		2	1			1		1			17
27																															1			1	1								3
28																									3		2			1	3			1	1	1		1	1		2		16
29																				1			1			1	2	2	3	3		2		1		2		1		1			20
30																									2				2	2					2			1					10
31																									1				1	2	3	1	1	3	4				3				23
32																	1	1							3	2	2	4		1	1	2		3	1		1	2	3				27
33														1														1				2	10			1		2		2			19
34															1	1									1	1		2			1	1	5			4	1	2		1			21
35																1												1			1	2			2	2	1	6	2	1	2		23
36																									1		1	1		2				1	1	1	2	2			1		13
37										1																			1			1	1	1	2	2	4					13
38																			1				1			2	1	1	1			2	1	1	2	2	1	3	2		5		26
39															1													3	2	1	4	1	1	3	1		2	1	2	6	1		29
40																													1		1		2	3		1	2	1	1	3			15
41																	1						1						1	1	1	1	2		1	2	4		1	1	6		23

# inds in group	5	23	23	14	10	22	16	35	18	7	17	14	25	15	23	35	12	15	26	20	16	14	22	15	18	15	7	15	22	13	18	19	22	19	22	17	19	25	19	17	24	*753*
# self-assigning inds	0	4	13	6	0	10	3	12	4	0	8	3	5	1	2	5	0	1	4	3	1	0	6	3	1	0	0	2	3	2	3	1	10	5	2	1	2	3	2	3	6	*140*
% group to self assign	**0**	**17**	**57**	**43**	**0**	**45**	**19**	**34**	**22**	**0**	**47**	**21**	**20**	**7**	**9**	**14**	**0**	**7**	**15**	**15**	**6**	**0**	**27**	**20**	**6**	**0**	**0**	**13**	**14**	**15**	**17**	**5**	**45**	**26**	**9**	**6**	**11**	**12**	**11**	**18**	**25**	**19**
																																										overall

Supplementary Table 3: Country reference sample assignment results

		Current Country Group					
		COD	COG	CMR	GHA	CIV	# inds assigning to group
Inferred Contry Group	COD	118	9	4			131
	COG	41	25	26			92
	CMR	7	4	217	4	3	235
	GHA			5	71	43	119
	CIV			4	34	138	176
# inds in group		166	38	256	109	184	*753*
# self assigning inds		118	25	217	71	138	*569*
% group to self assign		**71**	**66**	**85**	**65**	**75**	***76***
							overall

Supplementary Table 4: Genetic groups reference sample assignment results

		Current Genetic Group			
		E_CDRC	CR	WR	# inds assigning to group
Inferred Genetic Group	E_CDRC	63			63
	CR	2	381	5	388
	WR		14	288	302
# inds in group		65	395	293	*753*
# self assigning inds		63	381	288	*732*
% group to self assign		**97**	**96**	**98**	***97***
					Overall

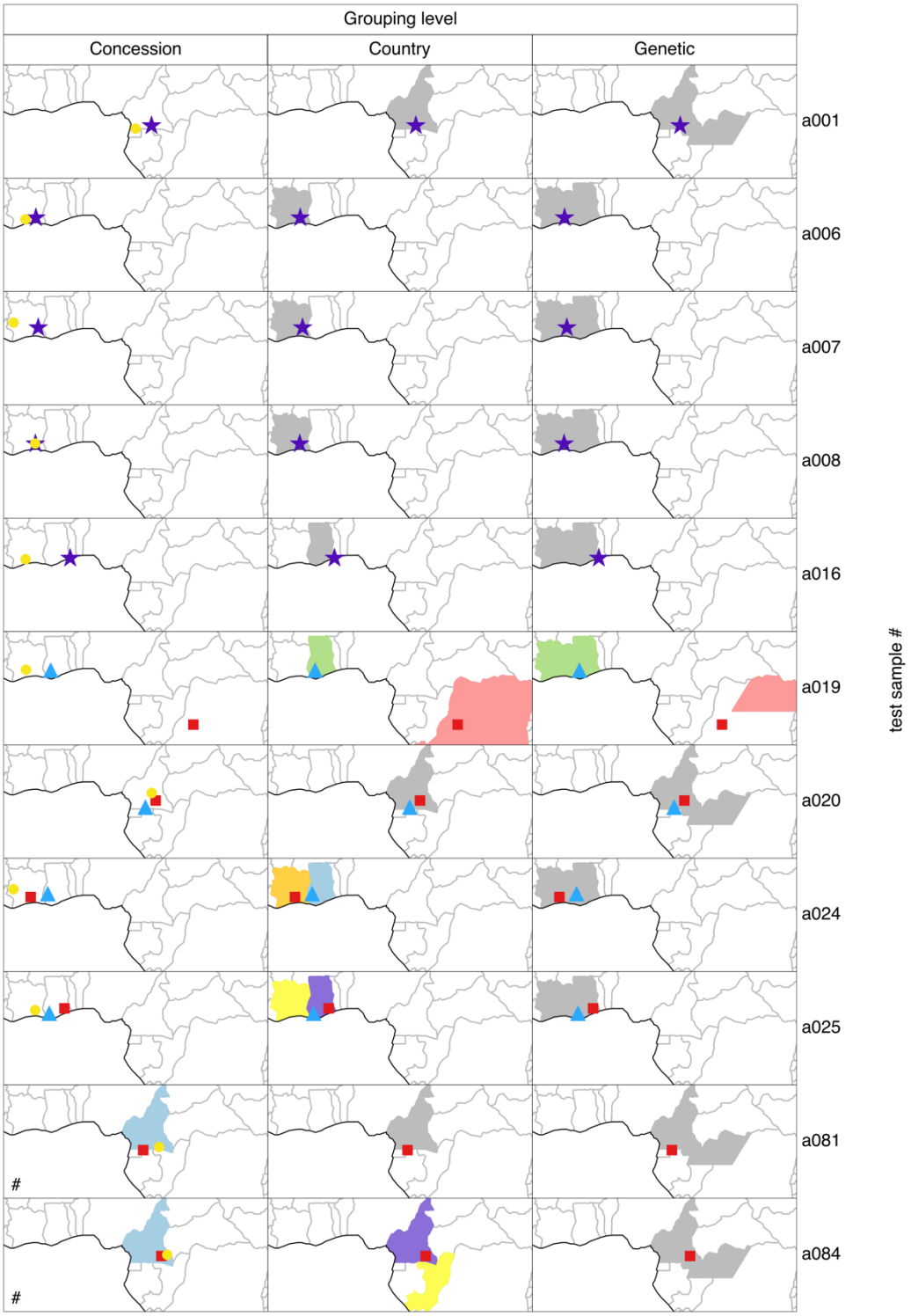
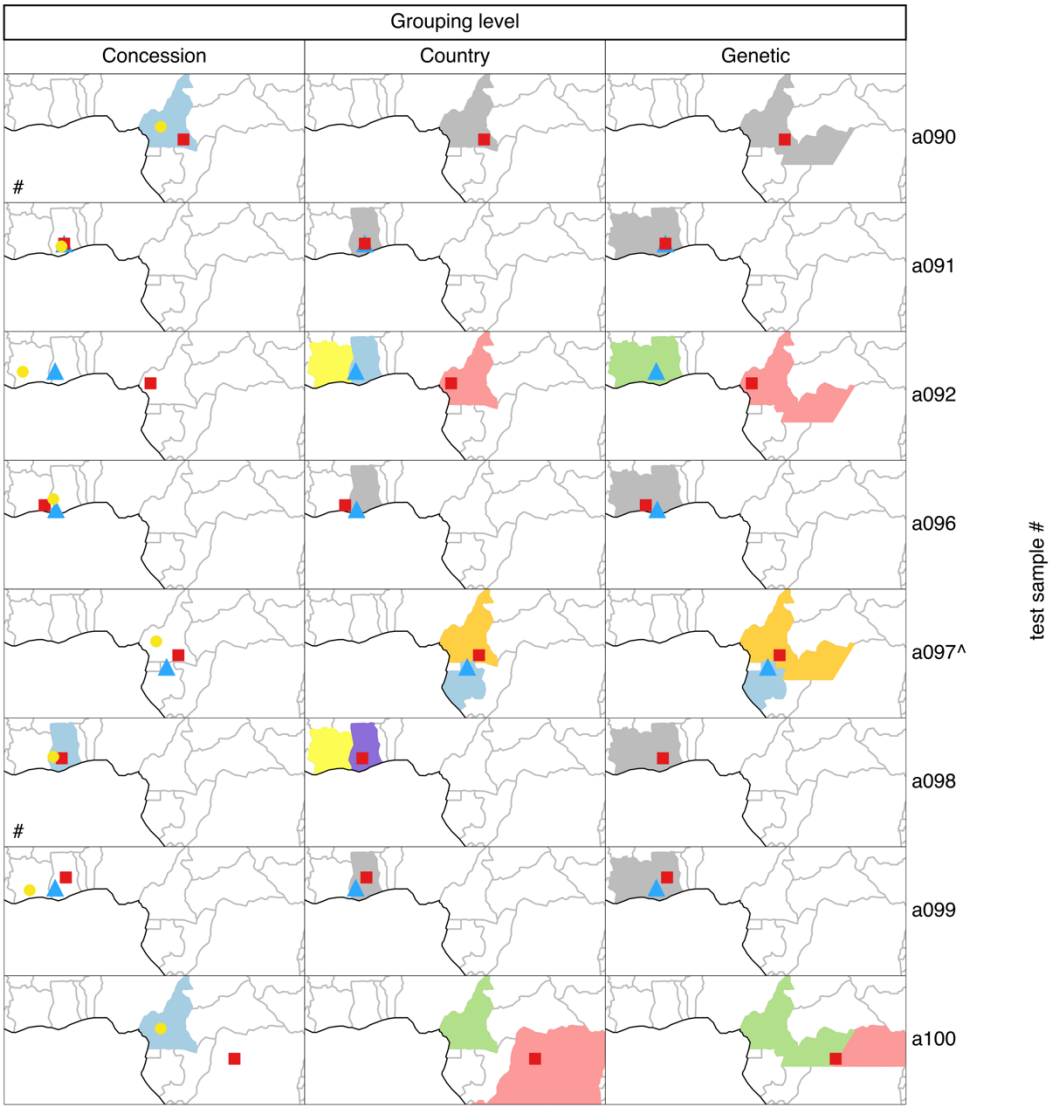


Figure S1 continues on next page

Continuation of Figure S1 from previous page



Supplementary Figure 1: Assignment results of test samples.
NB: Coordinates of test samples claimed (■), actual (▲) (or ★ if same location) origin, inferred reference concession group assignment (●). # = test samples not assessed for concession claims (only true country known (concession unknown) claim and inferred concession groups within same country), a097^ = test sample not analysed (true origin is Gabon, not encompased in reference sample extent), country/genetic group shading for: claim group ( ), inferred group ( ), actual (this includes concession tests where only country is known) ( ) if different, alternatively country/genetic group shading for: claim & inferred group ( ), inferred & actual group ( ), claim & actual group ( ), claim & inferred & actual groups ( ) if the same.

Chapter 6: General Thesis Discussion

The principal motivation of research presented in this thesis is to demonstrate how genetic methods, such as assignment testing to verify the origin of timber harvested from natural forest populations, can be used for forensic identification as part of timber trading and protection legislation (i.e. *Ch. 5*). For this outcome to occur, there are other research components that also needed to be conducted, including; the successful extraction of DNA from timber material (*Ch. 2*), the development of the genetic markers that can be used for timber identification (*Ch. 3*), and understanding the natural pattern(s) of genetic diversity distribution that is the basis of successful genetic assignment capacity (*Ch. 4*). Framed around the various data chapters, this thesis consists of research undertaken to facilitate the genetic identification of timber samples.

Rather than repeat the significant findings of each chapter individually, this section discusses the thesis' research more broadly. The results have been framed in the context of two themes, technical and applied research. This discussion will also explain how the findings relate to understanding and protecting timber species. Along with the data chapters, three other bodies of work (included in the appendices) will be referred to in this discussion. These additional documents are: Appendix V, the patent for the BOTAB DNA extraction protocol (hereafter the *patent*); Appendix IV, a brief conference paper on marker development and genetics analysis for ayous (*Triplochiton scleroxylon* K. Schum), presented at the 2015 World Forestry Congress in Durban, South Africa (hereafter the *WFC paper*); and Appendix II, a manuscript for the forensic identification of bigleaf Maple (*Acer macrophyllum* Pursch) timber samples (hereafter the *maple paper*).

Technical Research

During my PhD a range of techniques and procedures were developed, optimised and utilised to allow the applied research to occur. This section incorporates two areas; the optimisation and comparison of the BOTAB DNA extraction protocol

(hereafter the BOTAB protocol) (*Ch. 2*) as part of the finalisation of the patent document (*patent*), and the development of genetic markers for both species (*Ch. 3*) as well as demonstration of the suitability of the markers in timber material in ayous (*Ch. 5*) and bigleaf maple (*maple paper*).

The main focus of the BOTAB protocol was the use of a proprietary DNA extraction buffer, the BOTAB buffer, for extracting DNA from degraded and lignified plant tissue. The DNA extraction study (*Ch. 2*) found that the current reagents are still the most suitable. Yet it was deduced that concentrations of some of the current reagents are flexible, as the alternative concentration of some reagents produced a similar outcome to the standard. This means the skillset of laboratory technicians and the accuracy of the measuring equipment (e.g. if the method is to be automated or miniaturised) may not be crucial to a successful DNA extraction. Furthermore, considering that the patent does not restrict the use by academic or non-commercial research, the BOTAB protocol could be promoted for use more widely, especially to facilities in timber exporting countries, where capacity and technical knowhow may be limited.

One of the biggest shortcomings of the BOTAB protocol is the length of the procedure, two days. Unfortunately no suitable mechanisms to streamline or shorten the protocol could be identified from my research. This issue is compounded by the fact that with minor modifications, the commercial plant DNA extraction kits (*DNeasy Plant Mini Kit* (QIAGEN; Hilden, Germany), *innuPREP Plant DNA Kit* (ANALYTIK JENA; Jena, Germany), *PowerPlant Pro DNA Isolation Kit* (MO BIO; Carlsbad, CA, USA)) were also able to extract DNA from timber material. The shorter timeframe, alongside the lack of a hazardous chloroform phase separation step, means these kits are more suitable for a wider use than the BOTAB protocol, and this has been noted in other studies (Jiao et al. 2012; Särkinen et al. 2012).

The other technical aspects of this thesis were; the development of genetic markers for both bigleaf maple and ayous (*Ch. 3*), and the suitability of these markers for genotyping DNA from timber (*Ch. 5, maple paper*). Marker

development protocols designed to utilise High Throughput Sequencing Techniques (HTS) such as those used for bigleaf maple (*Ch. 3$_{I}$*) and ayous (*Ch. 3$_{II}$, WFC paper*) have facilitated the use of genetic markers such as Single Nucleotide Polymorphism (SNP) and Insertion/Deletion (INDEL) loci. These methods are potentially very useful for application to DNA extracted from timber and the simple variable patterns that they highlight can be easily identified in the genotyping analysis (MassARRAY® iPLEX™ platform (AGENA BIOSCIENCES™, San Diego, CA, USA)). Furthermore, the amplicon lengths are all short (~100 base pairs) and are thus more likely to work with degraded materials and DNA template, such as old timber.

The marker development protocol has shown to be highly appropriate. Regardless of the location or number of samples used in the development ((31 bigleaf maple, 48 ayous), the loci identified could amplify in samples from across the species' range (e.g. the genotyping of a timber sample from Gabon (*Ch. 5*) (but also see *Ch. 3*). The genetic markers were also highly species-specific, failing to amplify in unrelated species and only partially in sister species (76% in *Acer circinatum*) yet not enough to be considered usable (*maple paper*).

This work also identified that the genotyping is repeatable (e.g. 100% concordance between 87 paired bigleaf maple samples (*maple paper*)) and reliable in timber (e.g. 19/25 ayous timber samples with usable genotypes (≥ 95% loci amplified) (*Ch. 5*)) and low concentration samples (e.g. successful and correct genotyping of samples with a DNA concentration of 0.625ng/µL (*maple paper*)).

This thesis demonstrates that some of the technical restrictions that may have been limiting the uptake of genetic methods as a tool for identification of timber in the past can be overcome. The technical research conducted for this thesis has identified that with appropriate modifications, commercial kits can be used successfully to extract DNA from timber, and if not, the BOTAB protocol can be used. Additionally, SNP and INDEL markers can be used to genotype timber samples with a high success rate.

Applied Research

There are two main outcomes that arose from the thesis' applied research; an understanding of the structure of genetic variation in natural populations (*WFC paper, Ch. 4*), and demonstration of the utilisation of genetic methods for forensic identification, which could be incorporated into timber trade and protection legislation (*Ch. 5, maple paper*).

Both applications utilised the same set of genetic markers (developed in *Ch. 3*). Furthermore, for ayous, the genotyping results for samples used to examine genetic structure (*Ch. 4*) were suitable as reference samples for genetic assignment tests of blind timber samples (*Ch. 5*). Considering the monetary cost of generating the dataset, the capacity for it to be used in multiple ways increases its utility, making the approach more cost effective.

Genetic diversity of ayous (*Ch. 4*) was found to be high (total heterozygosity = 0.272), which is indicative of long lived, wind dispersed tree species (Dick et al. 2008). High genetic diversity is a mechanism that provides a species with genetic resilience and capacity to adapt to changing circumstances over time (Degen and Sebbenn 2016; Hamrick et al. 1992). However, the downside of this diversity is that there is low population differentiation (G_{ST}=0.058). The subsequent clustering analyses identified that between 2-5 genetic populations exist for ayous, with three clusters being the easiest to define (in *Ch. 4*), as was subsequently used for assignment tests (*Ch. 5*). There was a separation of two of the clusters across the Dahomey Gap, a 200 km wide break in the forest vegetation, with individuals from the Ivory Coast and Ghana forming one cluster in the west and samples from the forested areas of Cameroon, Republic of the Congo and north-western Democratic Republic of the Congo forming the other. This clustering pattern concurs with previous findings of other African tropical rainforest tree species (Hardy et al. 2013), and highlights the influence that the Dahomey gap has had on genetic structuring of tree species in the region. The third cluster identified was an isolated grouping of samples from central Democratic Republic of the Congo. Genetic differentiation between north-western and central areas of Democratic Republic of the Congo has not been

identified previously. In the Discussion of that chapter (*Ch. 4*), it was hypothesised that the reason for this genetic distinction, was due to the two locations being from different refugial groups. Looking more broadly, the clustering results across the species range are a combination of life history strategies used by ayous and the impact of changing climatic conditions in recent geological times.

When the genetic markers were used for timber identification it was to demonstrate their applicability in a legal setting and to support timber trade and protection legislation. For ayous, despite the project not being used in casework, the results obtained demonstrate that there is a place for genetic assignment testing to verify the claimed origins of samples (*Ch. 5*). The assignment testing study concluded that the claimed country of origin could be verified, as long as an appropriate assessment strategy is used. Despite having limited support for assignment to country level, due to a lack of geo-genetic/geopolitical alignment, the success when using this assessment measure indicates that reliable results can still be obtained.

For bigleaf maple, the genetic markers were applied in an actual legal case[17] and the results were used to support an illegal logging prosecution. The markers were successfully used for genetic individualisation testing (i.e. DNA profiling) (*maple paper*) to match confiscated timber from a lumberyard with a tree stump that was illegally felled. Both the ayous and bigleaf maple applications of our developed markers demonstrate the capacity of genetic identification. Genetic techniques alongside the other scientific techniques are key in enforcing, controlling and ultimately stopping illegal logging activities and can also be used to promote the sustainable trade of timber (Sasaki et al. 2016; Vlam et al. 2018).

It is worth mentioning the DNA extraction work again within the context of applied research. The DNA extraction comparison test (*Ch. 2*) identified that the innuPREP kit was the most successful kit protocol; and was used to extract

[17]www.justice.gov/usao-wdwa/pr/tree-thieves-and-mill-owner-indicted-theft-big-leaf-maples-national-forest

usable DNA from ayous blind timber samples alongside the BOTAB protocol (*Ch. 5*). The BOTAB protocol was also used successfully for DNA extraction from bigleaf maple evidence samples (*maple paper*). These examples demonstrate the capacity of both protocols and will hopefully facilitate the use of them more frequently.

Limitations identified during candidature

No thesis would be complete without the inclusion of a section discussing the limitations that were identified throughout the candidature and this thesis is no different. While there are specific problems that have been identified in each of the data chapters, they will not be re-addressed. Instead, this section aims to address broader issues and limitations that had a more fundamental impact on the work.

By far the biggest challenges encountered were in regard to issues associated with the marker development protocol, specifically the University of Adelaide approach (see *Ch. 3* for more information). Ayous and bigleaf maple were some of the first species in the laboratory to have markers developed using this protocol. Consequently, there was significant amounts of on-the-spot learning and troubleshooting required. Many of these issues were ironed out for later species (see published examples, (Cross et al. 2016; Hernawan et al. 2017)). The protocol was also limited because of the sequencing platform used, with sequencing performed using the Ion Torrent™ Personal Genome Machine™ (PGM) (THERMO FISHER SCIENTIFIC, Waltham, MA, USA). The PGM requires a high degree of technical proficiency and problems can easily arise because of it. Firstly, the chemistry used to read genetic sequences has been found to have a higher error rate than alternative platforms and reduce the quality of sequences produced (Quail et. al 2012). The impact of this means that a high proportion of sequences might have been removed due to low quality issues. Additionally, at the time the marker development for ayous and bigleaf maple were undertaken, the PGM had a limited data capacity, so the sequencing output was small in comparison to what can be generated on more recent platforms.

Because of the limiting factors and complexities associate with the PGM, sequencing is now performed using the Illumina® (San Diego, CA, USA) platform instead. The Illumina system is considered to be a superior system compared to the PGM, because it has a lower sequencing error rate and larger output, which means a greater number of high-quality sequences can be generated. Additionally, the marker development protocol has been updated (see Peterson et. al (2012) for overview). It is safe to assume that more polymorphic loci would be obtained if the marker development for ayous and bigleaf maple were repeated using the contemporary marker development protocol. In an ideal world, the issues that were affecting the clarity of results (assignment testing, spatial genetic structure) or preventing their use (phylogeography) would be overcome. But in all likelihood, these limitations, especially assignment resolution, are more likely to be a factor of the species and its inherent genetic structuring rather than too few loci.

Separate from marker development protocol issues, there are limitations associated with the lack of legal exposure of genetic timber identification. The results sections from both the ayous assignment testing (*Ch. 5*) and the bigleaf maple individualisation testing (*maple paper*) were robust and of sufficient quality to be court worthy. However, the legal support and robustness of the approaches used remain unknown. Notwithstanding the bigleaf maple data being submitted and accepted by the prosecution as evidence, it was not specifically challenged by defence or supported by prosecution lawyers (all defendants pleaded guilty), meaning that it may be possible that some aspect of the work could be legally challenged.

Despite the existence of DNA tests that could control or identify illegal logging, it is clear that these tests are still slow to be applied by the authorities and the timber industry. With more exposure and awareness raising to law enforcement, government authorities and forestry and timber sector stakeholders it is likely that these services will be increasingly taken up. The future publication of the bigleaf maple paper represents an excellent opportunity to achieve this. However, notwithstanding several recent publications on the use of genetic

techniques for timber identification (Blanc-Jolivet et al. 2017; Degen et al. 2017; Nowakowska et al. 2015; Tereba et al. 2017; Vlam et al. 2018), the application of these examples into a legal setting is yet to occur and the technology still remains legally unchallenged (Dormontt et al. 2015).

Future directions

Whilst the data chapters have highlighted some specific avenues of future work, this section takes a more philosophical and integrated approach. Continuing with the two themes (*Technical and Applied*) this section presents some ideas of where genetic methods can go with regard to understanding forests and how to utilise the technology to assist timber trading and forest protection legislation.

Technical Research

This thesis has shown that the extraction of DNA from timber is becoming more routine and can be undertaken using either the proprietary BOTAB extraction protocol (*the patent*), other in-house procedures, or by modifying commercial DNA extraction kits. Additionally, the use of short amplicons with relatively simple variants (such as SNPs and INDELs) has allowed the genotyping of timber samples. Nevertheless, regardless of the extraction protocol, or genetic marker used, there are still constraints that restrict genetic timber identification being employed more widely. For any of the procedures used in genetic timber identification there are three areas that would benefit from being streamlined.

Firstly, to facilitate their inclusion in frontline identification applications, DNA extraction and amplification/genotyping technologies would benefit from improvements to shorten the times. Presently the extraction protocols that are suitable for timber samples take 1-2 days to complete, and results from the amplification/genotyping of new samples takes 1-2 weeks. This all translates into a long turnaround time for answers to be provided to authorities or industry.

Secondly, the processes need to be simplified. Presently there is a significant reliance on equipment/machinery (such as centrifuges, vortexes and incubators)

in the extraction protocols, which increases the infrastructure requirements of a facility. Additionally the number and complexity of steps within each protocol means a certain level of technical proficiency is required. It also increases the chance of mistakes being made (which could result in negative legal implications).

The scale/size of each protocol and the time considerations also influence the third area for improvement, cost. While not necessarily the most important factor, cost considerations might still be inhibiting the use of genetic identification of timber samples.

Robotics and automated equipment have the capacity to go some way towards making DNA extractions and amplification/genotyping protocols easier, cheaper and faster to conduct. Many commercial extraction kit suppliers have already developed mechanised equipment that is capable of doing kit extractions (see footnotes[18,19,20] for examples). However, extractions using this equipment take a similar amount of time to complete as manual extractions. Additionally, the cost of purchasing and maintaining machinery comes with financial constraints.

The best way making DNA extractions and amplification/genotyping protocols easier, cheaper and faster is the development of a set of small handheld or portable devices that are cheap and easy to run with fast turnaround times (Masters *et. al.* (2019)). Ideally, it would be a single unit that can extract DNA and genotype samples. That level of simplicity would rapidly increase the usability of genetic identification techniques. These portable systems have

[18]www.qiagen.com/ca/shop/automated-solutions/sample-preparation/qiacube/#productdetails
[19]www.analytik-jena.de/en/life-science/products/prod/cat/automated-extraction/prod/innupure-c16-touch.html
[20]www.hamiltoncompany.com/products/automated-liquid-handling/standard-solutions/genomic-starlet

already been prototyped and are commercially available for human identification[21,22].

Instead of attempting to linkup DNA extraction and amplification/genotyping there is another approach that has already been utilised successfully for human identification. Known as direct PCR, it is an approach that eliminates the extraction procedure by amplifying/genotyping the tissue directly. This approach has been successfully used to generate a genetic profile of individuals by amplifying the cells left behind on a fingerprint (Haines and Linacre 2017; Templeton et al. 2015; Templeton et al. 2017).

To allow these sorts of technologies to be used in genetic timber identification, several breakthroughs need to be achieved. Firstly, unlike for humans, the tough cell walls of plants (especially timber) may require additional steps to break open before any extraction can be done. This issue also inhibits the capacity to perform direct PCR's. Additionally, without an extraction stage, chemical inhibitors would not be removed and may impact the ability to amplify the DNA. Finally, these devices can only perform a small number of genotyping tests for one species at present. To be truly useful, the genotyping capacity of the equipment needs to be able to work on a wide range of timber species and be able to infer the species name and where it comes from.

If the realisation of handheld devices is not achievable for timber species, then a complementary approach to solving timber crime could be to utilise human DNA. For example, the recovery of genetically identified human fingerprints from wildlife species (Mcleish et al. 2018). The genetic identification and profiling of a human fingerprint could theoretically be used to link illegal loggers with certain timbers. However, there are further questions that would need answering before this technique could be deemed suitable for timber samples. Hard, smooth surfaces are relatively easy to obtain a fingerprint from, but the rough surface of

[21]www.lgcgroup.com/products/paradna-technology/#.Wt6cXC97HjB
[22]https://techcrunch.com/2016/03/30/citizen-scientists-you-can-now-diy-your-own-dna-analysis/

a timber sample may not be suitable. Additionally, in the examples given, DNA profiles were obtained from samples up to 10 days old (Mcleish et al. 2018), yet many timber shipments would be in transit for much longer than that. Ultimately, human DNA could be used to link suspects to objects, such as timber, but identification of the timber itself will still be required in order to demonstrate that it is, in fact, illegal.

Another method that has potential is microbiome analysis. Because of a range of factors (including species, geographical location, and environmental conditions) microbial signatures are unique, and can be exploited for species, origin or individualisation testing. With less tough cell walls, microbes are more easily able to be utilised in portable systems. However, contamination issues need to be considered with microbe analysis. This would make it more unapproachable to less trained labs may have implications for any portable application. Additionally, to date there is insufficient evidence to verify that the signals obtained are repeatable and unique enough to meet forensic standards.

Regardless of how future studies can increase the speed from which identification can be obtained; they will all ultimately be limited by the amount of reference material and DNA sequence data that is available (Ahlers et. al 2017). Future work should seek to generate reference data for all traded timber species, regardless of their value or level of endangerment (Dormontt et al. 2015). As such, global initiatives that focus on generating as much reference data should be prioritised (Ahlers et. al 2017; Dormontt et al. 2015). The raw genetic sequences and samples generated from studies such as the ones in this thesis can be reanalysed at a later date as and when improved genetic techniques become feasible for broad scale application.

Applied Research

The research conducted as part of this thesis was undertaken with the assumption that the species studied will be eventually be at risk of population decline. The three ways in which scarcity is most likely to occur for these species is via: demand (legal or not) for timber products, deforestation (other than from

timber trading), and impacts of climate change. Considering the economic importance of both ayous and big leaf maple, future research should focus on ways of controlling impacts whilst also maintaining sustainable trade. It is important to understand how to maintain the trade of each species while protecting areas at risk and also allow for the movement of individuals of the species as climatic conditions change (McCallum et al. 2014; Olorode 2004). Further knowledge is needed to identify the resilience of specific populations to buffer the loss of individuals (through logging or dieback) and identify if that population should be protected. Also, information on which individuals would be best to relocate, are suitable for vegetative propagation or beneficial to maintain genetic diversity (i.e. for restoration) would be valuable, benefiting both *in-situ* the conservation and *ex-situ* revegetation of the species (Uyoh et al. 2003).

Concluding remarks

Irrespective of what the future might hold for the protection and management of timber species globally, this thesis highlights that genetic methods can be a key part of measures used to verify the identification of timber products. We successfully developed genetic markers for bigleaf maple and ayous that can be used in timber samples; additionally, these markers can be employed in a forensic context to identify the geographical region of origin or to match samples together. This thesis clearly demonstrates a successfully outcome in using genetic techniques for timber harvested tree species. Work on other species, especially those presently facing scarcity issues, will further enhance the use of genetic techniques to identify timber products.

I am proud of the work undertaken for this PhD and believe it to be a robust thesis.

Thank you for your consideration.

Chapter References

Ahlers N, Creecy J, Frankham G, Johnson RN, Kotze A, Linacre A, McEwing R, Mwale M, Rovie-Ryan JJ, Sitam F, Webster LMI (2017) 'ForCyt'DNA database of wildlife species. Forensic Science International: Genetics Supplement Series 6:e466-e468.

Blanc-Jolivet C, Kersten B, Daïnou K, Hardy OJ, Guichoux E, Delcamp A, Degen B (2017) Development of nuclear SNP markers for genetic tracking of Iroko, *Milicia excelsa* and *Milicia regia*. Conservation Genetics Resources:1-3 doi:10.1007/s12686-017-0716-2

Cross H, Biffin E, Lowe A, Waycott M (2016) Effective application of next-generation sequencing (NGS) approaches in systematics and population genetics: case studies in Eucalyptus and Acacia. Australian Systematic Botany 29:235-246 doi:10.1071/SB16019

Degen B, Blanc-Jolivet C, Stierand K, Gillet E (2017) A nearest neighbour approach by genetic distance to the assignment of individual trees to geographic origin. Forensic Science International: Genetics 27:132-141 doi:10.1016/j.fsigen.2016.12.011

Degen B, Sebbenn AM (2016) Genetics and Tropical Forests. In: L P, Köhl M (eds) Tropical Forestry Handbook. Springer-Verlag, Berlin Heidelberg, pp 885-920. doi:10.1007/978-3-642-54601-3_75

Dick CW, Hardy OJ, Jones FA, Petit RJ (2008) Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. Tropical Plant Biology 1:20-33 doi:10.1007/s12042-007-9006-6

Dormontt EE et al. (2015) Forensic timber identification: It's time to integrate disciplines to combat illegal logging. Biological Conservation 191:790-798 doi:10.1111/1755-0998.12387

Haines AM, Linacre A (2017) Detection of latent DNA on tape-lifts using fluorescent in situ detection. Australian Journal of Forensic Sciences:1-11 doi:10.1080/00450618.2017.1416174

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. Population Genetics of Forest Trees:95-124 doi:10.1007/978-94-011-2815-5_7

Hardy OJ et al. (2013) Comparative phylogeography of African rain forest trees: a review of genetic signatures of vegetation history in the Guineo-Congolian region. Comptes rendus Geoscience 345:284-296 doi:10.1016/j.crte.2013.05.001

Hernawan UE, Dijk Kj, Kendrick GA, Feng M, Biffin E, Lavery PS, McMahon K (2017) Historical processes and contemporary ocean currents drive genetic structure in the seagrass Thalassia hemprichii in the Indo-Australian Archipelago. Molecular ecology 26:1008-1021 doi:10.1111/mec.13966

Jiao L, Yin Y, Xiao F, Sun Q, Song K, Jiang X (2012) Comparative analysis of two DNA extraction protocols from fresh and dried wood of *Cunninghamia lanceolata* (*Taxodiaceae*). IAWA Journal 33:441-456 doi:10.1163/22941932-90000106

Masters A, Ogden R, Wetton JH, Dawnay N (2019) Defining end user requirements for a field-based molecular detection system for wildlife forensic investigations. Forensic science international, 301:231-239
McCallum KP, Guerin GR, Breed MF, Lowe AJ (2014) Combining population genetics, species distribution modelling and field assessments to understand a species vulnerability to climate change. Austral Ecology 39:17-28 doi:10.1111/aec.12041
Mcleish K, Ferguson S, Gannicliffe C, Campbell S, Thomson P, Webster L (2018) Profiling in wildlife crime: recovery of human DNA deposited outside. Forensic Science International: Genetics doi:10.1016/j.fsigen.2018.04.002
Nowakowska JA, Oszako T, Tereba A, Konecka A (2015) Forest tree species traced with a DNA-based proof for illegal logging case in Poland. In: Evolutionary biology: biodiversification from genotype to phenotype. Springer, pp 373-388. doi:10.1007/978-3-319-19932-0_19
Olorode O (2004) Invited Review-CONSERVATION OF PLANT GENETIC RESOURCES. African Journal of Traditional, Complementary and Alternative Medicines 1:4-14
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS ONE 7:e37135 doi:10.1371/journal.pone.0037135
Quail MA et al. (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13:341 doi:10.1186/1471-2164-13-341
Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT (2012) How to open the treasure chest? Optimising DNA extraction from herbarium specimens. PloS ONE 7:e43808 doi:10.1371/journal.pone.0043808
Sasaki N et al. (2016) Sustainable management of tropical forests can reduce carbon emissions and stabilize timber production. Frontiers in Environmental Science 4:50 doi:10.3389/fenvs.2016.00050
Templeton JEL, Taylor D, Handt O, Linacre A (2015) DNA profiles from fingermarks: A mock case study. Forensic Science International: Genetics Supplement Series 5:e154-e155 doi:10.1016/j.fsigss.2015.09.062
Templeton JEL, Taylor D, Handt O, Linacre A (2017) Typing DNA profiles from previously enhanced fingerprints using direct PCR. Forensic Science International: Genetics 29:276-282 doi:10.1016/j.fsigen.2017.05.006
Tereba A, Woodward S, Konecka A, Borys M, Nowakowska JA (2017) Analysis of DNA profiles of ash (Fraxinus excelsior L.) to provide evidence of illegal logging. Wood Science and Technology 51:1377-1387 doi:10.1007/s00226-017-0942-5
Uyoh E, Nkang A, Eneobong E (2003) Biotechnology, genetic conservation and sustainable use of bioresources. African Journal of Biotechnology 2:704-709 doi:10.5897/AJB2003.000-1129
Vlam M et al. (2018) Developing forensic tools for an African timber: Regional origin is revealed by genetic characteristics, but not by isotopic signature. Biological Conservation 220:262-271 doi:10.1016/j.biocon.2018.01.031

Thesis Appendix

LIST OF THESIS APPENDICES

APPENDIX I: LIST OF ADDITIONAL WORKS UNDERTAKEN THROUGHOUT CANDIDATURE

*NB: *Denotes full text included (and which appendix)*

JOURNAL ARTICLES

- EE Dormontt, **DI Jardine**, K-J van Dijk, BF Dunker, RRM Belton, V Hipkins, S Tobe, A Linacre, AJ Lowe (submitted) *Genetic profiling of timber provides tools for the prosecution of illegal logging crimes*. Nature: Sustainability. *(Appendix II)*
- Z Baruch, MJ Christmas, MF Breed, GR Guerin, J McDonald, **DI Jardine**, E Leitch, N Gellie, S Caddy-Retalic, KE Hill, KP McCallum, AJ Lowe (2017) *Leaf trait associations with environmental variation in the wide-ranging shrub Dodonaea viscosa subsp. angustissima (Sapindaceae)*. Austral Ecology; 42(5) 553-561
- GR Guerin, E Biffin, **DI Jardine**, HB Cross, AJ Lowe (2014) *A spatially predictive baseline for monitoring multivariate species occurrences and phylogenetic shifts in Mediterranean southern Australia*. Journal of Vegetation Science; 25(2) 338-348

OTHER PUBLICATIONS

- (Blog article) **DI Jardine**, AJ Lowe (2016) "Are you a threat to the Siberian tiger?" In: Biodiversity Revolution https://biodiversityrevolution.wordpress.com/2016/06/19/are-you-a-threat-to-the-siberian-tiger/. *(Appendix III)*
- (Conference paper) **DI Jardine**, EE Dormontt, J Gerlach, B Degen, AJ Lowe (2015) *Development of a Genographic map of ayous (Triplochiton scleroxylon K. Schum) for concessional level identification*. XIV World Forestry Congress, Durban, South Africa, 7-11 September 2015. *(Appendix IV)*
- (Patent) AJ Lowe, **DI Jardine**, HB Cross, B Degen, L Schindler, AM Holtken (2015). *A Method of Extracting Plant Nucleic Acids from Lignified Plant Tissue*. Patent number: WO/2015/070279A1 (Australia). Application number: AU2014/001047. *(Appendix V)*
- (Online dataset) GR Guerin, E Biffin, **DI Jardine**, HB Cross, AJ Lowe (2015). TREND (PSRF) vegetation plot data 2011, Version 1. DOI 10.4227/05/54AB6B443D1D3. Obtained from Australian Ecological Knowledge and Observation System Data Portal (ÆKOS, http://www.portal.aekos.org.au/).
- (Book section) **DI Jardine** (2014) *Wood extraction: The Basics*, In The Migrant Ecologies Project: Jalan Jati (Teak Road) Yu-Mei Balasingamchow (Ed) 1: 186-189 (2013) ISBN: 9789810755317. *(Appendix VI)*
- (Technical Report) EE Dormontt, DI Jardine, AJ Lowe, D Thomas (2013) *Big Leaf Maple Final Technical Report*, version 002. *(Appendix VII)*

PRESENTATIONS

- **DI Jardine**, EE Dormontt, J Austin, AJ Lowe (05-08 June 2017). *Genetic assignment testing on tree species*. Society for Wildlife Forensic Sciences (SWFS) Conference. Edinburgh United Kingdom
- **DI Jardine**, EE Dormontt, J Austin, AJ Lowe (18-23 September 2016). *Genetic assignment testing on tree species*. Australian and New Zealand Forensic Sciences Society Conference, Auckland, New Zealand (theme keynote)
- **DI Jardine**, EE Dormontt, BF Dunker, A Linacre, S Tobe, AJ Lowe, J Austin (18-23 September 2016). *When a tree is felled in the forest, will anybody stand up for it in court? Forensic validation of genetic methods for timber species*. Australian and New Zealand Forensic Sciences Society Conference, Auckland, New Zealand
- **DI Jardine**, EE Dormontt, J Gerlach, K-J van Dijk, B Degen, AJ Lowe (7-11 September 2015). *Genographic mapping of an African tropical Hardwood - application to DNA tracking*. World Forestry Congress. Durban, South Africa
- **DI Jardine**, EE Dormontt, AJ Lowe, J Austin (1-4 September 2014). *Development of genetic techniques for use in the detection of illegal logging of valuable timber species*. Australian and New Zealand Forensic Sciences Society Conference. Adelaide, SA, Australia (theme keynote)

APPENDIX II: MAPLE PAPER

Genetic profiling of timber provides tools for the prosecution of illegal logging crimes

Illegal logging drives substantial negative environmental, social and economic consequences, yet timber identification remains a challenge for law enforcement. Here we forensically validated 131 genetic markers for individualisation testing of bigleaf maple (*Acer macrophyllum* Pursch) and applied them in a successful illegal logging prosecution case. Our work demonstrates how forensic timber identification can support detection and prosecution of illegal logging and could verify legality and sustainability in reputable supply chains.

Illegal logging is a pernicious threat to biodiversity, to the protection and sustainable use of forests, and to the communities and economies who rely upon legal utilisation of forest products for their livelihoods. Globally, illegal logging has an estimated worth beyond US$30 billion annually and in some countries can constitute up to 90% of timber traded[1]. Until recently, there was little incentive for traders to actively ensure the legality of their timber. However, the amended *Lacey Act 2008* (USA) outlawed both international and domestic trade in illegally harvested wood, and was followed by the European Union's Timber Regulations (2010) and the *Illegal Logging Prohibition Act 2012* (Australia), which both have the same broad intent and application[2].

Enforcement of anti-illegal logging laws is challenging due to the lack of methods for identification of timber[3,4]. Despite recent interest in method development [5-7], forensic validation studies required to demonstrate legal suitability are generally lacking in the published literature, hindering wider uptake by law enforcement.

DNA profiling provides a promising prospect, however the application of genetic markers to timber presents several challenges. Timber contains limited quantities of low quality DNA which further degrades with age[8]. Furthermore, traditional forensic markers such as microsatellites are often not reliably amplified from timber extracted DNA[9]. There is also ambiguity regarding how best to apply forensic validation to genetic identification of timber. Existing guidelines were developed predominantly for human identification[10] and are not straightforwardly adapted to non-humans. This problem has been considered in some depth for animals[11,12], but so far the same consideration has not been given to plants[13].

Across the Pacific Northwest of the USA, bigleaf maple is regularly stolen from national forests (Fig. 1), finding its way into supply chains for the music wood industry. Here we present the first forensically validated individualisation test for timber and document its use in the first successful domestic prosecution of the *Lacey Act 2008*.

A set of 131 genetic markers (Supplementary Table 1) were forensically-validated based on a reference database of 394 trees from 43 separate sites (Fig. 1, Supplementary Table 2). Two quality control (QC) thresholds were applied; a maximum assay fail rate of 5% was determined through examination of all profiles generated for *A. macrophyllum* individuals (plus negative controls) and selected to maximise informative and minimise erroneous inclusions (Supplementary Fig. 1). A minimum DNA concentration threshold of 0.625 ng/μL was determined through sensitivity testing. Of 183 genotyped loci, 135 were selected that conformed to Hardy-Weinberg equilibrium expectations, were not significantly linked to other loci and had an observed heterozygosity of more than 0.1. To determine heritability of the markers, genotypes of mother trees and seeds were examined. Twelve of the 20 mother trees genotyped passed QC along with 98 seeds in total, collected from around these trees. Simulations to identify seeds incorrectly assigned to a mother-tree (Supplementary

Methods) led to the exclusion of 23 seeds from comparisons (Supplementary Fig. 2). For final assessment, 10,074 loci comparisons were made between 75 seeds and 10 mother-trees. Discordance was observed in five comparisons across four loci (0.05%) which were subsequently excluded from further analyses giving a final set of 131 loci (Supplementary Table 1).

To determine sensitivity, a range of DNA concentrations (10, 5, 2.5, 1.25 and 0.625 ng/µL) were tested in two arbitrarily selected individuals which all returned the expected genotype, the lower limit of sensitivity is likely less than 0.625ng/µL but this has been applied as a minimum QC threshold as lower concentrations have not been tested. To assess species specificity, DNA from non-target species with increasingly distant relatedness to *A. macrophyllum* were analysed and failed to produce genotypes that passed QC (vine maple (*Acer circinatum*), African teak (*Pericopsis elata*), speargrass (*Austrostipa* sp.), red alga (*Schizymenia dubyi*), spectacled fruit bat (*Pteropus conspicillatus*)). The congener *A. circinatum* did returned genotype calls in 76% of loci, as may be expected from closely related species. All other species tested did not amplify at any loci.

For the assessment of repeatability, 87 technical replicate pairs passing QC were checked for concordance in 11,360 loci comparisons. No discordant loci were observed, giving an error rate per locus of <0.009%. Reproducibility was assessed using five paired cambium and sawn timber samples, and seven paired leaf and cambium samples. Four cambium/leaf pairs passed QC and showed 100% concordance. Four of the five cambium/timber pairs extracted using different methods in different laboratories by different analysts passed QC and showed 100% concordance. Fourteen unpaired sawn timber samples were genotyped to further assess the reliability of the test when applied to DNA from case-type samples. Nine of 19 samples passed QC, giving a success rate for profile generation from DNA from sawn timber of 47%.

Probability of identity (PID) is the probability that two randomly selected individuals within a population have the same genotype at a given set of markers. The most conservative (highest estimate) of PID was derived using a dataset F_{ST} correction of the most common genotype and was 1.785×10^{-25}, equating to a random match probability of less than 1 in 5 septillion (Supplementary Methods and Results).

The forensic test was applied to samples collected from a site of bigleaf maple theft in the Gifford Pinchot National Forest, Washington, USA. DNA was extracted from the resprouting stumps and the remaining logs of two illegally felled trees along with four offcuts seized by US Forest Service (USFS) officers from the premises of a suspect. Between two and three replicates per evidence item returned genotypes passing QC thresholds for all four timber offcuts. Quality control samples passed indicating no contamination and that all reagents were working correctly. Genotypes of two offcuts were identical. The genotype from a third was identical to that obtained from the resprouting shoots of the stump of an illegally felled bigleaf maple tree from the Gifford Pinchot National Forest and from felled wood found around this stump. A forensic report provided to the USFS detailing the analysis was submitted as evidence to support the prosecution of four suspects over the timber theft. By early 2016, all defendants pled guilty.

This work was undertaken to address the needs of law enforcement in a specific criminal case; the DNA test developed was applied under forensic conditions to case samples and presented to court. The use of genetic profiling of timber as supporting evidence in an illegal logging prosecution sets a precedent for wider acceptance in criminal cases. The stringent

forensic validation process employed is designed to ensure that the strict criteria for court acceptance are met, however as the defendants pled guilty, the test has yet to withstand the scrutiny of a criminal defence team in court.

Broader application of DNA testing in timber could provide a means for legitimate traders to demonstrate their compliance with the law and sustainability practices, through the linking of products to their original felling sites providing supply chain verification[2,9]. Our work demonstrates how genetic identification of timber can be effectively employed to support law enforcement to detect and deter illegal logging and could also be used to support industry compliance through verification of legality and sustainability.

Methods:

Sampling and genetic analysis

Sampling was conducted in 2014 in Washington, Oregon, California and Southern Canada (Fig.1, Supplementary Table 2). Field collections were taxonomically verified with two voucher specimens (University of Washington Herbarium (WTU), accession numbers WTU403124, WTU403125). Cambium from each tree was sampled using a leather punch and mallet. Seven individuals from different sites were sampled for leaves, five were sampled at a sawmill for cambium and sawn timber, and 14 for only sawn timber. For heritability studies, cambium was sampled from 20 mother trees and up to 50 mature seeds collected from around their base (Supplementary Methods). All cambium, leaf and timber samples were stored in silica gel prior to DNA extraction. Seeds were stored in paper bags prior to germination. Of the 199 previously published single nucleotide polymorphism (SNP) and five insertion/deletion (INDEL)[14], 183 were successfully multiplexed for the MassARRAY platform (Agena Bio-science) and used to genotype all samples. DNA was extracted using either the Nucleospin Plant II Kit (Machery-Nagel) with the PL2/PL3 buffer system (for cambium and leaf, undertaken at the Australian Genome Research Facility) or a patented timber extraction method[15] (for sawn timber, undertaken at the University of Adelaide). Samples for extraction were arranged into 96 well plates, each containing an identical sample as a positive control. Separate negative controls for DNA extraction and genotyping procedures were analysed. Any loci which consistently failed to amplify were excluded. Conformity to Hardy-Weinberg equilibrium expectations and linkage disequilibrium between loci were assessed with Genepop[16] using Fisher's exact tests. Observed heterozygosity was calculated using GenoDive[17]. The probability of identity (PID) was calculated in GenAlEx[18,19] and F_{ST} correction (calculated in Genepop[16]) applied (Supplementary Methods). We interpreted and applied the SWGDAM validation guidelines for DNA analysis methods[10] specifically for the developmental validation of individualisation tests for timber species.

Application to illegal logging investigation

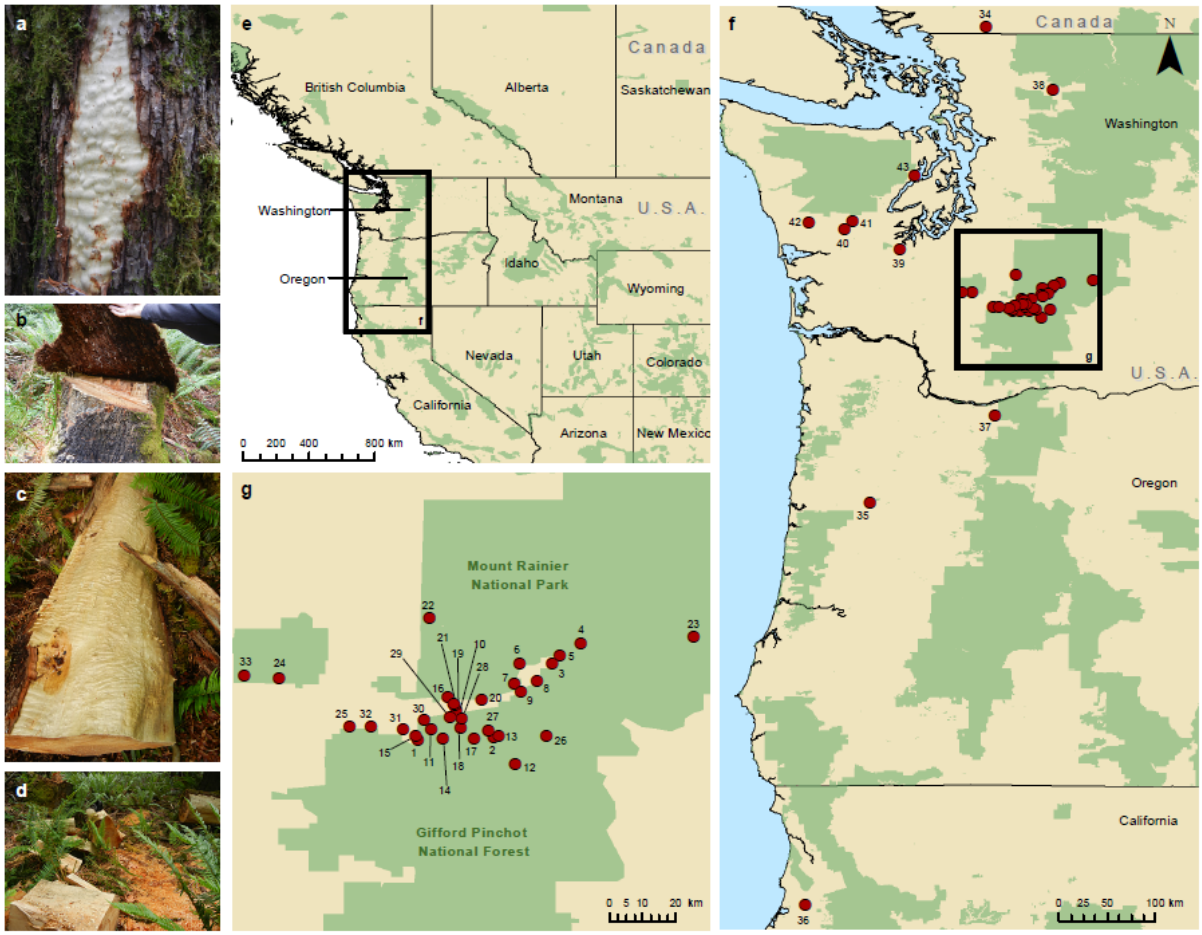
The forensically validated genetic individualisation test was applied to samples collected from a site of bigleaf maple theft in the Gifford Pinchot National Forest, Washington State, USA. Collection and DNA extraction was undertaken by the USFS with DNA shipped to Australia for analysis. DNA from the resprouting stumps of two illegally felled trees was extracted using the DNeasy 96 Plant kit (Qiagen) with the liquid nitrogen modification. Wood was also collected from the remaining felled logs associated with these stumps and DNA extracted using the DNeasy Plant Maxi Kit (Qiagen). Four bigleaf maple offcuts were seized by USFS officers from a suspect's premises and sent to Australia for comparison to the DNA extracted from the illegally felled trees. DNA was extracted from the timber offcuts using the patented timber extraction method[16]. Three extractions were undertaken per offcut and positive and negative controls were included.

References

1 Nellemann, C. & INTERPOL in *Green Carbon, Black Trade: Illegal Logging, Tax Fraud and Laundering in the Worlds Tropical Forests* (UNEP 2012)
2 Lowe, A. J. *et al. BioScience* **66**, 990-998 (2016).
3 Dormontt, E. E. *et al. Biol. Conserv.* **191**, 790-798 (2015).
4 UNODC *Best practice guide for forensic timber identification*. (United Nations, 2016).
5 Ng, K. K. S. *et al*. *Forensic Sci. Int.-Gen.* **23**, 197-209 (2016).
6 Hassold, S. *et al. PLoS One* **11**, e0157881, doi:10.1371/journal.pone.0157881 (2016).
7 Bergo, M. C. *et al*. *IAWA J.* **37**, 420-430 (2016).
8 Jiao, L. *et al*. *Holzforschung* **69**, 925-931, doi:10.1515/hf-2014-0224 (2015).
9 Lowe, A. *et al. Silvae Genet.* **59**, 263 (2010).
10 SWGDAM *Validation guidelines for DNA analysis methods* (2016).
11 Budowle, B. *et al. Int. J. Legal Med.* **119**, 295-302 (2005).
12 Linacre, A. *et al*. *Forensic Sci. Int.-Gen.* **5**, 501-505 (2011).
13 Iyengar, A. *J. Nat. Conserv.* **22**, 195-205 (2014).
14 Jardine, D. I. *et al*. *Conserv. Genet. Resour.* **7**, 797-801 (2015).
15 Lowe, A. J. *et al. A method of extracting plant nucleic acids from lignified plant tissue*. (2015).
16 Rousset, F. *Mol. Ecol. Resour.* **8**, 103-106 (2008).
17 Meirmans, P. G. & Van Tienderen, P. H. *Mol. Ecol. Notes* **4**, 792-794 (2004).
18 Peakall, R. & Smouse, P. E. *Bioinformatics* **28**, 2537-2539 (2012).
19 Peakall, R. & Smouse, P. E. *Mol. Ecol. Notes* **6**, 288-295 (2006).

Figure 1. Bigleaf maple illegal logging modus operandi and sampling locations. a, Trees are selected based on 'figured' patterning, observed through removal of the bark in a small area, 'quilted' figuring shown here. **b**, Once felled, the stump is often covered in moss to evade detection. **c**, Bark is removed from the felled log revealing the extent of figured patterning. The log is often then cut into blocks. **d**, In some cases little attempt is made to conceal the crime scene, only the most valuable (and portable) blocks are removed with the remaining timber left to rot in the forest. **e**, Western North America showing National Parks and Forests (green). **f**, Total study area with sampling locations (red circles). **g**, Area of most extensive sampling. Photo credits: Anne Minden.

NB. High resolution figure can be downloaded from
https://figshare.com/s/6d7bc936353772aa16ef



Supplementary Information

Supplementary Methods

Heritability

The heritability of the genetic markers was assessed by observing the genotypes of mother-trees and their offspring. DNA was extracted from the cambium of the mother trees as previously described. Offspring genotypes were assessed through germination, DNA extraction and genotyping of the seeds collected from around the mother trees.

Prior to germination, the seeds were dried and stored in paper bags at room temperature. Germination was achieved using a combination of stratification and gibberellic acid. The 'wing' of each seed was removed. Seeds were soaked in 4% bleach solution for five minutes before rinsing thoroughly with distilled water to protect against fungal infection. Seeds from around the base of each mother tree were germinated separately by placing eight seeds into a 90 mm diameter Petri dish on top of two filter papers. Filter papers were soaked with a 200 mg/l of gibberellic acid potassium salt solution[1] and dishes were incubated in a closed box for one month at 4 °C. For three mother-trees, an additional 16 seeds were germinated in the same way. Seeds were checked every 4–6 days, maintaining moist filter papers with the gibberellic acid solution and changing filter papers if a large amount of brown exudate was present.

Seedling embryos were dissected from the pericarp and seed coat using a scalpel and tweezers (Supplementary Fig. 2) and DNA extracted as described in the main paper. Non-linked loci conforming to Hardy-Weinberg expectations and with an observed heterozygosity >0.1 were assessed in the genotypes of mother-trees and seeds which passed QC (methods described in the main paper). Each locus was scored as concordant when the seedling and mother-tree genotypes shared at least one allele, according to Mendelian inheritance expectations. The winged samaras of *A. macrophyllum* are adapted to disperse seeds away from the mother-tree, so collections on the ground around the base of trees may occasionally include seeds from different trees. To identify seedlings incorrectly assigned to a mother-tree, a simulation approach was employed using Resampling Stats for Excel v4.0 (Statistics.com). This approach was based on the premise that discordance associated with incorrect parentage assumptions will be non-randomly distributed throughout the individual seed genotypes. Ten thousand artificial seed genotypes were generated containing the same number of loci as genotyped in the real samples. Loci were randomly assigned as concordant or discordant at each locus from the pool of data observed in the real mother-tree and seed comparisons, and selected without replacement. Numbers of discordant loci in the artificial seedlings were calculated and their distributions observed (Supplementary Fig. 3). In the real seed and mother-tree data set, any seed containing more discordant loci than observed in 99% of the simulated seeds was removed. Loci where discordance was observed in the remaining seed and mother-tree comparisons were removed from the final data set.

Probability of identity

The probability of identity (PID) and the more conservative probability of identity for siblings (PID_{sib}) were calculated for each locus using GenAlEx[2,3] both within each population and across all samples. F_{ST} per locus was calculated using Genepop[4] and used as correction factors to the PID at each locus[5]. A standard data set F_{ST} value was also calculated by taking the mean across all loci plus three standard deviations and applied as a single correction factor to the PID at each locus. No correction factors were added to the PID_{sib} calculations.

References

1 Stejskalová, J. *et al. J. For. Sci.* **61**, 325-331 (2015).
2 Peakall, R. & Smouse, P. E. *Mol. Ecol. Notes* **6**, 288-295 (2006).
3 Peakall, R. & Smouse, P. E. *Bioinformatics* **28**, 2537-2539 (2012).
4 Rousset, F. *Mol. Ecol. Resour.* **8**, 103-106 (2008).
5 Balding, D. J. & Nichols, R. A. *Forensic Sci. Int.* **64**, 125-140 (1994).

Supplementary Results

Probability of identity

The uncorrected probability of identity (PID) across all loci and individuals was 7.576×10^{-43}, which equates to a random match probability of less than 1 in a tredecillion. The PID of the most common and rarest genotypes were 6.187×10^{-30} and 1.227×10^{-120} respectively, calculated using locus specific F_{ST} corrections. These figures equate to random match probabilities of less than 1 in 161 octillion, and less than 1 in 81 quintillion googol, respectively. When a dataset F_{ST} correction was used, the PID of the most common and rarest genotypes were 1.785×10^{-25} and 2.616×10^{-81} respectively, equating to random match probabilities of less than 1 in 5 septillion and less than 1 in 382 quadrillion vigintillion respectively.

The PID for siblings (PID_{sib}) across all loci and individuals was 2.496×10^{-22}, equivalent to a random match probability of less than 1 in 4 sextillion. The most conservative PID_{sib} within individual populations was 7.512×10^{-17} found in population 38 located in northern Washington (Main Article Fig. 1). This equates to a random match probability of less than 1 in 13 quadrillion. The least conservative PID_{sib} within individual populations was 5.127×10^{-22} found in population 31 located in the Gifford Pinchot National Forest in Washington (Main Article Fig. 1). This equates to a random match probability of less than 1 in a sextillion.

Supplementary Table 1. List of 131 loci used for individualisation in bigleaf maple, a subset of those developed by Jardine *et al.* [1]. H_O: Observed heterozygosity per locus calculated using Genodive[2]. F_{ST}; Wright's fixation index per locus calculated using Genepop[3].

Locus type	Locus name	Alleles	H_O	F_{ST}
SNP	AM1f_1134	A/T	0.446	0.0956
	AM1f_1228	A/G	0.432	0.1044
	AM1f_1283	A/G	0.417	0.0357
	AM1f_1351	A/G	0.247	0.0778
	AM1f_2_2725	C/G	0.296	0.0440
	AM1f_2_389	C/T	0.486	0.0813
	AM1f_215	C/T	0.263	0.0549
	AM1f_2335	C/T	0.435	-0.0132
	AM1f_287	C/T	0.117	0.0508
	AM1f_441	A/C	0.187	0.0200
	AM1f_481	C/T	0.378	0.0851
	AM1f_5144	C/T	0.426	0.0668
	AM1f_524	C/T	0.253	0.0779
	AM1f_5928	C/T	0.269	0.0850
	AM1f_924	C/G	0.262	0.1003
	AM1f_927	A/G	0.277	0.0447
	AM1f_984	A/T	0.451	0.0703
	AM2f_18	A/G	0.100	0.0691
	AM2f_2_123	C/T	0.098	0.0736

Locus type	Locus name	Alleles	H_O	F_{ST}
	AM2f_218	C/T	0.452	0.0392
	AM2f_234	A/G	0.365	0.0224
	AM2f_290	A/T	0.456	0.0342
	AM2f_346	C/T	0.359	0.0476
	AM2f_49	C/T	0.438	0.0592
	AM2f_617	A/G	0.469	0.1180
	AM2f_629	C/G	0.204	0.0628
	AM2f_9	C/T	0.404	0.0518
	Maple_1086	G/T	0.431	0.0874
	Maple_10862	C/T	0.291	0.1574
	Maple_1191_e	G/T	0.300	0.0422
	Maple_121	A/G	0.410	0.0781
	Maple_12182	A/C	0.207	0.0238
	Maple_13_bis	A/T	0.177	0.0353
	Maple_1308	G/T	0.364	0.0612
	Maple_1481	A/G	0.085	0.0490
	Maple_1489	G/T	0.454	0.0507
	Maple_1557	C/G	0.397	0.0835
	Maple_1643	G/T	0.295	0.0199
	Maple_1752	C/T	0.428	0.0861
	Maple_1856	C/T	0.320	0.0094
	Maple_1906	C/T	0.300	0.0568
	Maple_20	C/T	0.265	0.0813
	Maple_2059	C/T	0.108	0.0354
	Maple_2074	C/T	0.295	0.0724
	Maple_2076	C/T	0.115	0.0182
	Maple_2109	C/T	0.265	0.0561
	Maple_2155	G/T	0.457	0.0632
	Maple_2417	A/G	0.432	0.0667
	Maple_2793	C/T	0.457	0.1023
	Maple_305	A/G	0.466	0.0399
	Maple_3075	A/G	0.469	0.0249
	Maple_3089	A/G	0.437	0.0320
	Maple_3090	A/C	0.351	0.0748
	Maple_3120	A/T	0.281	0.0944
	Maple_3136	A/C	0.303	0.0412
	Maple_3252	C/T	0.374	0.0558
	Maple_3773	A/G	0.217	0.0483
	Maple_3784	A/G	0.400	0.0937
	Maple_3814	C/T	0.282	0.0540
	Maple_3882	C/T	0.229	0.0683
	Maple_3918	C/T	0.393	0.0331
	Maple_3941	C/T	0.181	0.0150
	Maple_3989	C/T	0.423	0.0871
	Maple_3999	A/G	0.430	0.0707
	Maple_4002	A/G	0.404	0.0411
	Maple_4034	C/T	0.149	0.0511
	Maple_4044	C/T	0.375	0.0630
	Maple_4049	A/C	0.448	0.0834

Locus type	Locus name	Alleles	H_O	F_{ST}
	Maple_4091	C/T	0.278	0.0546
	Maple_4144	C/T	0.398	0.0323
	Maple_4174	C/T	0.322	0.0456
	Maple_4186	C/T	0.354	0.1032
	Maple_4218	A/G	0.366	0.0502
	Maple_4258	A/G	0.443	0.0890
	Maple_4278	C/T	0.326	0.0402
	Maple_4308	C/T	0.180	0.0690
	Maple_4381	C/T	0.424	0.0854
	Maple_4385	C/G	0.432	0.0354
	Maple_4393	C/T	0.477	0.0588
	Maple_4438	A/G	0.187	0.0681
	Maple_4455	C/T	0.133	0.0865
	Maple_4472	C/T	0.096	0.0434
	Maple_4484	A/C	0.301	0.0871
	Maple_4512	C/T	0.188	0.1584
	Maple_4566	C/T	0.417	0.0546
	Maple_4604	A/G	0.388	0.0356
	Maple_4663	C/T	0.477	0.0619
	Maple_4665	A/G	0.457	0.0402
	Maple_4696	A/G	0.416	0.0351
	Maple_4723	C/G	0.177	0.0184
	Maple_4724	C/T	0.443	0.0875
	Maple_4731	C/T	0.357	0.0339
	Maple_4840	G/T	0.179	0.0249
	Maple_4847	C/T	0.327	0.0352
	Maple_4850	C/G	0.418	0.1130
	Maple_4896	A/G	0.111	0.0410
	Maple_4923	C/T	0.411	0.0429
	Maple_4998	A/G	0.451	0.0718
	Maple_5062	A/G	0.441	0.0786
	Maple_5095	G/T	0.335	0.0821
	Maple_5112	A/G	0.418	0.0629
	Maple_5227	A/C	0.466	0.0710
	Maple_5231	G/T	0.234	0.0577
	Maple_5287	A/G	0.403	0.0826
	Maple_5345	A/G	0.468	0.0464
	Maple_5418	C/T	0.377	0.0438
	Maple_5421	C/G	0.321	0.0665
	Maple_5646	C/T	0.372	0.0698
	Maple_5761	A/G	0.418	0.0856
	Maple_5820	A/G	0.438	0.0916
	Maple_6002	C/T	0.389	0.0752
	Maple_6157	A/C	0.386	0.1529
	Maple_6318	A/T	0.291	0.0474
	Maple_6339	A/G	0.477	0.0382
	Maple_65	A/T	0.441	0.0840
	Maple_6560	A/G	0.457	0.0413
	Maple_659	A/G	0.427	0.0455

Locus type	Locus name	Alleles	H_O	F_{ST}
	Maple_6626	A/T	0.321	0.0175
	Maple_6682	C/G	0.479	0.0502
	Maple_679	C/T	0.407	0.0911
	Maple_7509	A/C	0.434	0.1070
	Maple_7588	C/T	0.323	0.0412
	Maple_7702	A/C	0.364	-0.0037
	Maple_7856	C/T	0.276	0.0176
	Maple_823	A/G	0.487	0.0524
	Maple_8688	C/T	0.452	0.1109
	Maple_9048	A/C	0.121	0.1139
	Maple_9291	C/T	0.381	0.1279
INDEL	Maple_4074	C/DEL	0.486	0.0243
	Maple_4829	AT/DEL	0.478	0.1144
	Maple_8509	TTCAGG/DEL	0.22	0.1037

1 Jardine, D. I. *et al*. *Conserv. Genet. Resour.* **7**, 797-801 (2015).
2 Meirmans, P. G. & Van Tienderen, P. H. *Mol. Ecol. Notes* **4**, 792-794 (2004).
3 Rousset, F. *Mol. Ecol. Resour.* **8**, 103-106 (2008).

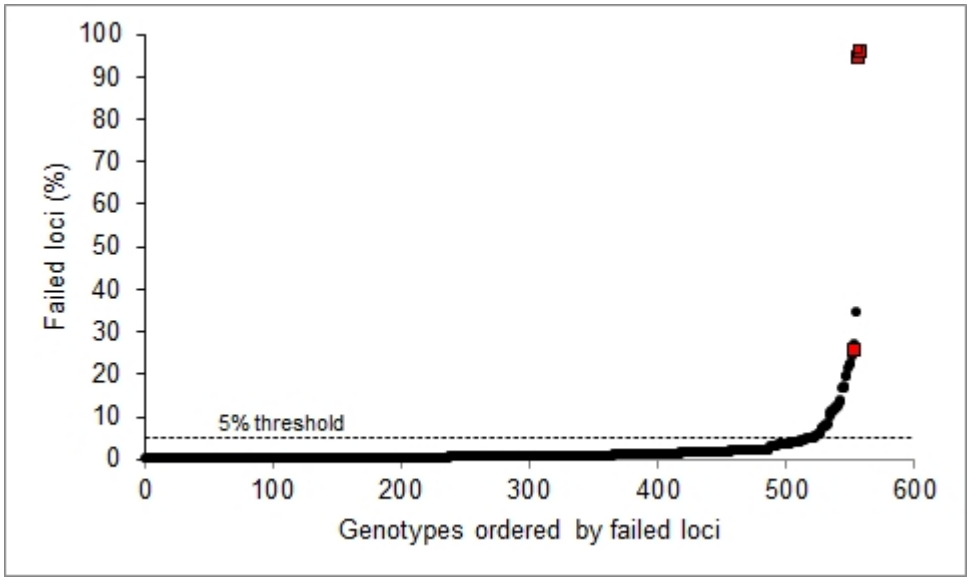
Supplementary Table S2. Summary information on location and number of individuals of bigleaf maple successfully genotyped at each sampling site.

Site number	Number of individuals	Latitude	Longitude
1	7	46.4284	-121.989
2	9	46.4363	-121.784
3	8	46.63667	-121.625
4	7	46.69267	-121.547
5	8	46.65992	-121.602
6	8	46.63739	-121.712
7	6	46.58306	-121.727
8	10	46.58944	-121.664
9	8	46.56144	-121.709
10	9	46.51083	-121.884
11	10	46.45961	-121.952
12	10	46.36533	-121.723
13	10	46.44133	-121.769
14	10	46.43278	-121.92
15	10	46.44111	-121.996
16	10	46.54536	-121.907
17	10	46.43247	-121.835
18	10	46.46325	-121.873
19	10	46.52544	-121.89
20	10	46.53953	-121.814
21	10	46.52664	-121.893
22	10	46.76211	-121.957

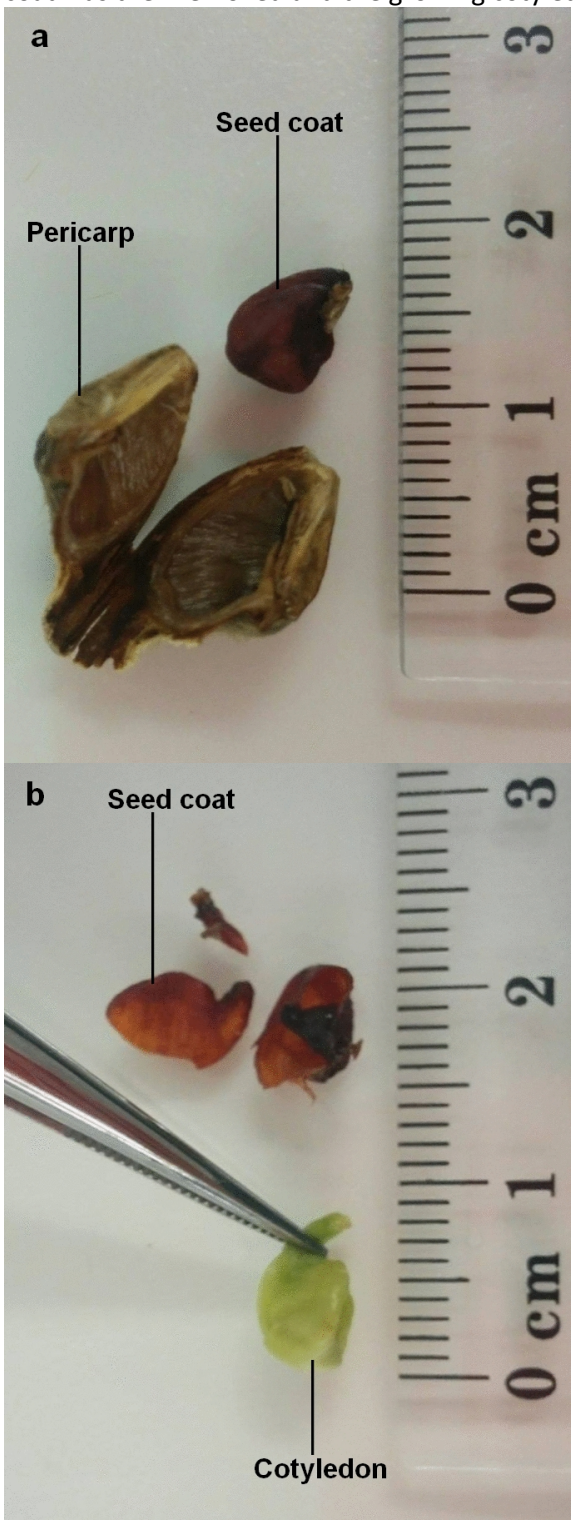
Site number	Number of individuals	Latitude	Longitude
23	7	46.70983	-121.239
24	9	46.59794	-122.367
25	10	46.46483	-122.174
26	10	46.43936	-121.639
27	10	46.45586	-121.797
28	10	46.48836	-121.869
29	10	46.49122	-121.901
30	10	46.48339	-121.972
31	10	46.45717	-122.028
32	10	46.46564	-122.117
33	10	46.60486	-122.46
34	10	49.07724	-122.239
35	10	44.64311	-123.322
36	9	40.89623	-123.925
37	10	45.44637	-122.157
38	8	48.48932	-121.616
39	9	46.99793	-123.041
40	8	47.18842	-123.562
41	8	47.26137	-123.48
42	7	47.24794	-123.892
43	9	47.68933	-122.903

Supplementary Figure 1

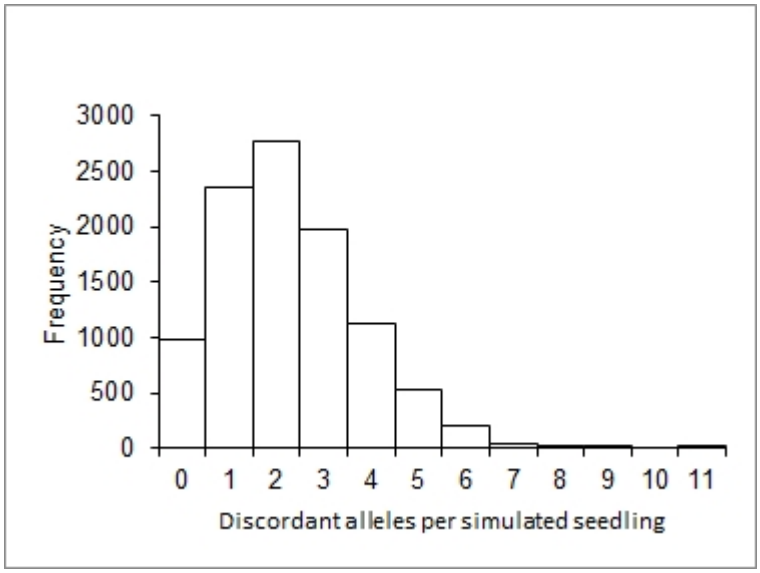
Assay fail-rate threshold. The percentage of loci which failed to amplify in each genotype (excluding 100%) with data sorted in order of increasing number of failed loci. Profiles from tested samples are represented (black circles), along with reagent blanks (red squares). The dashed line represents the maximum 5% fail-rate threshold applied to sample genotypes as part of quality control to ensure exclusion of erroneous genotypes in subsequent analyses.



Supplementary Figure 2
Dissection of germinating *Acer macrophyllum* seeds. **a**, After cold stratification, germinating seeds were dissected to reveal the seed coat and separate it from the pericarp. **b**, The seed coat was then removed and the growing cotyledons used for DNA extraction.



Supplementary Figure 3. Distribution of discordant allele calls in 10,000 simulated seedlings.


3000
2500
2000
1500
1000
500
0
Frequency
0 1 2 3 4 5 6 7 8 9 10 11
Discordant alleles per simulated seedling

APPENDIX III: ARE YOU A THREAT TO THE SIBERIAN TIGER?

Blog article

post from Andrew Lowe's Blog *Biodiversity Revolution*

Are you a threat to the Siberian tiger?

Posted on June 19, 2016 by Prof Andy Lowe

You are probably unintentionally contributing to the future demise of the Siberian tiger. Tiger habitat, predominantly Mongolian oak, is being destroyed by illegal logging. As consumers of oak furniture, which has potentially been illegally sourced from tiger habitat, we are all part of the problem, but we can also be part of the solution. The next time you buy solid oak furniture, ask where it comes from?

The Siberian tiger, at home in the Mongolian oak forests (www.worldwildlife.org)

The recent announcement by the WWF that global tiger numbers have increased for the first time in a century has been widely celebrated. There is no doubt that this is a significant step toward increasing the tiger population to a more sustainable level. Yet, there is more to be done to reach the Global 2020 conservation targets.

In Australia, you may wonder how you can help. Given Australia's tough laws and adherence on CITES species, there is little to no tiger products making their way onto our shores, and most Australians are not seeking out tigers or tiger products when overseas. Thus, you may conclude that you have no impact on the tiger species.

Yet you probably are unintentionally contributing to the future demise of this species. Tiger habitat, predominantly Mongolian oak, is being destroyed by illegal logging and is one of the major limitations to the increase in tiger numbers. Australians are part of this problem, but we can also be part of the solution. Most people have some sort of solid timber furniture in their houses or workplaces. Do you know what species it is? Oak is a popular and abundant timber for furniture. Yet, there are many species of oak. Do you know which ones are in your home?

Furniture is typically just sold as "oak". If you ask a retailer to name the species, or even the origin of the "oak", you will most likely be met with a blank stare. There are at least 600 known species of oak, but only 19 species are used for their timber. The most commonly traded oak species are northern Red Oak (Quercus rubra) from North America, and the European Oaks (Q. petraea, Q. robur). Due to the high demand for these species, the lesser known Mongolian Oak (Q. mongolica) is being substituted for these species on a regular basis. Mongolian oak is a CITES appendix III (Russian Federation) listed species. This means that Russia is seeking international assistance to control the trade of the species. A requirement of this legislation is that certification to allow the trade of the species must be attached to a consignment.

Mongolian Oak is being illegally logged from the Far East Russian forests by Russian criminal organisations and smuggled into China, where the timber is then sold as European or American Oak species. Estimates from both WWF and EIA indicate that the trade of Mongolian oak from Russia into china is between 200-400% of the permitted volumes. However it is not all doom and gloom, and recently a US company, Lumber liquidators, were fined US$13 million for illegally trading in Mongolian oak.

The effects of illegally logging Mongolian Oak are profound. The oak forests of Russia's Primorsky Province, is home to the last known wild population of Siberian tigers (also known as the Amur tiger), of which only about 500 are known to exist. These forests are also the habitat of the highly endangered Amur Leopard (only ~60 known animals). These predators depend on intact forests for their primary source of food, Wild Boar and Red

Deer, which eat the acorns of Mongolian oak. The loss of habitat also means the tigers are more easily spotted by poachers.

Effects of illegal logging in Russian Far east Forest

One of the main limitations to controlling trade of Mongolian oak is that it is difficult to distinguish the wood of oak species by eye, especially with no formal training. However, genetic markers have been developed that can routinely distinguish between the major species of white oak (Q. rubra, Q. petraea, Q. robur, and Q. mongolica).

We as timber consumers, especially oak furniture, can help out. Be an informed buyer of oak goods. Asking questions of the retailer is a good place to start:

- What species of oak is this?
- What is the origin of this timber?
- Is there any certification on the origin of this product?
- Has the company used due diligence to identify the origin of their products?
- What is the company doing to comply with timber trade legislation?

Alternately, purchasing only products that have certification (such as FSC or PEFC) is a good way to start.

The more we as consumers can put pressure on companies to identify the origin of their timber and provide sustainable products, the more we can achieve, not just for the Siberian tigers, but also for all plant and animal species that are affected by illegal logging.

By Duncan Jardine

Key References

- Dormontt, EE, et al. “Forensic timber identification: It’s time to integrate disciplines to combat illegal logging.” Biological Conservation 191 (2015): 790-798. doi:10.1016/j.biocon.2015.06.038
- EIA (2013). Liquidating the Forests: Hardwood flooring, organised crime and the world’s ast Siberian tigers. eia-global.org/campaigns/forests-campaign/liquidating-the-forests/
- WWF (2013). Illegal logging in the Russian Far East: global demand and Taiga destruction. Milakovsky B (ed). http://www.worldwildlife.org/publications/illegal-logging-in-the-russian-far-east-global-demand-and-taiga-destruction

APPENDIX IV: DEVELOPMENT OF A GENOGRAPHIC MAP OF AYOUS (*TRIPLOCHITON SCLEROXYLON* K. SCHUM) FOR CONCESSIONAL LEVEL IDENTIFICATION

Conference paper

2015 World Forestry Congress (WFC), Durban, South Africa

XIV WORLD FORESTRY CONGRESS, Durban, South Africa, 7-11 September 2015

Development of a Genographic map of Ayous (*Triplochiton scleroxylon* K. Schum) for concessional level identification

Duncan Jardine[1], Eleanor Dormontt[2], Josephine Gerlach[3], Bernd Degen[4], Andrew Lowe[5]

[1] School of Biological Sciences, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia, duncan.jardine@adelaide.edu.au; (+61)883131257

[2] School of Biological Sciences, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia; eleanor.dormontt@adelaide.edu.au; (+61)407762169

[3] School of Biological Sciences, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia; josephine.gerlach@adelaide.edu.au; (+61)403992570

[4] Thünen Institute of Forest Genetics; Sieker Landstraße 2, 22927 Großhansdorf, Germany; bernd.degen@ti.bund.de; (+49)4102696101

[5] School of Biological Sciences, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia; andrew.lowe@adelaide.edu.au; (+61)883131149

Abstract

One of the major threats to global biodiversity is deforestation, particularly in tropical regions. Whilst timber production is an essential component of the economy of many developing countries, a large proportion of timber production comes from illegal sources. Consumer countries are now confronting the problem of illegal logging through the implementation of legislation, such as the USA *Lacey Act* (2008 amendment), the *European Union Timber Regulation* (EUTR) (2012), and the *Australian Illegal Logging Prohibition Act* (2013). Given that the results of DNA analyses are routinely used as evidence in court proceedings for human identification, and are frequently employed in animal identification cases, the application of DNA evidence in a forest legality framework is a logical move. DNA evidence to identify species or source origin can also provide law enforcers with a robust and reliable form of evidence. As part of an International Tropical Timber Organisation funded global collaboration to develop genetic markers for African timber species, we have been working on Ayous (*Triplochiton scleoxylon* K. Schum), one of the most economically important species from the Tropical West African region. To date, we have generated a preliminary genographic map using variable Single Nucleotide Polymorphism (SNP) loci that could be used to verify region of origin. Here we describe the protocol we have used to create the genographic map and demonstrate our preliminary findings with respect to provenance identification for Ayous using DNA. The DNA techniques described have the potential to be used by both the timber industry and law enforcement in their efforts to prevent illegal timber trading.

Keywords: Genographic map, Triplochiton scleroxylon, DNA fingerprinting, timber tracking, SNP

Introduction, scope and main objectives

Deforestation is considered to be one of the main threats to global biodiversity, with the rate of worldwide deforestation estimated at 14-16 million hectares per year (Finkeldey et al., 2007, Zahnen, 2007). The felling of trees for timber is one of the significant drivers of deforestation (Zahnen, 2007), and a large proportion of traded timber is estimated to come from illegal sources, where the global cost of illegal logging is estimated at 30% of a global trade value of €180 Billion per year (Degen, 2012, Degen et al., 2013). Up to fifty percent of timber exported from some major forestry regions including the Amazon, Central Africa, Russia and South East Asia are suspected as coming from illegally logged

sources (Degen, 2012, Degen et al., 2013, Tnah et al., 2009). The majority of illegally harvested timber are for those species that have high economic importance, with many of these timber species being protected under the UN Convention on international trade in Endangered Species (CITES) or listed on the International Union for Conservation of Nature and Natural Resources (IUCN) red list of endangered species (Degen, 2012, Jolivet and Degen, 2012).

Tropical West Africa is considered to be the second largest tropical forested area globally, and as for the Amazon, is at risk from the effects of illegal logging. According to ITTO statistics, in 2013, approximately 2.3 million m^3 of roundwood, 1.7 million m^3 of sawn timber, 230 thousand m^3 of veneer and 130 thousand m^3 of plywood was exported from Tropcial West African Countries (ITTO, 2014). Timber exports in Tropical West Africa come from range of species, including many that are CITES listed, such as *Pericopsis elata,* and several *Khaya* species. Ayous (*Triplochiton scleroxylon* K. schum) is one of the most economically important tree species in Africa (Hall and Bada, 1979, Orwa et al., 2009), and goes by other trade names including, African whitewood, African maple, Obeche, Wawa or Samba. Ayous wood has a number of uses ranging from construction and ornamental items, through to more specific uses such as sauna panelling and in table tennis bats (Bosu and Krampah, 2005, Hall and Bada, 1979, Orwa et al., 2009). Estimates from the early 2000's indicate that the combined exports of sawn timber from Ghana, Cameroon and Côte d'Ivoire were approximately 300,000m^3 per year.

Ayous is easily identifiable by its lobed leaves, as most associated species have entire leaves, and is seasonally deciduous. Trees can grow up to 50 m tall with a straight trunk up to 30 m in height. It has a broad distribution across the tropical West African rainforest, occurring from the Democratic Republic of Congo (DRC) in the East and extending Westwards through Central African Republic (CAR), Congo, Cameroon, Gabon, Equatorial Guinea, Nigeria, Benin, Ghana, Côte d'Ivorie, Liberia, Guinea and into the Eastern parts of Sierra Leone. Its occurrence is largely split into three major components, Sierra Leone-Togo, Benin-Nigeria and Cameroon-DRC; with the Dahomey Gap separating Togo from Benin and Cameroon Highlands splitting Nigeria and Cameroon (Bosu and Krampah, 2005, Hall and Bada, 1979, Igboanugo and Iversen, 2004, Orwa et al., 2009).

This project has built upon genetic techniques used previously to identify the source species and origin of biological products in a number of organisms, including fish, meat (cattle), elephants (ivory), plants, and Humans (Alaeddini et al., 2010, Gitzendanner, 2012, Gugerli et al., 2005, Jolivet and Degen, 2012, Lowe and Cross, 2011). Our technique uses DNA extracted from timber, which has already been shown to work as a tool for monitoring illegal logging and timber tracking (Degen and Fladung, 2007, Finkeldey et al., 2007, Lowe, 2007). Our approach involves the development and creation of genographic maps, which are made up of a set of genetically variable markers that can be used to identify the origins of an individual. These timber genographic maps are similar to those already implemented for human genetic analysis. Yet, it is acknowledged that the creation of such maps to solve the problem of illegal logging for all timber species is an endeavour several orders of magnitude larger than dealing with one species, such as humans. The datasets contributing to timber species genographic maps are still a relatively recent technique, and the greater the sampling intensity and genetic information provided for each species, the more precise and thorough the results (Tnah et al., 2009). The creation of a genetic database of timber species is a significant step toward accurately describing and identifying important timber species in the logging trade.

Methodology/approach

For this analysis a total of 48 individuals from 10 populations were used. These individuals were collected from three countries, which represent a broad part of the distribution of Ayous: Democratic Republic of the Congo (DRC) in the East, Cameroon in the centre and Ghana to the West. The full list of samples used, and their respective information, can be found in Table 1. DNA from all samples was extracted from leaf tissue in house at Thünen Institute of Forest Genetics (TFG) (Großhansdorf, Germany) and the extracted genomic DNA was then sent to the University of Adelaide (Adelaide, Australia) for analysis.

Table 1: List of individuals used in the preliminary genographic map analysis, as well as population, country and geographical coordinates.

Sample ID	Population code	Population #	Country	Latitude	Longitude
NB674	C-01	1	Cameroon	2.211668	10.264498
03_TRI_06	C-03	2	Cameroon	2.67112	12.83753
03_TRI_07	C-03	2	Cameroon	2.5799	12.93968
03_TRI_08	C-03	2	Cameroon	2.58193	12.93812
03_TRI_09	C-03	2	Cameroon	2.60122	12.74347
03_TRI_10	C-03	2	Cameroon	2.6029	12.74433
03_TRI_13*	C-03	2	Cameroon	2.74948	12.61883
13_TRI_01	C-13	3	Cameroon	4.62918	11.25585
13_TRI_02	C-13	3	Cameroon	4.82718	11.32100
13_TRI_03	C-13	3	Cameroon	4.85284	11.32797
13_TRI_04	C-13	3	Cameroon	4.64697	11.24887
13_TRI_05	C-13	3	Cameroon	4.66827	11.25397
13_TRI_11*	C-13	3	Cameroon	4.88375	11.35302
14_TRI_07	C-14	4	Cameroon	4.59007	13.22319
14_TRI_08	C-14	4	Cameroon	4.64413	12.55812
14_TRI_09	C-14	4	Cameroon	4.57043	13.43135
14_TRI_10	C-14	4	Cameroon	4.65138	12.43874
14_TRI_11	C-14	4	Cameroon	4.64825	12.51636
14_TRI_12*	C-14	4	Cameroon	4.60509	12.67308
15_TRI_11	C-15	5	Cameroon	5.23456	13.56328
15_TRI_12	C-15	5	Cameroon	5.26181	13.56083
15_TRI_13	C-15	5	Cameroon	5.35004	13.50260
15_TRI_14	C-15	5	Cameroon	4.70178	13.82965
15_TRI_15	C-15	5	Cameroon	4.71379	13.83376
15_TRI_17	C-15	5	Cameroon	4.66555	13.65435
02_TRI_04	DRC-02	6	DRC	0.75674	24.49406
02_TRI_05	DRC-02	6	DRC	0.75687	24.49420
02_TRI_07	DRC-02	6	DRC	0.75754	24.49486
02_TRI_08	DRC-02	6	DRC	0.75739	24.49488
02_TRI_15	DRC-02	6	DRC	0.75885	24.49553
30_TRI_03	DRC-30	7	DRC	3.17679	19.80938
30_TRI_04	DRC-30	7	DRC	3.17714	19.80947
30_TRI_05	DRC-30	7	DRC	3.17746	19.80884
30_TRI_10	DRC-30	7	DRC	3.17920	19.81211
30_TRI_14	DRC-30	7	DRC	3.16384	19.81551
2-TRI 2	GH-02	8	Ghana	5.77024	2.56987

2-TRI 3	GH-02	8	Ghana	5.76752	2.57389
2-TRI 4	GH-02	8	Ghana	5.76114	2.58019
2-TRI 6	GH-02	8	Ghana	5.77607	2.62477
2-TRI 7	GH-02	8	Ghana	5.77576	2.62502
3-TRI 4*	GH-03	9	Ghana	6.03973	1.52491
3-TRI 5	GH-03	9	Ghana	6.03893	1.52369
3-TRI 7	GH-03	9	Ghana	6.03942	1.52443
3-TRI 8	GH-03	9	Ghana	6.04074	1.52797
3-TRI 11	GH-03	9	Ghana	6.05672	1.57200
NB628	GH-05	10	Ghana	6.691371	-1.318834
NB632	GH-05	10	Ghana	6.690194	-1.326118
NB627	GH-05	10	Ghana	6.691371	-1.318834

NB: * individuals that failed to amplify in the SEQUENOM MassArray

Two reduced representation libraries were used in this study, with both of them utilising the Cross et al (2015) AFLPseq Library prep protocol to generate a reference library of SNP markers. The specifications to the protocol are as follows: +2(CA)/+2(CA) selective base additions to the ligation adapters, size selection of pooled samples using E-Gel® (Life Technologies, Carlsbad, California), purification of selected product with AMPure™ XP (Agencourt, Beckman Coulter, Inc., Brea, California) and quantification using a 2200 TapeStation™ (Agilent, Santa Clara, California) with high-sensitivity ScreenTapes. The Emulsion PCR, enrichment and sequencing was undertaken on the Ion Torrent™ Platform (Life technologies) (as per Cross et al 2015) under manufacturers protocol and conditions.

Analysis of the initial data used the bioinformatics software program CLC-Bio (Qiagen). The raw sequences from the two runs were initially kept separate. Firstly the raw sequences were de-multiplexed according to their barcodes, with non-informative priming sequences, including barcodes and the original ligation adapters, trimmed from the reads. Once the excess sequence was trimmed from the reads, the two sequencing runs could be amalgamated. At this point, a *de novo* assembly was used to identify and remove any paralogous or duplicated loci present between the two runs and produce a single reference dataset. A second *de novo* assembly that discriminated each sample's raw reads against the reference dataset was then completed. By specifying a minimum coverage of reads per loci, per individual, only high coverage loci were used. Consensus sequences from each individual against the reference dataset were generated and mapped back to the reference dataset. The results were then exported into the software Geneious (Biomatters) where manual selection of appropriately variable loci was employed. Only loci that were suitable for use in the Sequenom MassArray iPLEX platform (Gabriel et al., 2009) were selected, which requires a single variable site located within ~100bp of non variable sequence. A BLAST search through GenBank of suitable loci was then done to detect contamination from other organisms. The final list of suitable loci was then sent to the Australian Genome Research Facility (AGRF) Brisbane node (University of Queensland, St Lucia, QLD, Australia) for primer design and genotyping.

The resulting data from the genotyping was stripped of non-informative samples and loci and then genetic diversity and differentiation statistics were calculated using GENODIVE (Meirmans and Van Tienderen, 2004). A STRUCTURE (Pritchard et al., 2000) analysis was performed to examine the number and significant of genetic clustering of populations. The number of cluster (K) was varied between K=1 to K=11, using default parameters, during a burnin period of 300,000 reps, 700,000 MCMC reps after burnin and five replicates for each K value. STRUCTURE HARVESTER (Earl, 2012) was then used to determine the most appropriate number of genetic clusters for the dataset.

Results

A total of 127 loci were identified as being suitable for Sequenom MassArray analysis. Of this, 107 loci could be used in the genotyping, with 96 of these producing results with high coverage across individuals. Of the 48 samples tested, only four failed to produce any genotyping results (marked with * in Table 1). A comparison of the Sequenom MassArray genotyping and sequencing results, using 17 loci and 22 individuals, found that only 4 discrepancies (1.13%) occurred between the two methods.

Calculation of genetic diversity (Table 2) in GENODIVE (Meirmans and Van Tienderen, 2004), as well as Analysis of MOlecular VAriance (both Stepwise Mutation Model (Table 3) and Infinite Allele Model (Table 4)), show that the results from the genotyping retain some genetic structure. The relatively high proportion of explained variance from *F_it* (Variance within individuals) (Table 4: IAM) could be explained by the low level of coverage from both within a population level as well as total number of populations used.

Table 2: Genetic Diversity Statistics from Sequenom MassArray genotyping

Statistic	Value	Std.Dev.	c.i.2.5%	c.i.97.5%	Description
Nm	1.979	0.015	1.948	2	Number of alleles
Eff_Nm	1.266	0.022	1.225	1.31	Effective number of alleles
Ho	0.203	0.016	0.173	0.234	Observed Heterozygosity
Hs	0.229	0.015	0.2	0.259	Heterozygosity Within Populations
Ht	0.249	0.017	0.217	0.281	Total Heterozygosity
H't	0.252	0.017	0.219	0.284	Corrected total Heterozygosity
Gis	0.115	0.039	0.04	0.191	Inbreeding coefficient

NB: Analysis done in Genodive (Meirmans and Van Tienderen, 2004)

Table 3: AMOVA Stepwise Mutation Model results for Sequenom MassArray genotyping

Source of Variation	%var	F-stat	F-value	Std.Dev.	c.i.2.5%	c.i.97.5%	P-value
Within Individual	0.862	R_it	0.138	0.039	0.065	0.216	--
Among Individual	0.08	R_is	0.085	0.038	0.013	0.16	0.001
Among Population	0.058	R_st	0.058	0.015	0.03	0.086	0.001

NB: Analysis done in Genodive (Meirmans and Van Tienderen, 2004) with 999 Permutations (reported statistics are equivalent to Rst).

Table 4: AMOVA Infinate Allele Model results from Sequenom MassArray genotyping

Source of Variation	%var	F-stat	F-value	Std.Dev.	c.i.2.5%	c.i.97.5%	P-value	F'-value
Within Individual	0.838	F_it	0.162	0.036	0.094	0.235	--	--
Among Individual	0.09	F_is	0.097	0.037	0.027	0.171	0.001	--
Among Population	0.072	F_st	0.072	0.011	0.05	0.094	0.001	0.094

NB: Analysis done in Genodive (Meirmans and Van Tienderen, 2004), with 999 Permutations.

A PCA by individuals (Figure 1) and by Populations (figure 2) show that there is some pattern emerging from the data. Whilst the relatively proportions explained by the individuals PCA data 10% for the first axis, 7% for the second axis, and 7% for the third axis are not especially high, some structural pattern can be detected. The population based PCA shows a greater level of explanation, with axis 1 (33%), axis 2 (19%) and axis 3 (12%), which is to be expected when amalgamating individuals together.

Figure 1: PCA output of individuals

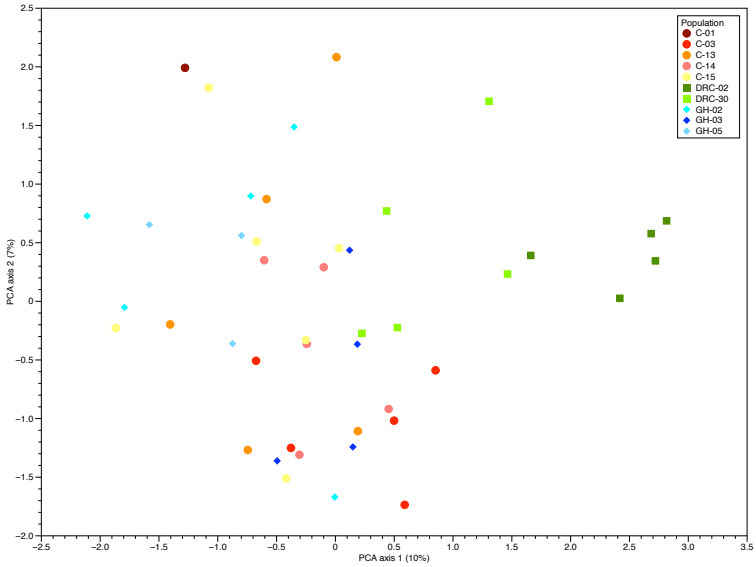
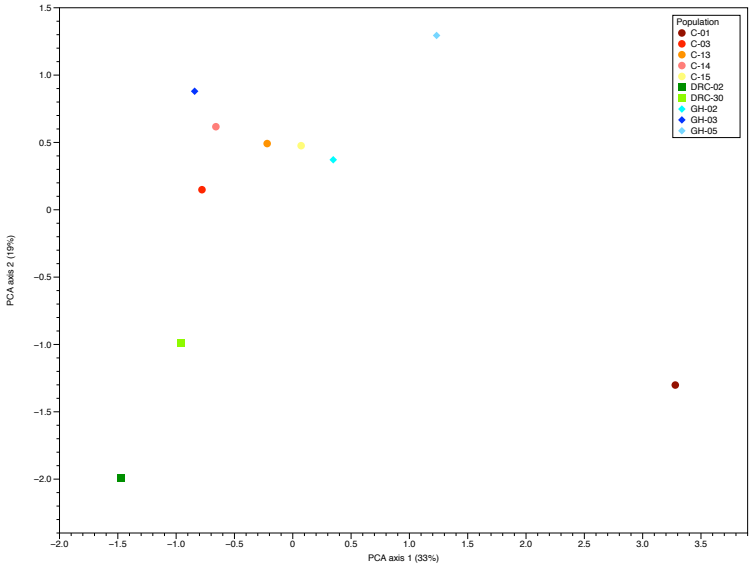


Figure 2: PCA output of Populations

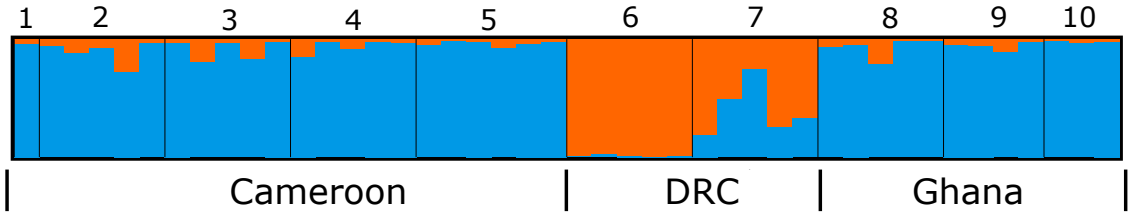


Using STRUCTURE HARVESTER (Earl, 2012), it was found that K=2 was the most likely number of populations for the genotyping results, with a Delta K score of 195.59, which was much greater than any of the other K value Delta K scores. The second most likely number of populations was K=3, with a Delta K score of 4.39. Yet this score was not much above the average score for all other K values. The bar graphs for K=2 (

Figure 3) and K=3 (

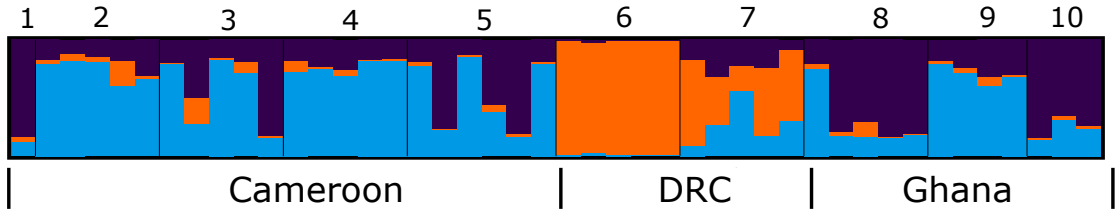
Figure 4) show the genetic differentiation of the two Eastern DRC populations (6 and 7) compared to the more Westerly populations of Cameroon (1-5) and Ghana (8-10). Therefore these initial findings support the accurate identification of samples from the broad geographical region sampled.

Figure 3: STRUCTURE bar graph of proportional population assignment when K=2



NB: graph produced using CLUMPAK (Kopelman et al., 2015) for all K values tested from STRUCTURE (Pritchard et al., 2000) analysis

Figure 4: STRUCTURE bar graph of proportional population assignment when K=3



NB: graph produced using CLUMPAK (Kopelman et al., 2015) for all K values tested from STRUCTURE (Pritchard et al., 2000) analysis

Discussion

This project was designed to develop a set of genetic markers that could be used routinely to identify Ayous samples of an unknown origin. Given that Ayous is one of the most economically important species in Africa, this research and the results found will greatly assist in using DNA based methods for origin identification purposes. The STRUCTURE results identified as K=2 (Figure 3) as the most likely number of populations for this dataset. The results identify that there is high levels of genetic distinction of the DRC populations in comparison to the other two countries, with the most easterly (DRC-02) being most dissimilar to all other populations. This genetic differentiation could be attributed to the presence of the Congo River Delta, a geographical barrier that has been previously known to restrict gene flow. The presence of small amounts of proportional assignment found in the Cameroon and Ghana populations (Orange sections in figures 3 and 4) show that a low level of gene flow exists across populations.. These results also show that the application of Next Generation Sequencing Techniques, in the development of a genographic map, of a relatively unknown species is appropriate. This pilot study has shown that even with a small number of individuals from a species that is distributed across a broad geographical area we were able to identify genetic structure in the data.

Conclusions/outlook

The congruence between the results obtained from the AFLP-seq procedure and Sequenom MassArray support the use of ALP-seq as a reliable and cost effective tool for genotyping. The conformity of the resulting genetic structure to the current theory of gene flow within Tropical West Africa further shows the promise of this technology and supports the ongoing roll out of this analysis across the entire distribution of Ayous. The expansion of this project will incorporate a much larger number of samples (~1000) with a focus on a larger number of loci, within population samples, and total number of populations. We expect the results from the extended study to provide a greater resolution of provenance identification for Ayous.

Acknowledgements

Associate Professor Jeremy Austin, Dr Korjent VanDijk, Dr Jillian Schedneck, Australian Research Council, ITTO, GTTN, Thünen Institute of Forest Genetics (Dr Celine Jolivet, Lasse Schindler, Vivian Kuhlenkamp)

References

1. Alaeddini, R., Walsh, S.J., and Abbas, A. (2010) Forensic implications of genetic analyses from degraded DNA—a review. *Forensic Science International: Genetics* 4, 148-157.
2. Bosu, P. and Krampah, E. (2005) Triplochiton scleroxylon K. *Schum. In. Louppe, D.(Eds). Prota* 7.
3. Cross, H.B., E; Van Dijk, KJ; Lowe, AJ; Waycott ,M (2015) AFLPseq. *(in Prep)*.
4. Degen, B. (2012) Practical applications of DNA technologies in timber identification and tracking. Forestry Research Network of Sub-Saharan Africa.
5. Degen, B. and Fladung, M. (2007) Use of DNA-markers for tracing illegal logging. In *Proceedings of the international workshop "Fingerprinting methods for the identification of timber origins" October*, pp. 8-9.
6. Degen, B., Ward, S.E., Lemes, M.R., Navarro, C., Cavers, S., and Sebbenn, A.M. (2013) Verifying the geographic origin of mahogany (Swietenia macrophylla King) with DNA-fingerprints. *Forensic Science International: Genetics* 7, 55-62.
7. Earl, D.A. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources* 4, 359-361.
8. Finkeldey, R., Rachmayanti, Y., Nuroniah, H., Nguyen, N.P., Cao, C., and Gailing, O. (2007) Identification of the timber origin of tropical species by molecular genetic markers–the case of dipterocarps. In *Proceeding of the International Workshop "Fingerprinting methods for the identification of timber origins". Bonn, Germany*, pp. 143-158.
9. Gabriel, S., Ziaugra, L., and Tabbaa, D. (2009) SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Current protocols in human genetics*, 2.12. 11-12.12. 16.
10. Gitzendanner, M.A. (2012) Use and guidelines for plant DNA analyses in forensics. In *Forensic Botany: A Practical Guide* (HAll, D.W.B., Jason H, ed), pp. 93-106, John Wiley & Sons.
11. Gugerli, F., Parducci, L., and Petit, R.J. (2005) Ancient plant DNA: review and prospects. *New Phytologist* 166, 409-418.
12. Hall, J.B. and Bada, S. (1979) The distribution and ecology of Obeche (Triplochiton Scleroxylon). *The Journal of Ecology*, 543-564.
13. Igboanugo, A. and Iversen, P. (2004) Conservation and use of Triplochiton scleroxylon in moist West Africa. *Forest Genetic Resources*, 50-52.
14. ITTO (2014) Annual Review Statiatics Database. http://www.itto.int/annual_review_output/?mode=searchdata.
15. Jolivet, C. and Degen, B. (2012) Use of DNA fingerprints to control the origin of sapelli timber (< i> Entandrophragma cylindricum</i>) at the forest concession level in Cameroon. *Forensic Science International: Genetics* 6, 487-493.
16. Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources*.
17. Lowe, A. (2007) Can we use DNA to identify the geographic origin of tropical timber. In *Proceedings of the international workshop "Fingerprinting methods for the identification of timber origins" October*, pp. 8-9.

18. Lowe, A.J. and Cross, H.B. (2011) The application of DNA methods to timber tracking and origin verification. *IAWA Journal-International Association of Wood Anatomists* 32, 251.
19. Meirmans, P.G. and Van Tienderen, P.H. (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* 4, 792-794.
20. Orwa, C., Mutua, A., Kindt, R., Jamnadass, R., and Simons, A. (2009) Agroforestree database: a tree reference and selection guide version 4.0.
21. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945-959.
22. Tnah, L.H., Lee, S.L., Ng, K.K., Tani, N., Bhassu, S., and Othman, R.Y. (2009) Geographical traceability of an important tropical timber (< i> Neobalanocarpus heimii</i>) inferred from chloroplast DNA. *Forest Ecology and Management* 258, 1918-1923.
23. Zahnen, J. (2007) Foreword from WWF-Germany. In *Proceedings of the international workshop "Fingerprinting methods for the identification of timber origins" October*, pp. 8-9.

APPENDIX V: A METHOD OF EXTRACTING PLANT NUCLEIC ACIDS FROM LIGNIFIED PLANT TISSUE

Patent

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 May 2015 (21.05.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/070279 A1

(51) International Patent Classification:
C12N 15/10 (2006.01)

(21) International Application Number:
PCT/AU2014/001047

(22) International Filing Date:
14 November 2014 (14.11.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2013904424 15 November 2013 (15.11.2013) AU

(71) Applicants: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, Massachusetts 02138 (US). **THE MINISTER FOR SUSTAINABILITY, ENVIRONMENT AND CONSERVATION, FOR AND ON BEHALF OF THE STATE OF SOUTH AUSTRALIA** [AU/AU]; Level 9, 91-97 Grenfell Street, Adelaide, South Australia 5000 (AU). **ADELAIDE RESEARCH & INNOVATION PTY LTD** [AU/AU]; Level 14, 115 Grenfell Street, Adelaide, South Australia 5000 (AU).

(72) Inventors: **LOWE, Andrew John**; 9 Melrose Avenue, Stirling, South Australia 5152 (AU). **JARDINE, Ducan Iain**; 6/50 Maud Street, Unley, South Australia 5061 (AU). **CROSS, Hugh Bradley**; Landaskollen 19, N-1430 As (NO). **DEGEN, Bernd**; Ahrensboeker Strasse 60, 23858 Reinfeld (DE). **SCHINDLER, Lasse**; Kamp 2b, 22941 Hammoor (DE). **HOLTKEN, Aki Michael**; Kuckuckskamp 2, 22927 Grosshandorf (DE).

(74) Agent: **WRAYS PTY LTD**; 56 Ord Street, West Perth, Western Australia 6005 (AU).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

WO 2015/070279 A1

(54) Title: A METHOD OF EXTRACTING PLANT NUCLEIC ACIDS FROM LIGNIFIED PLANT TISSUE

(57) Abstract: A method for extracting plant nucleic acids from lignified plant tissue comprising the steps of: a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids; b) purifying the nucleic acids extracted in (a).

A Method of Extracting Plant Nucleic Acids from Lignified Plant Tissue

TECHNICAL FIELD

[0001] A method for extracting plant nucleic acids from lignified plant tissue, and methods for using the extracted nucleic acids to determine the source of the lignified plant tissue.

BACKGROUND ART

[0002] The following discussion of the background art is intended to facilitate an understanding of the present invention only. The discussion is not an acknowledgement or admission that any of the material referred to is or was part of the common general knowledge as at the priority date of the application.

[0003] Extraction of high quality DNA or RNA from fresh plant leaf and bud tissue has become routine, rapid and can now be semi- or completely automated with robotic workstations and specialised extraction 'kits' (*e.g.* from Qiagen, or other manufacturers). DNA may also be routinely extracted from fungal cells using techniques such as that described in Wolfe et al (New Phytologist; 2010)

[0004] DNA extracted from freshly harvested wood, particularly if cambium tissue is incorporated, can also yield high quality DNA, which can be comparable to leaf sources (Colpaert *et al.* 2005). However, extraction of DNA from dried wood is more problematic, with lower yields farther from the cambium (Lowe and Cross 2011).

[0005] Overall, the challenges with using dried wood for genetic work are twofold: both the quantity and quality of the nucleic acids extracted are low. The number of cells in wood that can be expected to contain nucleic acids is low compared to leaves or buds, thus the quantity of nucleic acids is low. Nucleic acid degradation in plants is characterized by lengthwise fractures, with the result that only smaller fragments can be amplified and sequenced and, for example, DNA in wood is typically degraded (Lowe and Cross 2011). Some of the standard extraction kits available are not designed to recover these small fragments, and therefore these methods often will not work for extremely degraded tissues.

[0006] Other aspects of wood anatomy will affect the quality of the nucleic acids that are isolated from wood. In addition to the difficulties posed by macromolecules like lignin and cellulose, there are a wide-range of co-extracted compounds that can also inhibit the polymerase chain reaction (PCR) and other laboratory analyses. In one study, wood DNA extracts added to a PCR reaction with fresh leaf DNA showed significant inhibitory effects, which decreased with increasing dilution of the wood DNA sample (Rachmayanti *et al.* 2009). These experiments showed that the inhibitory effects of wood components on the nucleic acids that can be extracted were worse in the outer sapwood and not as bad in the inner heartwood. A range of approaches have improved the success rate for nucleic acid extraction from wood and limited the effects of inhibitors. Several additives can also be incorporated at the PCR step to improve amplification in the presence of some potentially inhibiting contaminants.

[0007] However, there does not seem to be any agreement as to the best approach, and results seem to vary widely. One factor may be that some woods are more amenable to nucleic acid extraction than others. Other studies have shown a great deal more success with extraction of nucleic acids from tropical woody species.

[0008] In general, these studies show that it is possible to extract nucleic acids such as DNA from wood using specifically designed laboratories, where contamination issues are minimized and a traceable process is implemented (Gugerli *et al.* 2005), and that success has been higher for non-tropical plants than for tropical species, but that the process is still not reliable. There is also still work to be done on how far down the timber treatment process (fresh wood, field-dried wood, sawn timber, kiln dried, or other processing) nucleic acids can be extracted.

[0009] It is against this background that the present description has been developed. It seeks to overcome, or at least ameliorate, one or more of the deficiencies of the prior art mentioned above to provide a nucleic acid extraction method that reliably works with a wide variety of lignified plant tissue from a range of plant species and on a range of timber products from along the processing chain (fresh timber, kiln dried timber and veneers), or to at least provide the consumer with a useful or commercial choice.

SUMMARY OF INVENTION

[0010] The present invention provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0011] The lignified plant tissue is preferably subjected to the extraction solution for between 0.5 and 16 hours, more preferably between 0.5 and 6 hours, between 1h and 6h, between 2h and 6h, between 3h and 6h, between 4h and 6h, or between 4.5 and 5.5 hours. Preferably, the lignified plant tissue is exposed to the solution for about 4h, 4.25h, 4.5h, 5h, 5.25 or 5.5 hours to enable extraction of the nucleic acids.

[0012] The exposure of the lignified plant tissue to the extraction solution should preferably occur at a temperature of between 40^0 and 65^0, between 45^0 and 65^0 between 47^0 and 60^0, between 49.5^0C and 60.5^0C, between 50^0C and 57^0C, or between 52^0C and 56^0C. Most preferably, the temperature of the solution during exposure of the lignified plant tissue to extract nucleic acids is about 52^0C, 53^0C, 54^0C, 55^0C, 56^0C or 57^0C.

[0013] The present invention also provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a chelating agent and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0014] The present invention also provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a salt and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0015] The present invention also provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, an excipient and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0016] The present invention also provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a biological antioxidant and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0017] The present invention also provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a serine protease and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0018] The present invention also provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a chelating agent, a salt, an excipient, a biological antioxidant, a serine protease and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0019] The present invention also provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: CTAB, Tris, NaCl,

EDTA, Boric Acid, PVP K30, DTT, and Proteinase K;

b) purifying the nucleic acids extracted in (a).

[0020] The present invention also provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: 55mM CTAB, 100mM Tris, 1.4M NaCl, 20mM EDTA, 1% (w/v) Boric Acid, 2% (w/v) PVP K30, 47mM DTT, and 1.86% Proteinase K;

b) purifying the nucleic acids extracted in (a).

[0021] The present invention also provides a method of determining the genus and/or species a sample of lignified plant tissue is from, comprising the steps of:

a) extracting plant nucleic acids from the lignified plant tissue using a method comprising the following steps:

i. exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids;

ii. purifying the nucleic acids extracted in (i);

b) using the nucleic acids from (a) to develop a genetic profile for the lignified plant tissue sample;

c) comparing the genetic profile from (b) with a library of genetic reference profiles.

[0022] The present invention also provides a method of determining the plant source of a sample of lignified plant tissue, comprising the following steps:

a) extracting plant nucleic acids from the lignified plant tissue using a method comprising the following steps:

i. exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids;

ii. purifying the plant nucleic acids extracted in (i);

b) using the plant nucleic acids from (a) to develop a genetic profile for the lignified plant tissue sample;

c) comparing the genetic profile from (b) with a library of genetic reference profiles.

[0023] The present invention also provides a method for developing a genetic profile for a sample of lignified plant tissue comprising the steps of:

a) extracting plant nucleic acids from lignified plant tissue using a method comprising the following steps:

i. exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids;

ii. purifying the nucleic acids extracted in (i);

b) subjecting the isolated plant nucleic acids of step (a) to a procedure chosen from the list comprising: PCR, RAPD analysis, AFLP analysis, CAPS analysis, RFLP analysis, SSR analysis, SNP analysis, SSCP analysis, sequencing.

[0024] The present invention also provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, boric acid and buffer;

b) instructions for use.

[0025] Preferably, the instructions for use of the kit designate that the extraction is carried out by exposing the lignified plant tissue to the extraction solution for about 5 hours. Preferably, the instructions for use of the kit designate that the extraction is carried out by exposing the lignified plant tissue to the extraction solution at a temperature of about 55^0C.

[0026] The present invention also provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, a chelating agent, boric acid and buffer;

b) instructions for use.

[0027] The present invention also provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, a salt, boric acid and buffer;

b) instructions for use.

[0028] The present invention also provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, an excipient, boric acid and buffer;

b) instructions for use.

[0029] The present invention also provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, a biological antioxidant, boric acid and buffer;

b) instructions for use.

[0030] The present invention also provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, a serine protease, boric acid and buffer;

b) instructions for use.

[0031] [The present invention also provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, boric acid, a buffer a chelating agent, a salt, an excipient, a biological antioxidant and a serine protease;

b) instructions for use.

[0032] The present invention also provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) CTAB, Tris, NaCl, EDTA, Boric Acid, PVP K30, DTT, Proteinase K;

b) instructions for use.

DETAILED DESCRIPTION

Detailed Description

[0033] It is often important to determine the source of lignified tissue such as timber, to answer questions including:

- Was this timber harvested from the declared country or region?
- Does this timber come from a conservation area?
- Does this timber come from natural forest or a plantation?
- Does this product come from the declared log or stump (is the chain-of-custody intact)?
- Has this log or lot been swapped with other plants of the same species (log laundering)?
- Is this timber of the declared species?
- Is this timber a Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) listed species or not?

Method of Extracting Nucleic Acids

[0034] In order to assist with determining the answer to question such as these, the present description provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0035] The lignified plant tissue is preferably subjected to the extraction solution for between 0.5 and 16 hours, more preferably between 0.5 and 6 hours, between 1h and 6h, between 2h and 6h, between 3h and 6h, between 4h and 6h, or between 4.5 and 5.5 hours. Preferably, the lignified plant tissue is exposed to the solution for about 4h, 4.25h, 4.5h, 5h, 5.25 or 5.5 hours to enable extraction of the nucleic acids.

[0036] The exposure of the lignified plant tissue to the extraction solution should preferably occur at a temperature of between 40^0 and 65^0, between 45^0 and 65^0 between 47^0 and 60^0, between 49.5^0C and 60.5^0C, between 50^0C and 57^0C, or between 52^0C and 56^0C. Most preferably, the temperature of the solution during exposure of the lignified plant tissue to extract nucleic acids is about 52^0C, 53^0C, 54^0C, 55^0C, 56^0C or 57^0C.

[0037] Without wishing to be bound by any particular theory, it is believed that the boric acid assists with the removal of carbohydrate impurities in the extraction solution.

[0038] The boric acid is preferably provided in the solution at between 0.1% (w/v) and 5% (w/v). Preferably, the boric acid is provided at between 0.25% and 4%, between 0.5% and 3%, between 0.5% (w/v) and 2% (w/v), or between 0.9% (w/v) and 1.1% (w/v). Most preferably, the boric acid is provided in the solution at 0.5%, 0.75%, 1%, 1.25% or 1.5% (w/v).

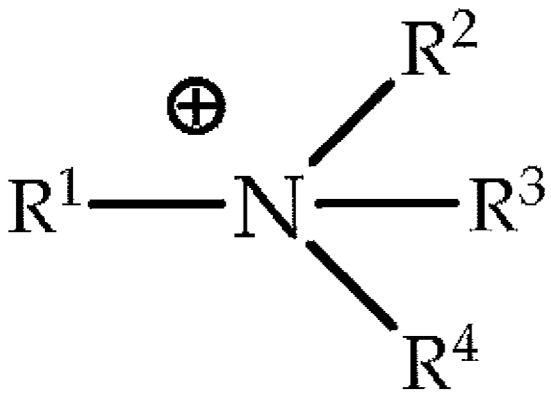
[0039] The cationic surfactant used in the method of the present invention functions to at least collate and emulsify the lipids in the reaction mixture. Any cationic surfactant that can perform at least this function may be used. For example, the cationic surfactant may be chosen from the list comprising: cetyltrimethylammonium bromide (CTAB), centrimonium chloride, benzalkonium chloride, and benzethonium chloride.

[0040] The cationic surfactant is preferably provided in the solution at between 10mM and 100mM. Preferably, the cationic surfactant is provided at between 20mM and 80mM, between 30mM and 70mM, between 45mM and 65mM, or between 49.5mM

and 60.5mM. Most preferably, the cationic surfactant is provided in the solution at 45mM, 50mM, 55mM, 60mM or 65mM.

[0041] As referred to herein, the term "cationic surfactant" refers to any surfactant molecule comprising a hydrophobic "tail" group and a hydrophilic "head" group, wherein the hydrophilic head group comprises a positive charge.

[0042] In some embodiments, the cationic surfactant comprises a quaternary ammonium cation as the head group. "Quaternary ammonium cations" are positively charged ions of the structure:



(I)

wherein each of R^1, R^2, R^3 and R^4 are independently selected from H, alkyl or aryl groups, with the proviso that at least one of R^1, R^2, R^3 and R^4 comprises a hydrophobic alkyl or aryl group.

[0043] "Alkyl" as a group or part of a group denotes an optionally substituted straight or branched aliphatic hydrocarbon group. The group may be a terminal group or a bridging group.

[0044] "Aryl" as a group or part of a group denotes (i) an optionally substituted monocyclic, or fused polycyclic, aromatic carbocycle (ring structure having ring atoms that are all carbon) preferably having from 5 to 18 atoms per ring. Examples of aryl groups include optionally substituted phenyl, optionally substituted naphthyl, and the like; (ii) an optionally substituted partially saturated bicyclic aromatic

carbocyclic moiety in which a phenyl and a C_{5-7} cycloalkyl or C_{5-7} cycloalkenyl group are fused together to form a cyclic structure, such as tetrahydronaphthyl, indenyl or indanyl.

[0045] The term "optionally substituted" as used throughout the specification denotes that the group may or may not be further substituted or fused with one or more non-hydrogen substituent groups. In certain embodiments the substituent groups are one or more groups independently selected from the group consisting of halogen, =O, =S, -CN, $-NO_2$, $-CF_3$, $-OCF_3$, alkyl, alkenyl, alkynyl, haloalkyl, haloalkenyl, haloalkynyl, heteroalkyl, cycloalkyl, cycloalkenyl, heterocycloalkyl, heterocycloalkenyl, aryl, heteroaryl, cycloalkylalkyl, heterocycloalkylalkyl, heteroarylalkyl, arylalkyl, cycloalkylalkenyl, heterocycloalkylalkenyl, arylalkenyl, heteroarylalkenyl, cycloalkylheteroalkyl, heterocycloalkylheteroalkyl, arylheteroalkyl, heteroarylheteroalkyl, hydroxy, hydroxyalkyl, alkyloxy, alkyloxyalkyl, alkyloxycycloalkyl, alkyloxyheterocycloalkyl, alkyloxyaryl, alkyloxyheteroaryl, alkyloxycarbonyl, alkylaminocarbonyl, alkenyloxy, alkynyloxy, cycloalkyloxy, cycloalkenyloxy, heterocycloalkyloxy, heterocycloalkenyloxy, aryloxy, phenoxy, benzyloxy, heteroaryloxy, arylalkyloxy, amino, alkylamino, acylamino, aminoalkyl, arylamino, sulfonylamino, sulfinylamino, sulfonyl, alkylsulfonyl, arylsulfonyl, aminosulfonyl, sulfinyl, alkylsulfinyl, arylsulfinyl, aminosulfinylaminoalkyl, C(O)OH, $-C(=O)R^a$, $-C(=O)OR^a$, $C(=O)NR^aR^b$, $C(=NOH)R^a$, $C(=NR^a)NR^bR^c$, NR^aR^b, $NR^aC(=O)R^b$, $NR^aC(=O)OR^b$, $NR^aC(=O)NR^bR^c$, $NR^aC(=NR^b)NR^cR^d$, $NR^aSO_2R^b$,$-SR^a$, $SO_2NR^aR^b$, $-OR^a$, $OC(=O)NR^aR^b$, $OC(=O)R^a$ and acyl, wherein R^a, R^b, R^c and R^d are each independently selected from the group consisting of H, C_1-C_{12}alkyl, C_1-C_{12}haloalkyl, C_2-C_{12}alkenyl, C_2-C_{12}alkynyl, C_2-C_{10}heteroalkyl, C_3-C_{12}cycloalkyl, C_3-C_{12}cycloalkenyl, C_2-C_{12}heterocycloalkyl, C_2-C_{12}heterocycloalkenyl, C_6-C_{18}aryl, C_1-C_{18}heteroaryl, and acyl, or any two or more of R^a, R^b, R^c and R^d, when taken together with the atoms to which they are attached form a heterocyclic ring system with 3 to 12 ring atoms.

[0046] In some embodiments each optional substituent is independently selected from the group consisting of: halogen, =O, =S, -CN, $-NO_2$, $-CF_3$, $-OCF_3$, alkyl, alkenyl, alkynyl, haloalkyl, haloalkenyl, haloalkynyl, heteroalkyl, cycloalkyl, cycloalkenyl, heterocycloalkyl, heterocycloalkenyl, aryl, heteroaryl, hydroxy,

hydroxyalkyl, alkyloxy, alkyloxyalkyl, alkyloxyaryl, alkyloxyheteroaryl, alkenyloxy, alkynyloxy, cycloalkyloxy, cycloalkenyloxy, heterocycloalkyloxy, heterocycloalkenyloxy, aryloxy, heteroaryloxy, arylalkyl, heteroarylalkyl, arylalkyloxy, amino, alkylamino, acylamino, aminoalkyl, arylamino, sulfonyl, alkylsulfonyl, arylsulfonyl, aminosulfonyl, aminoalkyl, -COOH, -SH, and acyl.

[0047] In some embodiments, at least one, at least two, at least three or each of R^1, R^2, R^3 and R^4 is alkyl.

[0048] Quaternary ammonium cations may be permanently charged, independent of the pH of their solution. Quaternary ammonium salts or quaternary ammonium compounds as referred to herein may include salts of quaternary ammonium cations with an anion.

[0049] As set out above, cationic surfactants, including those comprising a quaternary ammonium cation, include one or more hydrocarbon tail groups. Typically, the hydrocarbon tail is of sufficient length to be lipophilic or hydrophobic. In some embodiments, the one or more hydrocarbon tail groups of the cationic surfactant, ie. one or more of R^1, R^2, R^3 and R^4 in structure (I), may comprise an alkyl chain at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 carbons in length.

[0050] The buffer used in the method of the present invention functions to at least maintain the pH of the solution during extraction of the nucleic acids. Any buffer that can perform at least this function may be used. For example, the buffer may be chosen from the list comprising: tris(hydroxymethyl)aminomethane (Tris), piperazine-N,N'-bis(2-ethanesulfonic acid) (PIPES), (M)ethanosulfonic acid and succonic acid.

[0051] The buffer is preferably provided in the solution at between 50mM and 150mM. Preferably, the buffer is provided at between 60mM and 140mM, between 70mM and 130mM, between 80mM and 120mM or between 90mM and 110mM. Most preferably, the buffer is provided in the solution at 70mM, 80mM, 90mM, 100mM, 110mM, 120mM or 130mM.

[0052] The solution that the lignified plant tissue is exposed to in the method of the present invention may additionally comprise a chelating agent for binding metal ions such as magnesium and/or calcium ions. Preferably, the chelating agent chelates

magnesium and/or calcium ions to reduce their ability to act as a co-factor to endogenous nucleases that may cause damage to the extracted nucleic acids. Any chelating agent that can perform at least this function may be used. For example, the chelating agent may be chosen from the list comprising: ethylene diamine tetraacetic acid (EDTA), ethylene glycol tetraacetic acid (EGTA), ethylenediamine-N,N'-disuccinic acid (EDDS) and diethylene triamine pentaacetic acid (DTPA).

[0053] The chelating agent is preferably provided in the solution at between 1mM and 50mM. Preferably, the chelating agent is provided at between 5mM and 40mM, between 10mM and 30mM, between 15mM and 25mM or between 18mM and 22mM. Most preferably, the chelating agent is provided in the solution at 10mM, 12mM, 15mM, 17mM, 20mM, 22mM, 25mM or 27mM.

[0054] The invention thus provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a chelating agent and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0055] The solution that the lignified plant tissue is exposed to in the method of the present invention may additionally comprise a salt for maintaining the osmolarity of the solution and assist in nucleic acid stabilisation. Any salt that can perform at least this function may be used. For example, the salt may be chosen from the list comprising: sodium chloride, sodium citrate.

[0056] The salt is preferably provided in the solution at between 0.5M and 3mM. Preferably, the salt is provided at between 0.5M and 2.0M, between 1M and 2M, between 1.26M and 1.54M or between 1M and 1.4M. Most preferably, the salt is provided in the solution at 1.4M.

[0057] The invention thus provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a salt and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0058] The solution that the lignified plant tissue is exposed to in the method of the present invention may additionally comprise an excipient to allow the components of the solution to react with the liginifed tissue and enhance precipitation of components other than the nucleic acids such as phenolic compounds. The excipient used in the present method should preferably form complex hydrogen bonds with phenolics and co-precipitates with cell debris upon cell lysis. Any excipient that can perform at least this function may be used. For example, the excipient may be chosen from the list comprising: polyvinylpyrrolidone (PVP including PVP 360, PVP K30, PVP-40); polyvinylpolypyrrolidone (PVPP) and polyethylene glycol (PEG 200).

[0059] The excipient is preferably provided in the solution at between 0.5% and 4% (w/v). Preferably, the excipient is provided at between 1% (w/v) and 3% (w/v), or between 1.8% (w/v) and 2.2% (w/v). Most preferably, the excipient is provided in the solution at 1%, 1.5%, 1.75%, 2%, 2.25%, 2.5% or 3.0% (w/v).

[0060] The invention thus provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, an excipient and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0061] The solution that the lignified plant tissue is exposed to in the method of the present invention may additionally comprise a biological antioxidant which will serve to reduce the disulfide bonds of at least the nucleic acids in the solution and/or to reduce the tannins and other polyphenols present in the extraction mixture. Any biological antioxidant that can perform at least this function may be used. For example, the biological antioxidant may be chosen from the list comprising: β-mercaptoethanol, dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP), ascorbic acid, and sodium ascorbate.

[0062] The biological antioxidant is preferably provided in the solution at between 5 and 100mM, or between 5 and 75mM. Preferably, the biological antioxidant is provided at between 20mM and 70mM, between 30mM and 60mM, 30mM and 65mM, or between 42.3mM and 51.7mM. Most preferably, the biological antioxidant is provided in the solution at 35mM, 37mM, 40mM, 45mM, 47mM, 50mM, 52mM or 55mM.

[0063] The invention thus provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a biological antioxidant and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0064] The solution that the lignified plant tissue is exposed to in the method of the present invention may additionally comprise a protease to assist in the breakdown of cellular wall material in the lignified plant tissue and inactivate nucleases.

[0065] Preferably, the protease in the solution that the lignified plant tissue is exposed to in the method of the present invention is a serine protease. Any serine protease that can perform to assist in the breakdown of cellular wall material in the lignified plant tissue and inactivate nucleases may be used. For example, the serine protease may be chosen from the list comprising: proteinase K, OB protease.

[0066] Alternatively, the solution that the lignified plant tissue is exposed to in the method of the present invention may additionally comprise a cysteine protease to assist in the breakdown of cellular wall material in the lignified plant tissue and inactivate nucleases. Any cysteine protease that can perform at least this function may be used. For example, the cysteine protease may be chosen from the list comprising: brofasin, cysteine proteinase. In another aspect of the invention, the solution that the lignified plant tissue is exposed to in the method of the present invention may additionally comprise a broad specificity protease such as nagarse,

[0067] The serine protease is preferably provided in the solution at between 0.1% and 3.5%. Preferably, the serine protease is provided at between 0.5% and 3.0%, between 1.0% and 2.4%, or between 1.67% and 2.05%. Most preferably, the biological antioxidant is provided in the solution at 1.25%, 1.5%, 1.75%, 1.86%, 1.95%, 2.0%, 2.15%, 2.25% or 2.5%.

[0068] The invention thus provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a serine protease and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0069] Alternatively, the solution that the lignified plant tissue is exposed to in the method of the present invention may additionally comprise two or more of the additional agents listed above. Therefore, the solution that the lignified plant tissue is exposed to may comprise (in addition to a cationic surfactant, boric acid and buffer) two or more of the following: a chelating agent, a salt, an excipient, a biological antioxidant, and a serine protease.

[0070] In one embodiment of the invention, the invention provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a chelating agent, a salt, an excipient, a biological antioxidant, a serine protease and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

[0071] The solution may comprise additional reagents such as RNAse, DNAse, or other reagents depending on the nature of the nucleic acids it is desired to extract.

[0072] Once the nucleic acids have been extracted from the lignified plant tissue, the technique of purification of those nucleic acids is dependent on the preference of the investigator. Any technique, from caesium chloride gradients to proprietary kits may

be used in this step of the method. For example, the extracted nucleic acids may be purified with chloroform or a phenol-chloroform mixture, followed by precipitation an alcohol such as ethanol or isopropanol. Alternatively, the nucleic acids may be precipitated with sodium or ammonium acetate.

[0073] In a preferred embodiment, the extracted nucleic acids are purified using more than one technique sequentially. For example, the extracted nucleic acids may be purified using chloroform/isoamyl alcohol, followed by precipitation using isopropanol (preferably in the presence of glycogen), followed by purification using a proprietary kit. If precipitation is carried out with isopropanol, this step is preferably carried out in the presence of glycogen either overnight at about -20^{0}C, or at -80^0 for between 0.5h and 2h, preferably between 0.5 and 1h, most preferably for 1h. Alternatively, the isopropanol precipitation in the presence of glycogen could be carried out overnight at -80^{0}C, or for 2h, 3h, 4h, 5h, 6h, 7h, 8h, 9h,12h etc. The isopropanol precipitation could also be carried out at -20^{0}C for 2h, 3h, 4h, 5h, 6h, 7h, 8h, 9h,12h etc.

[0074] Preferably the nucleic acids are purified using the proprietary kit MO BIO Ultra Clean® 15 DNA Purification Kit [catalogue #:12100; MO BIO Laboratories, Inc., Carlsbad, USA]. However, other DNA extraction kits from a variety of sources can be used in the present method.

[0075] The nucleic acids extracted using the method of the present description may be DNA or RNA, from either the genome of the cells of the lignified plant tissue, or from chloroplasts or mitochondria.

[0076] In one aspect of the invention, the lignified plant tissue is exposed to a solution comprising: CTAB, Tris, NaCl, EDTA, Boric Acid, PVP K30, DTT,and Proteinase K.

[0077] In one embodiment of the invention, the invention provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: CTAB, Tris, NaCl, EDTA, Boric Acid, PVP K30, DTT, and Proteinase K;

b) purifying the nucleic acids extracted in (a).

[0078] In one example of the invention, the lignified plant tissue is exposed to a solution comprising 55mM CTAB, 100mM Tris, 1.4M NaCl, 20mM EDTA, 1% (w/v) Boric Acid, 2% (w/v) PVP K30, 47mM DTT; and 1.86% Proteinase K.

[0079] The invention therefore provides a method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: 55mM CTAB, 100mM Tris, 1.4M NaCl, 20mM EDTA, 1% (w/v) Boric Acid, 2% (w/v) PVP K30,47mM DTT, and 1.86% Proteinase K;

b) purifying the nucleic acids extracted in (a).

[0080] Before the method of nucleic extraction is performed on the lignified plant tissue, the tissue may be exposed to a sample preparation method that breaks down at least part of the structure of the lignified plant tissue. For example, the tissue may be flaked, diced, cut or grated, and/or ground with a mortar and pestle, or exposed to a ball mill. The sample preparation method may be carried out at room temperatures, or may be performed at low temperatures such as in the presence of dry ice or liquid nitrogen. It is believed that low temperatures such as those provided by exposing the tissue to liquid nitrogen may help preserve the nucleic acids during extraction.

[0081] In the present specification, the term "lignified plant tissue" is understood to encompass plant tissue comprising lignin. For example, the lignified plant tissue may be chosen from tissue such as stems, roots, seeds, seed coats, flower buds, leaf buds, wood and timber.

[0082] In the present specification, the term "trees" is understood to encompass bushes, shrubs and other sources of lignified plant tissue.

[0083] The terms "wood" and "timber" encompasses any hard, fibrous structural tissue containing lignin found in the stems and roots of trees and other woody plants. This includes heartwood and sapwood. The wood or timber maybe from hardwood trees or softwood trees. The terms "wood" and "timber" also encompasses

manufactured or engineered wood including glued laminated timber (glulam), wood structural panels (including plywood, oriented strand board and composite panels), laminated veneer lumber (LVL) and other structural composite lumber (SCL) products, parallel strand lumber, I-joists, chipboard, hardboard, and medium-density fiberboard (MDF).

[0084] Preferably, the lignified plant tissue is wood or hard tissue such as stems, roots, seeds, seed coats, flower buds and leaf buds from sources such as trees, shrubs and bushes. Alternatively, the lignified plant tissue may be wood from dried sources such as dried wood, furniture, wooden barrels, wood pulp, plywood, laminated wood, lumber or logs.

Use of Extracted Nucleic Acids

[0085] It is known to use nucleic acid-based techniques to investigate natural variation within populations of plants to identify genus and species (DNA barcoding). However, difficulties in extracting suitable nucleic acid material from lignified plant tissue renders these processes less effective when applied to timber products, particularly dried timber products. Therefore, the present method of nucleic acid extraction may be used to assist in the determination of the genus and/or species of the source tree that the lignified plant tissue comes from.

[0086] There is therefore provide a method of determining the genus and/or species a sample of lignified plant tissue is from, comprising the steps of:

a) extracting plant nucleic acids from the lignified plant tissue using a method comprising the following steps:

   i) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids;

   ii) purifying the nucleic acids extracted in (i);

b) using the nucleic acids from (a) to develop a genetic profile for the lignified plant tissue sample;

c) comparing the genetic profile from (b) with a library of genetic reference profiles.

[0087] It is known to use the natural variation within populations of plants to determine whether that tree comes from a specific geographical location or population of plants using nucleic acid-based techniques. The nucleic acid based techniques allow the determination of whether the lignified plant tissue comes from sustainable, or at least legal, sources and/or whether the lignified plant tissue comes from the geographical source the provider has declared it originates from. It also allows the determination of whether the lignified plant tissue comes from the declared log or stump that the provider has been certified the tissue to come from. Thus, there are a number of different genetic methods to be used for verification of source, both at the regional (phylogeography) and concession scales (population genetic assignment), and for tracking individual logs or timber products (DNA fingerprinting). However, difficulties in extracting suitable nucleic acid material from lignified plant tissue renders these processes less effective when applied to timber products, particularly dried timber products. Therefore, the present method of nucleic acid extraction may be used to assist in the determination of the geographical location of the source tree that the lignified plant tissue comes from.

[0088] There is therefore provide a method of determining the plant source of a sample of lignified plant tissue, comprising the following steps:

a) extracting plant nucleic acids from the lignified plant tissue using a method comprising the following steps:

i) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids;

ii) purifying the nucleic acids extracted in (i);

b) using the nucleic acids from (a) to develop a genetic profile for the lignified plant tissue sample;

c) comparing the genetic profile from (b) with a library of genetic reference profiles.

[0089] The genetic profile may be developed for a sample of lignified plant tissue by:

a) extracting plant nucleic acids from lignified plant tissue using the method described above; and

b) subjecting the isolated nucleic acids of step (a) to a procedure that allows determination of genomic, mitochondrial or chloroplast molecular markers.

[0090] The molecular markers may be chosen from the list comprising: randomly amplified polymorphic DNA (RAPD); amplified fragment length polymorphisms (AFLP); restriction fragment length polymorphisms (RFLP); cleaved amplified polymorphic sequences (CAPS), simple sequence repeats/microsatellites (SSR), single nucleotide repeats (SNP), single strand conformational polymorphisms (SSCP). Alternatively, other molecular markers known to the skilled reader may be used in the present method. The nucleic acids may be subject to sequencing, and/or PCR in its various forms (such as RT-PCR, Allele-specific PCR, Dial-out PCR, Asymmetric PCR, Intersequence-specific PCR, RACE LaNe RAGE etc).

[0091] There is therefore provided a method for developing a genetic profile for a sample of lignified plant tissue comprising the steps of:

a) extracting plant nucleic acids from lignified plant tissue using a method comprising the following steps:

   i) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids;

   ii) purifying the nucleic acids extracted in (i);

b) subjecting the isolated nucleic acids of step (a) to a procedure chosen from the list comprising: PCR, RAPD analysis, AFLP analysis, CAPS analysis, RFLP analysis, SSR analysis, SNP analysis, SSCP analysis, sequencing.

[0092] When carrying out the method of: determining the genus and/or species a sample of lignified plant tissue is from; the plant source of a sample of lignified plant tissue; or developing a genetic profile for a sample of lignified plant tissue, the lignified plant tissue may be exposed to one or more of the following (in addition to a

cationic surfactant, boric acid and buffer): a chelating agent, a salt, an excipient, a biological antioxidant, or a serine protease.

[0093] In one aspect of the invention, the lignified plant tissue is exposed to a solution comprising: CTAB, Tris, NaCl, EDTA, Boric Acid, PVP K30, DTT, and Proteinase K.

[0094] In one example of the invention, the lignified plant tissue is exposed to a solution comprising: 55mM CTAB, 100mM Tris, 1.4M NaCl, 20mM EDTA, 1% (w/v) Boric Acid, 2% (w/v) PVP K30, 47mM DTT, and 1.86% Proteinase K.

Kits

[0095] The present description further provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, boric acid and buffer;

b) instructions for use.

[0096] Preferably, the instructions for use designate that the extraction is carried out by exposing the lignified plant tissue to the extraction solution for between 0.5 and 16 hours, more preferably between 0.5 and 6 hours, between 1h and 6h, between 2h and 6h, between 3h and 6h, between 4h and 6h, or between 4.5 and 5.5 hours. Preferably, the lignified plant tissue is exposed to the solution for about 4h, 4.25h, 4.5h, 5h, 5.25 or 5.5 hours to enable extraction of the nucleic acids.

[0097] Preferably, the instructions for use designate that the extraction is carried out by exposing the lignified plant tissue to the extraction solution at a temperature of between 40^0 and 65^0, between 45^0 and 65^0 between 47^0 and 60^0, between 49.5^0C and 60.5^0C, between 50^0C and 57^0C, or between 52^0C and 56^0C. Most preferably, the temperature of the solution during exposure of the lignified plant tissue to extract nucleic acids is about 52^0C, 53^0C, 54^0C, 55^0C, 56^0C or 57^0C.

[0098] The present description further provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, a chelating agent, boric acid and buffer;

b) instructions for use.

[0099] The present description further provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, a salt, boric acid and buffer;

b) instructions for use.

[00100] The present description further provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, an excipient, boric acid and buffer;

b) instructions for use.

[00101] The present description further provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, a biological antioxidant, boric acid and buffer;

b) instructions for use.

[00102] The present description further provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, a serine protease, boric acid and buffer;

b) instructions for use.

[00103] The present description further provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, boric acid, a buffer a chelating agent, a salt, an excipient, a biological antioxidant and a serine protease;

b) instructions for use.

[00104] The present description further provides a kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) CTAB, Tris, NaCl, EDTA, Boric Acid, PVP K30, DTT, Proteinase K;

b) instructions for use.

General

[00105] Those skilled in the art will appreciate that the description herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the description includes all such variations and modifications. The description also includes all of the steps, features, compositions and compounds referred to or indicated in the specification, individually or collectively and any and all combinations or any two or more of the steps or features.

[00106] The present description is not to be limited in scope by the specific embodiments described herein, which are intended for the purpose of exemplification only. Functionally equivalent products, compositions and methods are clearly within the scope of the description as described herein.

[00107] The entire disclosures of all publications (including patents, patent applications, journal articles, laboratory manuals, books, or other documents) cited herein are hereby incorporated by reference. No admission is made that any of the references constitute prior art or are part of the common general knowledge of those working in the field to which this invention relates.

[00108] Each document, reference, patent application or patent cited in this text is expressly incorporated herein in their entirety by reference, which means that it should be read and considered by the reader as part of this text. That the document, reference, patent application or patent cited in this text is not repeated in this text is merely for reasons of conciseness.

[00109] Any manufacturer's instructions, descriptions, product specifications, and product sheets for any products mentioned herein or in any document incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention.

[00110] As used herein the term "derived" and "derived from" shall be taken to indicate that a specific integer may be obtained from a particular source albeit not necessarily directly from that source.

[00111] As used herein, the singular forms "a," "an" and "the" include plural references unless the context clearly dictates otherwise. Thus, for example, reference to a terpene synthase that catalyzes the formation of a terpene includes synthases that catalyze the productions of one or a plurality of terpenes.

[00112] Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

[00113] Other than in the operating example, or where otherwise indicated, all numbers expressing quantities of ingredients, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about". Accordingly, unless indicated to the contrary, the numerical parameters set forth in the specification and claims are approximations that may vary depending upon the desired properties sought to be obtained by the present invention. Hence "about 80 %" means "about 80 %" and also "80 %". At the very least, each numerical parameter should be construed in light of the number of significant digits and ordinary rounding approaches.

[00114] Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the description are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value; however, inherently contains certain errors necessarily resulting from the standard deviation found in their respective testing measurements

[00115] Other definitions for selected terms used herein may be found within the detailed description and apply throughout. Unless otherwise defined, all other scientific and technical terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which the description belongs.

[00116] The following examples serve to more fully describe the manner of using the invention, as well as to set forth the best modes contemplated for carrying out various aspects of the invention. It is understood that these methods in no way serve to limit the true scope of this invention, but rather are presented for illustrative purposes.

EXAMPLES

Example 1

Method

Sample preparation

1. Remove the outer part of the timber sample (removal of contaminated tissue)
2. Make small (curly) flakes with a clean scalpel
3. Fill a 2mL microcentrifuge tube about 1/2 to 2/3 of the volume with flakes
4. Add 2 clean steel balls
5. Freeze the sample in liquid nitrogen
6. Homogenate the sample with a Bead Beater (Retsch® mixer mill) at 17.5s^{-1} for 100 Seconds
7. Repeat steps 5 and 6 once (or twice) more until fine sawdust can be seen
8. Remove the steel balls

Extraction

1. Add 900µL BoTab buffer, 45µL DTT^ (1M) and 18µL Proteinase K and mix well (vortex)
2. Incubate for 5 hours at 55°C (mixing occasionally)

Purification

1. Add 450µL Chloroform/Isoamyl alcohol (24:1 v/v) and mix (vortex) the sample
2. Incubate for 10 minutes at room temperature, mixing regularly
3. Centrifuge for 15 minutes at 16 200g (4°C)
4. Prepare a 1.5mL microcentrifuge tube with 225µL 3M NaOAc (pH 5.2)

5. Add 600µL of the aqueous phase to the NaOAc
6. Add 600µL Isopropanol and 0.5-4µL Glycogen
7. Incubate at -20°C overnight
8. Centrifuge for 15 minutes at 16 200g (4°C)
9. Discard the supernatant
10. Add 300µL 70% EtOH
11. Centrifuge for 10 minutes at 16 200g (4°C)
12. Discard the supernatant
13. Add 25µL H_2O to the pellet (do not dry the sample)

^ DTT was substituted for β-Mercapto-ethanol because of its lower toxicity

Cleaning DNA using MO BIO UltraClean® 15 DNA Purification Kit

1. Add 90-100µL **ULTRA SALT** (3 times the volume of the sample)
2. Add 5.8µL **ULTRA BIND** (homogenated) and mix well
3. Incubate for 5-10 minutes at room temperature
4. Centrifuge for 5 seconds at 10 000g (room temperature)
5. Discard the supernatant
6. Add 900µL **ULTRA WASH** and mix well
7. Centrifuge for 5 seconds at 10 000g (room temperature)
8. Discard the supernatant
9. Centrifuge for 5 seconds at 10 000g (room temperature)
10. Remove the supernatant
11. Add 23µL of H_2O and mix well
12. Incubate for ~5 minutes at room temperature
13. Centrifuge for 1 minute at 10 000g (room temperature) - the supernatant contains the DNA and can be diluted approximately 1:10

NB: The pellet from step 13 is unstable, use supernatant quickly. Rerun the centrifuge every two samples if extracting multiple samples of supernatant.

[MO BIO Ultra Clean® 15 DNA Purification Kit (catalogue #:12100); MO BIO Laboratories, Inc., Carlsbad, USA].

BoTab buffer (for 250mL):

Component	Mass (g)	Final concentration
CTAB	5.01	55 mM
Tris	3.03	100 mM
NaCl	20.45	1.4 M
EDTA	1.46	20 mM
Boric Acid	2.5	1% (w/v)
PVP K30	5.0	2% (w/v)

Results

[00117] To evaluate the success of the mBO extraction method it has been compared with the most commonly used plant DNA extraction protocol (CTAB) as well as five selected commercial kits:

- *Nucleon Phytopure Genomic DNA Extraction kit* (NPh)
- *DNeasy® Plant mini Kit* (DNP)
- *GenElute™ Plant Genomic DNA Miniprep Kit* (GEP)
- *innuPREP Plant DNA Kit (iPP)*
- *sbeadex mini plant kit* (smp)

[00118] These kits were chosen because of their successful application by other groups working with plant material:

[00119] Samples of saw timber of four individuals for each of the four tree species *Swietenia sp.* (Sw1, Sw2, Sw3, Sw4), *Intsia sp.* (In1, In2, In3, In4), *Quercus sp.* (Qu1, Qu2, Qu3, Qu4) and *Larix sp.* (La1, La2, La3, La4) were used in the present test.

[00120] For each sample, chloroplast fragments of a size between 120-300bp were amplified.

[00121] The success of the PCR and thus the DNA extraction was checked by agarose-gel electrophoresis. A strong PCR band was classified as 1 point, a weak PCR band by a 0.5 and no PCR product by 0 points. For each extraction method the sum of values was calculated as an overall indicator of the quality of the DNA extraction protocol. Three tests could not been done because of lack of chemicals (*).

Probe	mBo	CTAB	NPh	GEP	smp	iPP	DNP
Sw1	0	0	0	0	0	1	0
Sw2	0	0.5	0	0	0	1	0
Sw3	1	0.5	0.5	0	0	1	1
Sw4	1	0.5	1	1	1	1	1
In1	1	1	0	0	0	0	0
In2	0.5	0	0	0	0	0	0
In3	0.5	0	0	0	0	0	0
In4	1	1	0	0	1	0	0.5
Qu1	1	0	0	0	0.5	0.5	1
Qu2	0	0	0	0	0	0	0
Qu3	1	0	0	0	0	1	1
Qu4	1	0	0	0	1	0	0
La1	1	0	0	0	1	1	1
La2	1	0	0	*	*	1	1
La3	1	0.5	0	0	*	1	0
La4	1	1	0	1	1	0.5	1
Sum	12	5	1,5	(2/15)	(5,5/14)	9	7,5

[00122] The mBo extraction method was the most successful, with a success rate of 75%. The commercial kits, as well as the common CTAB protocol, all showed clearly lower success rates. The *innuPREP Plant DNA Kit* is the best commercial kit, but it only had a success rate of 56 %, all the others are below the 50 % rate.

CLAIMS

1. A method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

   a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids;

   b) purifying the nucleic acids extracted in (a).

2. The method of claim 1 wherein the lignified plant tissue is exposed to the extraction solution for about 5 hours.

3. The method of claim 1 wherein the lignified plant tissue is exposed to the extraction solution at a temperature of about 55^0C.

4. A method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

   a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a chelating agent and buffer to extract the nucleic acids;

   b) purifying the nucleic acids extracted in (a).

5. A method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

   a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a salt and buffer to extract the nucleic acids;

   b) purifying the nucleic acids extracted in (a).

6. A method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, an excipient and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

7. A method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a biological antioxidant and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

8. A method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a serine protease and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

9. A method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid, a chelating agent, a salt, an excipient, a biological antioxidant, a serine protease and buffer to extract the nucleic acids;

b) purifying the nucleic acids extracted in (a).

10. A method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: CTAB, Tris, NaCl, EDTA, Boric Acid, ,PVP K30, DTT, and Proteinase K;

b) purifying the nucleic acids extracted in (a).

11. A method for extracting plant nucleic acids from lignified plant tissue comprising the steps of:

a) exposing the lignified plant tissue to a solution comprising: 55mM CTAB, 100mM Tris, 1.4M NaCl, 20mM EDTA, 1% (w/v) Boric Acid, 2% (w/v) PVP K30, 47mM DTT, and 1.86% Proteinase K;

b) purifying the nucleic acids extracted in (a).

12. A method of determining the genus and/or species a sample of lignified plant tissue is from, comprising the steps of:

a) extracting plant nucleic acids from the lignified plant tissue using a method comprising the following steps:

i. exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids;

ii. purifying the nucleic acids extracted in (i);

b) using the nucleic acids from (a) to develop a genetic profile for the lignified plant tissue sample;

c) comparing the genetic profile from (b) with a library of genetic reference profiles.

13. A method of determining the plant source of a sample of lignified plant tissue, comprising the following steps:

a) extracting plant nucleic acids from the lignified plant tissue using a method comprising the following steps:

i. exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids;

ii. purifying the nucleic acids extracted in (i);

b) using the nucleic acids from (a) to develop a genetic profile for the lignified plant tissue sample;

c) comparing the genetic profile from (b) with a library of genetic reference profiles.

14. A method for developing a genetic profile for a sample of lignified plant tissue comprising the steps of:

a) extracting plant nucleic acids from lignified plant tissue using a method comprising the following steps:

i. exposing the lignified plant tissue to a solution comprising: a cationic surfactant, boric acid and buffer to extract the nucleic acids;

ii. purifying the nucleic acids extracted in (i);

b) subjecting the isolated nucleic acids of step (a) to a procedure chosen from the list comprising: PCR, RAPD analysis, AFLP analysis, CAPS analysis, RFLP analysis, SSR analysis, SNP analysis, SSCP analysis, sequencing.

15. A kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, boric acid and buffer;

b) instructions for use.

16. A kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, a chelating agent, boric acid and buffer;

b) instructions for use.

17. The kit of claim 16 wherein the instructions for use designate that the extraction is carried out by exposing the lignified plant tissue to the extraction solution for about 5 hours.

18. The kit of claim 16 wherein the instructions for use designate that the extraction is carried out by exposing the lignified plant tissue to the extraction solution at a temperature of about 55^0C.

19. A kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, a salt, boric acid and buffer;

b) instructions for use.

20. A kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, an excipient, boric acid and buffer;

b) instructions for use.

21. A kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, a biological antioxidant, boric acid and buffer;

b) instructions for use.

22. A kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, a serine protease, boric acid and buffer;

b) instructions for use.

23. A kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) a cationic surfactant, boric acid, a buffer a chelating agent, a salt, an excipient, a biological antioxidant and a serine protease;

b) instructions for use.

24. A kit for extracting plant nucleic acids from lignified plant tissue, said kit comprising:

a) CTAB, Tris, NaCl, EDTA, boric acid, PVP K30, DTT, Proteinase K;

b) instructions for use.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU2014/001047

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/10 (2006.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MEDLINE, CA, EMBASE, BIOSIS, AGRICOLA, WPI, EPODOC, TXTE: WOOD, TIMBER, LIGNIN, LIGNIFIED, TREE ROOT, HEARTWOOD, SAPWOOD, HARDWOOD, SOFTWOOD, GLULAM, PLYWOOD, LUMBER, CHIPBOARD, FIBREBOARD, LOG, CAMBIUM, BORIC ACID and like terms and synonyms.

Inventor name search.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	

[X] Further documents are listed in the continuation of Box C [ ] See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 January 2015	09 January 2015
Name and mailing address of the ISA/AU	**Authorised officer**
AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA Email address: pct@ipaustralia.gov.au	Susan MacKenzie AUSTRALIAN PATENT OFFICE (ISO 9001 Quality Certified Service) Telephone No. 0262832690

INTERNATIONAL SEARCH REPORT C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		International application No. **PCT/AU2014/001047**
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NOVAES, R.M.L. 'An efficient protocol for tissue sampling and DNA isolation from the stem bark of Leguminosae trees', Genetics and Molecular Research, 2009, Vol. 8, pages 86-96 Whole document	15-24
X	LOWE, A.J and CROSS, H.B. 'The application of DNA methods to timber tracking and origin verification', IAWA Journal, 2011, Vol. 32, pages 251-262 Whole document	15-24
X	TANG, X. et al., 'Wood identification with PCR targeting noncoding chloroplast DNA', Plant Molecular Biology, 2011, Vol. 77, pages 609-617 Whole document	15-24
X	ASIF, M.J. and CANNON, C.H., 'DNA extraction from processed wood: a case study for the identification of an endangered timber species (Gonystylus bancanus)', Plant Molecular Biology Reporter, 2005, Vol. 23, pages 185-192 Whole document	15-24
X	RACHMAYANTI, Y. et al., 'Extraction amplification and characterization of wood DNA from *Dipterocarpaceae*', Plant Molecular Biology Reporter, 2006, Vol. 24, pages 45-55 Whole document	15-24
X	REYNOLDS, M.M. and WILLIAMS, C.G., 'Extracting DNA from submerged pine wood', Genome, 2004, Vol. 47, pages 994-997 Whole document	15-24

APPENDIX VI: WOOD EXTRACTIONS: THE BASICS

Book section

The Migrant Ecologies Project: Jalan Jati (Teak Road)

WOOD EXTRACTION: THE BASICS

Duncan Jardine
Research Assistant for the Lowe Lab Group
University of Adelaide/State Herbarium of South Australia

The following is an interpretation of the methodology we use:

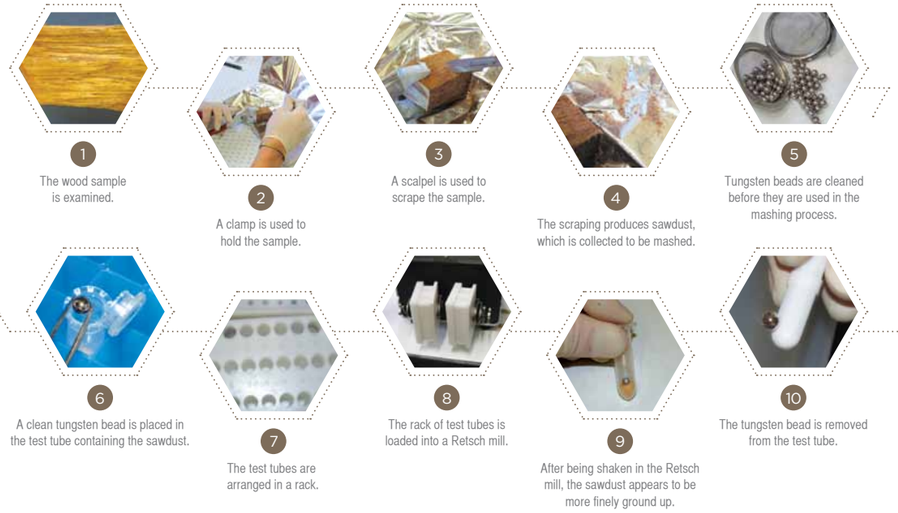
The wood sample is first examined (Fig. 1) and the most appropriate surface from which to scrape the sample is chosen. This is usually the outermost or youngest surface. Then, using a clamp to hold the sample (Fig. 2), a scalpel is used to scrape the sample (Fig. 3) to generate what is essentially sawdust (Fig. 4). Once enough sawdust has been collected, the material is transferred into a tube for mashing.

The mashing process involves tungsten beads (Figs. 5 and 6) and a device known as a Retsch mill (Figs. 7 and 8), that shakes the samples at high speeds to break open the cell walls of the tissue and allow the DNA to be extracted. The result looks like a slightly more ground-up form of the original sawdust (Fig. 9). The bead is then removed from the sample (Fig. 10) and the DNA extraction process can begin.

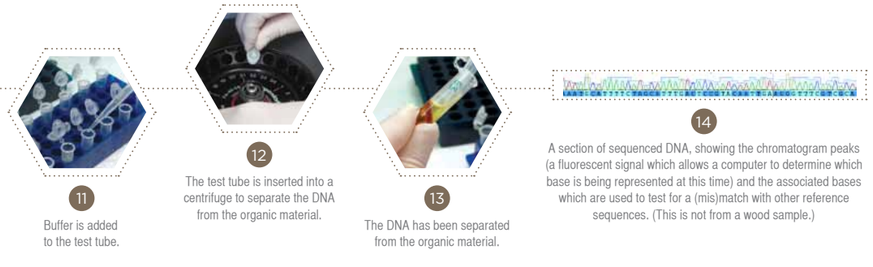
Buffer (Fig. 11) is added to the sample, along with chloroform, to separate the organic matter containing DNA from the other material. This is done using a centrifuge (Fig. 12) and a heat block. The separation can be seen in Fig. 13. The separated DNA is then purified, removing essentially everything except the DNA, so it can be amplified in a polymerase chain reaction (PCR).

The PCR increases the number of DNA fragments of a particular genetic region—in our cases, various barcoding regions. These regions help to identify the wood sample being extracted to a list of reference 'barcode' sequences (see Fig. 14 for an idea of this), so as to create a barcode that is unique for a particular genus, species or population. These barcodes are used to determine the (mis)match of the test piece to a barcode reference library.

To test the extraction barcode fragment to the barcode library, the amplified DNA needs to be sequenced, which allows a computer (and therefore the scientists) to see the sequence of bases (A, C, G and T) that make up a barcode. The computer-generated sequence can then be statistically tested against the reference library to determine the level of matching.



1 The wood sample is examined.
2 A clamp is used to hold the sample.
3 A scalpel is used to scrape the sample.
4 The scraping produces sawdust, which is collected to be mashed.
5 Tungsten beads are cleaned before they are used in the mashing process.
6 A clean tungsten bead is placed in the test tube containing the sawdust.
7 The test tubes are arranged in a rack.
8 The rack of test tubes is loaded into a Retsch mill.
9 After being shaken in the Retsch mill, the sawdust appears to be more finely ground up.
10 The tungsten bead is removed from the test tube.



11 Buffer is added to the test tube.
12 The test tube is inserted into a centrifuge to separate the DNA from the organic material.
13 The DNA has been separated from the organic material.
14 A section of sequenced DNA, showing the chromatogram peaks (a fluorescent signal which allows a computer to determine which base is being represented at this time) and the associated bases which are used to test for a (mis)match with other reference sequences. (This is not from a wood sample.)

APPENDIX VII: BIG LEAF MAPLE FINAL TECHNICAL REPORT

DoubleHELIXXX
#02-12/25 The Franklin – Science Park I
Singapore 118223 T +65 6227 9706
www.doublehelixtracking.com
GST/Reg No. 200812136R

Big Leaf Maple Final Technical Report

Version 002

DNA Fingerprinting Analysis of Big Leaf Maple (*Acer macrophyllum*)

Prepared by

Dr. Eleanor Dormontt, Duncan Jardine, Professor Andrew Lowe and Darren Thomas

This report supersedes all previous reports relating to this project.

Relevant Findings

- The DNA fingerprint from evidence sample LD210072 was a significant match to the DNA sample JS79 across 96 novel SNP loci developed for *A. macrophyllum*. Through population simulations generated using the reference sample data; we estimate the chance of obtaining such a match to be less than 1 in 100,000, if evidence sample LD210072 originated from an *A. macrophyllum* tree chosen at random.

- The DNA fingerprint from evidence sample LD210070 was a significant match to that from evidence sample LD210071 across 100 novel SNP loci developed for *A. macrophyllum*. Through population simulations generated using the reference sample data; we estimate the chance of obtaining such a match to be less than 1 in 100,000, if evidence sample LD210072 originated from an *A. macrophyllum* tree chosen at random.

- No other significant matches were detected.

- Sanger sequencing at a subset of seven loci showing good amplification and readable sequences confirmed 100% of the SNP calls made using the CRoPS method at five loci. One loci showed some indication of allele dropout occurring in the Sanger sequencing and another showed an indication of allele dropout occurring in CRoPS sequencing.

Objectives

Double Helix Tracking Technologies Pte Ltd (DoubleHelix) was contracted by the World Resources Institute to develop genetic marker resources for the Big Leaf Maple (*Acer macrophyllum*) suitable for genetic fingerprinting purposes.

Specifically, as part of an ongoing investigation by the USDA Forest Service Law Enforcement and Investigations, we have been asked to compare the DNA fingerprints of four separate wood samples to those of a reference group of 21 individuals of *A. macrophyllum* provided as pre-extracted DNA by the USFS National Forest Genetics Laboratory (NFGEL) in order to identify any DNA fingerprint matches.

We were informed by NFGEL that DNA from the wood of two felled trees found next to two tree stumps were also included. DNA extracted from leaves sprouting from these stumps were included in the original reference group of samples. It was assumed that the felled trees matched the stumps they were found next to, although this assumption was not independently verified by other means.

Samples Received

Wood material from evidence samples (Control No's: LD210070, LD210071, LD210072, LD210073) were sent from RA Malamphy (Law Enforcement Officer #1602 -US Forestry Services, 10024 US Hwy 12, PO Box 670, Randle, WA 98377, United States of America) to Mr. Duncan Jardine (Laboratory Technician, the University of Adelaide, L12 Shulz Building, North Terrace, Adelaide 5005) and received on the 11th of August 2012.

In addition, 25 tubes of pre-extracted, precipitated DNA from living *A. macrophyllum* trees were sent in two batches (two tubes sent 26th November 2012, followed by an additional 23 tubes sent 28th January 2013) by Dr. Valerie Hipkins (NFGEL Lab Director, 2480 Carson Road, Placerville, CA 95667, United States of America) addressed to Ms. Alison Jobling (batch 1) (ACEBB Administration Coordinator) and Mr. Duncan Jardine (batch 2) at the University of Adelaide (both to the same address as above). A list of sample names can be found in Table 1.

Table 1: List of all sample names and their status as either an evidence wood sample or a reference DNA sample.

Sample Name	Evidence wood sample	Reference DNA sample
LD210070	✓	
LD210071	✓	
LD210072	✓	
LD210073	✓	
JS79*		✓
JS80*		✓
JS81†		✓
JS82†		✓
JS83		✓
JS84		✓
JS85		✓
JS86		✓
JS87		✓
JS88		✓
JS89		✓
JS90		✓
JS91		✓
JS92		✓
JS93		✓
JS94		✓
JS95		✓
JS96		✓
JS97		✓
JS98		✓
JS99		✓
JT01		✓
JT02		✓
JT03		✓
JT04		✓

* and † DNA assumed to be from the same individual (stump and felled log respectively)

Personnel and project duration

All works were carried out by Dr. Eleanor Dormontt, Mr. Duncan Jardine or Professor Andrew Lowe at the University of Adelaide's Centre for Evolutionary Biology and Biodiversity (ACEBB) except where explicitly stated otherwise. Work for this project was formally carried out between December 4th 2012 and December 4th 2013.

Methodology

DNA extraction

DNA extractions from the evidence samples followed a modified BoTab DNA extraction protocol (patent pending). Twelve extraction reactions were undertaken in three batches. The extraction regime consisted of an initial negative extraction followed by the extraction of wood samples. A final negative extraction was carried out for further quality control.

DNA concentrations

All samples were quantified to determine their nucleic acid concentration, using a Qubit. We used 10 µl of DNA, with batch 1 and 2 samples quantified using the high sensitivity chemistry, and batch 3 samples quantified with the broad range chemistry. Three separate DNA quantification readings were taken and an average calculated (Appendix 1).

Genetic marker development - CRoPS library preparation and Ion Torrent sequencing

A modified complexity reduction of polymorphic sequences (CRoPS) approach (van Orsouw *et al,* 2007) was used to generate single nucleotide polymorphism (SNP) data. This protocol is based on those used in amplified fragment length polymorphisms (AFLP) analyses (Vos *et al*, 1995), with several additional steps employed to generate DNA sequence data through next generation sequencing (NGS). The technique consists of a restriction digest of genomic DNA, ligation of blunt end adapters, followed by two rounds of polymerase chain reactions (PCR) designed to selectively amplify a range of DNA fragments. These fragments are then processed using next generation sequencing. Full details of the exact protocols are available on request. Samples were run in three batches; run 1 (14 *A. macrophyllum* samples and a negative control); run 2 (22 *A. macrophyllum* samples and a negative control) and; run 3 (32 *A. macrophyllum* samples and a negative control). A full list of the samples used in each run is in Appendix 2.

The *Ion Torrent*™ platform was used for NGS, which at the time of this report was capable of sequencing up to 12 million 200-bp reads in one sequencing run. For sequencing to run efficiently, pooled products need to be size selected to remove either product that is too small (which will be preferentially sequenced) or product that is too large (which will impact on the accuracy of quantification). An Egel™ or

Ampure® size selection method was used to select fragements of appropriate size for NGS. Size selected products were quantified on both a Qubit and a Tape station 2200®.

The laboratory at the University of Adelaide currently has an *Ion Torrent One Touch* and *Personal Genome Machine (PGM)* in house for NGS. For all three sequencing runs, the *One Touch* run was performed in house. The first two runs were also sequenced in house using the *PGM*, however the third and final run was performed at Australian Genome Research Facility (AGRF) in Adelaide using their *PGM*.

All resulting data was analysed in house using the CLC genomics workbench program (CLCbio.com). Sequences were separated according to their unique sample identifier, trimmed of primer product and then run through a modified PRGmatic (Hird *et al,* 2011) pipeline program capable of detecting single nucleotide polymorphisms (SNPs). SNPs were visually verified in Tablet (Milne *et al,* 2013) then sorted and concatenated in Geneious (www.geneious.com), before being analysed by our in-house genotype matching software. Simulations were undertaken using Resampling Stats (statistics.com).

[π]Life Technologies Corporation, 5791 Van Allen Way, Carlsbad, CA 92008, United States of America

[φ]Beckman Coulter Australia Pty Ltd 23-27 Chaplin Drive Lane Cove NSW 2066 Australia

[β]Agilent Technologies Australia Pty Ltd, 679 Springvale Road, MULGRAVE Victoria 3170, Australia

Hird, S. M., Brumfield, R. T., & Carstens, B. C. (2011). PRGmatic: an efficient pipeline for collating genome-enriched second-generation sequencing data using a 'provisional-reference genome'. *Molecular Ecology Resources*, *11*(4), 743-748.

Milne, I., Stephen, G., Bayer, M., Cock, P.J.A., Pritchard, L., Cardle, L., Shaw, P.D. and Marshall, D. 2013.Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*, *14*(2), 193-202.

van Orsouw, N. J., Hogers, R. C., Janssen, A., Yalcin, F., Snoeijers, S., Verstege, E., ... & van Eijk, M. J. (2007). Complexity reduction of polymorphic sequences (CRoPS™): a novel approach for large-scale polymorphism discovery in complex genomes. *PloS one*, *2*(11), e1172.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van De Lee, T., Hornes, M., ... & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic acids research*, *23*(21), 4407-4414.

Sanger Sequencing

To assess our confidence in the CRoPS method we verified a selection of SNP calls via traditional capillary separation (Sanger sequencing). Independent primer sequences (forward and reverse) were designed from the flanking regions of 13 SNPs identified through the CRoPS approach as applied to *A. macrophyllum*. Sequencing PCR reactions were undertaken using these primer pairs with eight different DNA samples, one repeated DNA sample, and a negative control. PCR products were sequenced at the Australian Genome Research Facility (AGRF) in Adelaide.

Results

Next Generation Sequencing runs

Of the three sequencing runs, the third and final run was the most informative. The first run was considered a pilot, with fewer samples included. The second run yielded no usable data. Due to the nature of the library preparation, the ability to merge sequence data between runs is limited. When we attempted to merge data from runs 1 and 3 it yielded no overlapping loci. Thus, the following results are derived only from the third sequencing run.

A combined total of 2.2 million usable sequences were amplified in the third run with an average of 70 thousand reads per sample. This equates to 2285 initial allele calls after the data were first passed through the PRGmatic pipeline. After uninformative loci and unusable SNPs were removed, a potential 241 SNPs were identified in the 30 test samples. Samples JS80 and JS82 (DNA extracted from felled trees by NFGEL) had insufficient sequencing coverage to be useful (only 12% and 26% of loci had data respectively).These poor results are likely caused by the DNA extraction method employed by NFGEL which was not optimised for degraded wood samples. Data from JS80 and JS82 were not included in any further analyses.

SNP selection

The final SNP dataset was finalised prior to any between sample comparisons being undertaken (except between replicated samples from LD210072 as these represent results from what we know to be 100% matching DNA). Any loci that showed inconsistent results between these replicate samples were removed from the entire dataset (for further details on the nature of these inconsistencies, please see the next section on 'defining expected match percentage'. All loci that called a SNP in 15 or less individuals were also removed to minimise missing data and facilitate genotype matching. Data from the repeated sample LD210072 was then combined to create a single fingerprint for that individual. Any loci where no data were available for LD210072 were also removed. The final number of suitable SNP loci available for downstream analyses was 101. A table of genotypes at each of those 101 loci in all individuals (excluding JS80 and JS82) can be found in Appendix 3.

Defining expected match percentage

Although theoretically we would expect to observe a 100% match between DNA profiles that originate from the same individual, this is not always the case. Possible explanations for a deviation from a 100% match include:

- **DNA degradation** - a particular problem with DNA extracted from non-living wood tissue. Degraded DNA can cause unexpected results; this is particularly a problem with length variation measures of genetic variability such as microsatellites. The problem is lessened with SNP loci but the effects on timber DNA fingerprinting analysis have not been exhaustively examined to date.

- **Paralogous loci** (areas where the same sequence is repeated in different parts of the genome) exhibiting different SNP variants. Every effort is made to avoid these when selecting loci (i.e. by rejecting SNP loci with three or more alleles), but their complete exclusion cannot be guaranteed.

- **Allele dropout**, where uneven signal strength between alleles can lead to the weakest allele signal being ignored by the SNP calling algorithms.

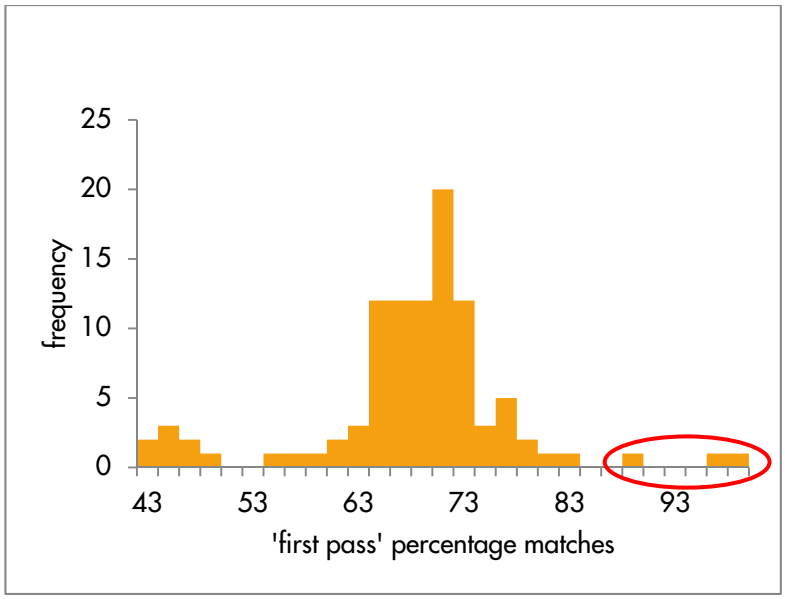
In cases of paralogous loci or allele dropout, we can expect that a heterozygote (an individual with two different alleles at a particular locus) may have been incorrectly assumed to be a homozygote (an individual with all the same alleles at a locus). In the same individual however, we would not expect loci to be incorrectly called as a homozygote for the alternative allele. Where homozygotes for alternative alleles are called at a particular locus, this is a strong indication that the DNA samples do not originate from the same individual.

In order to quantify the match percentage that we can expect between samples taken from the same individual, we analysed the results from the sequencing of the two separate extractions of evidence sample LD210072. We analysed a total of 114 loci (prior to removal of the loci that were not consistently scored between the repeated samples). A percentage match of 93.4% was calculated between the samples. Ninety nine loci (86.8%) matched both alleles, the remaining 15 loci (13.2%) matched at one of two alleles. As expected there were no loci (0%) that did not match any alleles between these two samples. This finding is consistent with our expectation that no alternative homozygotes would be called in two samples originating from the same individual. All inconsistencies identified between the two samples are explained by allele dropout.

Comparing DNA fingerprints between samples

In-house genotype matching software was used to calculate the percentage matches obtained when comparing the evidence and reference samples. We began by comparing each evidence sample with all other samples, restricting loci to those present in the evidence sample in question. In this 'first pass' any missing data in the samples for comparison were treated as 'no match'. The results from the first pass of each evidence sample against all other samples were assessed for any unusually high values (outliers) indicating possible matches. Three percentage matches were identified as being particularly high (Figure 1). These were LD210072 and JS79 (89.6%), and LD210070 and LD210071 (97.5% and 98.5% depending on the direction of the comparison due to the inclusion of different loci with missing data).

Figure 1. Frequency histogram of 'first pass' genotype matches between evidence and reference samples. Any missing data in the samples for comparison were treated as 'no match'. Three comparisons stand out as being particularly high (circled).



Where a potential match was identified in the first pass comparisons, the samples were individually compared again, this time using only loci present in both samples ('no missing data'). Once missing data were removed from the comparisons, the percentage matches for LD210072 and JS79 increased to 94.3%. LD210070 and LD210071 remained at 98.5%..

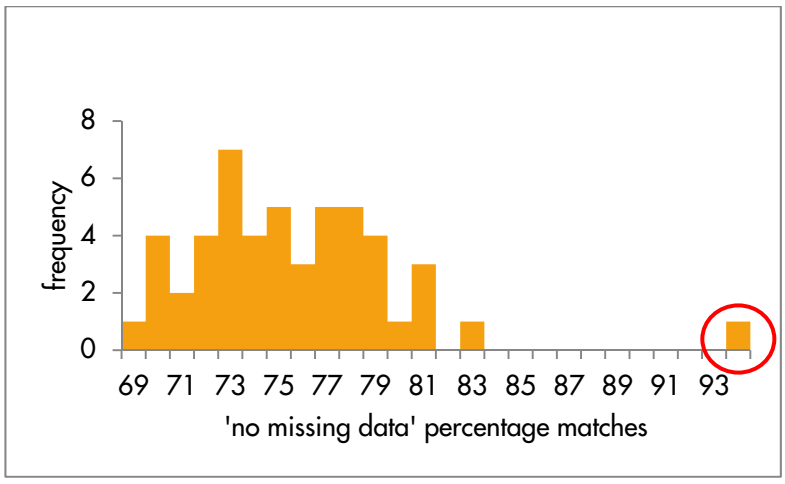
Of the 11 loci matching only one allele between LD210072 and JS79, six loci called homozygotes in DNA from leaf material, and heterozygotes in the wood material. Conversely five loci called homozygotes in DNA from wood material as opposed to leaf material. These results show that allele dropout did not preferentially occur in the DNA profile obtained from wood, so is unlikely to be caused by DNA degradation in the wood sample. These comparisons also revealed no loci (0%) that did not match any alleles between the two compared samples

Based on the very high match between LD210070 and LD210071 and the absence of any completely mismatched loci between these two evidence samples, we treated them as a single data point (hereafter referred to as LD210071) in order to avoid any psuedoreplication (except where the match between LD210070 and LD210071 was itself being assessed).

To ascertain the importance of the increased match percentage obtained between LD210072 and JS79 with the removal of all missing data, we compared it to the match percentage obtained in other comparisons between the evidence samples (LD210072 and LD210071) and the rest of the samples, again with no missing data. This approach (dynamic removal of missing data for each individual comparison) is computationally more intensive but maximises the utility of the available data without compromising on accuracy or requiring any assumptions about missing data values.

The results of these more accurate comparisons provided even stronger evidence that the 94.3% match obtained between LD210072 and JS79 was outside of the normal range of matches obtained when comparing against other samples (Figure 2). A table of all the performed 'no missing data matches' and their outcomes can be found in Appendix 4.

Figure 2. Frequency histogram of 'no missing data' genotype matches between evidence samples LD210072 , LD210071 and reference samples. Any loci with missing data in the samples for comparison were removed. The comparison between LD210072 and JS79 stands out as being particularly high (circled).



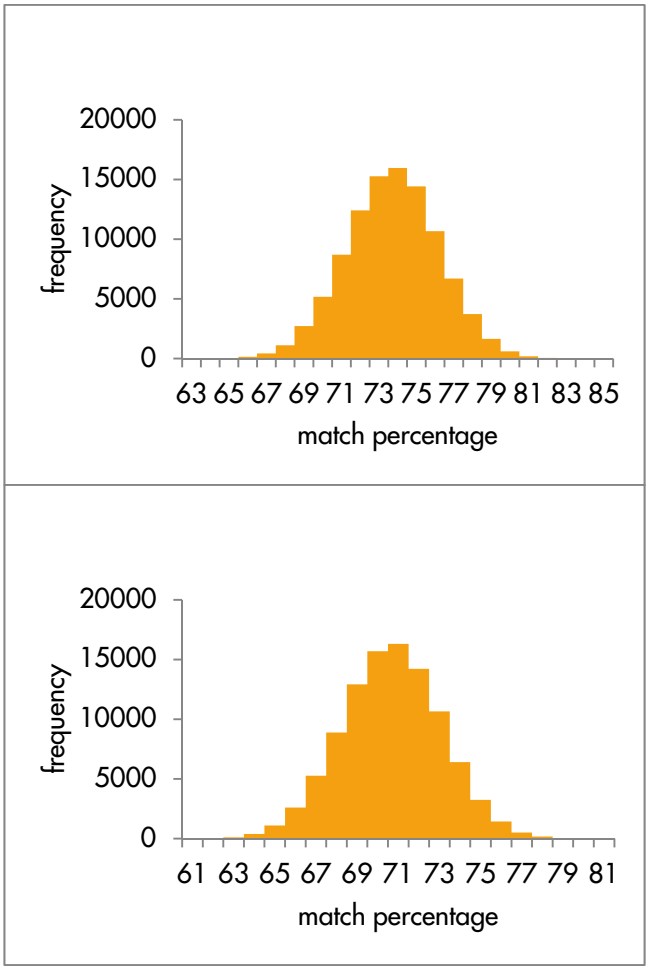
Assessing the likelihood of observed matches by chance

In order to quantify the significance of the observed matches between LD210070 and LD210071, and LD210072 and JS79 we used a simulation approach. We generated a pool of available genotypes at each locus from those present in the reference samples. The loci included in each simulation were only those present in the evidence sample in question (LD210071 or LD210072).

For each match, we simulated an individual by randomly selecting a genotype at each locus (with replacement). The more conservative approach was taken of choosing complete loci genotypes rather than individual alleles as units for sampling, to avoid reliance on assumptions of Hardy Weinberg equilibrium at each locus. We then compared our simulated individual to the matched sample (either LD210071 or LD210072) and generated a percentage match.

This process was repeated 100,000 times and the distribution of observed match percentages assessed (Figure 3, Appendix 5). In each case, the generated distribution failed to include a match percentage as high as those observed in the real data. This equates to a chance of less than 1 in 100,000 of the observed percentage matches occurring by random chance, based on the reference samples provided. A chance of less than 1 in 100,000 is equivalent to a statistical significance of $P<0.00001$.

Figure 3. Frequency histograms of match percentages obtained via resampling of 100,000 randomly simulated individuals. A) Distribution generated from comparison with LD210071. Mean = 73.74 %, Standard deviation = 2.45%. B) Distribution generated from comparison with LD210072. Mean = 70.71%, Standard deviation = 2.39%. Frequency tables can be found in Appendix 5.



Sanger Sequencing

The results obtained from Sanger sequencing in this case are consistent with those we obtained through CRoPS. Details of individuals and loci analysed through Sanger sequencing can be found in Appendix 6. Sanger sequencing is a robust and well validated method for analysing DNA fingerprints and has the established confidence of the scientific and legal community. However, it is still not without its potential problems, which include those previously described for CRoPS (degraded DNA, paralogous loci, allele dropout). Of the 13 SNPs we attempted to Sanger sequence, we were able to get readable sequence data for nine across all or most of the individuals analysed. The remaining four loci did not amplify correctly or did not have enough sequence coverage to accurately call the base pairs at each position. This is a problem caused by primer design, but is not a reflection on the accuracy of the original SNP calls made with the CRoPS approach.

Of the nine readable sequences, two SNP loci found within the same stretch of sequence (one set of primers used to amplify both) gave inconsistent results in some individuals, both against the CRoPS results but also within Sanger runs (forward and reverse sequencing results different). In this instance it is difficult to identify the cause of these mismatched results but we suspect this may be due to paralogous sequences elsewhere in the genome contributing to the signal being detected in one or both of the sequencing methods.

Five of the seven remaining loci which produced good sequence data were 100% consistent with the results obtained from the CRoPS method. In the other two loci there were some mismatches between Sanger and CRoPS.

At the first mismatched locus, individual LD210071 was called as a homozygote by the Sanger sequencing and a heterozygote by CRoPS sequencing. All other allele calls at that locus were identical between the two methods. As we have shown earlier in the report, this individual (LD210071) is a highly significant match to evidence sample LD210070, which was also Sanger sequenced at this locus. Results for Sanger sequencing of this second evidence sample call a heterozygote and are consistent with CRoPS. Therefore, we believe the most plausible explanation for this discrepancy is actually allele dropout occurring in the Sanger sequencing reaction for LD210071 at this locus, rather than an error in the CRoPS allele calls.

The second mismatched locus showed consistent results across both Sanger and CRoPS in two of seven individuals where sequences were available, but called heterozygotes in the Sanger sequencing and homozygotes in the CRoPS sequencing in the remaining five individuals. So in this case the results are a match for one allele and possibly reflect allele dropout occurring in the CRoPS sequencing at this locus. All negative controls showed no DNA sequences.

Overall we consider these results a successful verification of the CRoPS method using Sanger sequencing and note that all inconsistencies identified between the methods can be explained by allele dropout, no loci were called as alternative homozygotes. Further, the allele dropout identified did not occur preferentially in results obtained through either method, suggesting they perform similarly and can be complementary when used together to confirm genotypes.

Conclusion

The successful amplification and sequencing of DNA from wood samples using the CRoPS method employed by Professor Andrew Lowe and his team at the University of Adelaide was able to generate and screen 101 SNPs across 25 samples of *A. macrophyllum*.

Results matched evidence sample LD210072 with that of reference sample JS79 and additionally matched evidence sample LD210070 with evidence sample LD210071.

Sanger sequencing verification on a subset of individuals and loci indicated that the CRoPS method performs as well as Sanger sequencing for producing robust sequencing data.

Future Scope

The development of 101 SNP loci represents a significant and valuable resource for future study and protection of *A. macrophyllum*.

In order to enable routine forensic identification of *A. macrophyllum* across its natural range, further work is recommended to sample and process additional populations in the western United States.

The testing process can be further optimised to increase the speed of analysis and minimise the cost of testing, enabling effective enforcement.

Appendix 1. Calculated DNA concentration readings

DNA concentration (µg ml^{-1})

Sample		Blank Extraction			LD210070		LD210071		LD210072		LD210073		H_20
		#1	#2	#3	#1	#2	#1	#2	#1	#2	#1	#2	
Reading #	1	0.02	0.02	-	8.36	8.18	2.33	7.20	8.30	9.32	7.32	8.74	-
	2	0.03	0.02	-	8.17	8.18	2.30	7.20	7.93	9.43	7.01	8.71	-
	3	0.03	0.02	-	8.16	8.00	2.30	7.11	7.94	9.23	7.10	8.72	-
	mean	0.03	0.02	-	8.23	8.12	2.31	7.17	8.06	9.33	7.14	8.72	-

Extraction batches: *1, †2, ‡3.

Appendix 2. Details of samples used in each Ion Torrent sequencing run

Sample	Sequencing Run 1	Sequencing Run 2	Sequencing Run 3
Blank Extraction #1	✓	✓	
Blank Extraction #2			✓
Blank Extraction #3			✓
LD210070 #1	✓	✓	
LD210070 #2			✓
LD210071 #1	✓	✓	
LD210071 #2			✓
LD210072 #1	✓		✓
LD210072 #2			✓
LD210073 #1	✓	✓	
LD210073 #2			✓
JS79	✓	✓	✓
JS80		✓	✓
JS81		✓	✓
JS82		✓	✓
JS83	✓		✓
JS84		✓	✓
JS85	✓		✓
JS86		✓	✓
JS87	✓	✓	✓
JS88		✓	✓
JS89		✓	✓
JS90	✓		✓
JS91		✓	✓
JS92		✓	✓
JS93	✓	✓	✓
JS94		✓	✓
JS95	✓		✓
JS96		✓	✓
JS97		✓	✓
JS98		✓	✓
JS99		✓	✓
JT01	✓		✓
JT02		✓	✓
JT03	✓		✓
JT04	✓		✓

Appendix 3. Table of allele calls at 101 loci in all individuals (excluding JS80 and JS82 which were removed due to very high levels of missing data). Base pair calls are coded as follows: A=1, C=2, G=3, T=4 and missing =5.

	LD210070	LD210071	LD210072	LD210073	JS79	JS81	JS83	JS84	JS85	JS86	JS87	JS88	JS89	JS90	JS91	JS92	JS93	JS94	JS95	JS96	JS97	JS98	JS99	JT01	JT02	JT03	JT04
MB3_2_114.1	1	1	1	1	1	1	2	2	2	2	1	1	1	1	1	2	1	1	1	1	1	2	1	2	1	1	2
MB3_2_114.1	2	2	2	2	2	2	1	1	1	1	2	2	2	2	2	1	2	2	2	2	2	1	2	1	2	1	1
MB3_2_114.2	1	1	1	1	1	1	3	1	3	3	1	1	3	1	1	3	1	1	1	1	1	3	1	3	1	1	3
MB3_2_114.2	3	3	3	3	1	3	1	3	1	1	3	3	1	3	3	1	3	3	3	3	3	1	3	1	3	3	1
MB3_2_23.1	1	1	1	1	3	1	1	1	1	3	1	1	3	1	3	1	1	3	1	1	1	3	3	1	1	1	1
MB3_2_23.1	1	1	3	1	1	3	1	1	1	3	1	1	3	1	3	1	1	1	1	1	1	1	1	1	1	1	1
MB3_2_23.2	1	1	2	2	2	1	1	2	2	2	2	2	2	1	2	1	2	2	2	2	1	2	2	2	1	1	2
MB3_2_23.2	1	1	2	2	2	2	2	2	2	2	1	1	2	2	2	1	1	2	1	1	1	2	2	2	2	2	1
MB3_2_23.3	4	4	2	2	4	4	4	2	2	4	4	2	4	2	4	4	2	4	2	2	4	4	4	2	4	4	2
MB3_2_23.3	4	4	4	2	2	4	2	2	2	4	2	4	4	4	4	4	4	2	4	4	4	2	2	2	2	2	4
MB3_2_23.4	3	3	3	1	3	3	3	3	3	3	3	3	3	1	3	3	3	3	1	3	3	3	3	3	3	3	3
MB3_2_23.4	3	3	3	1	3	3	1	1	1	3	3	3	3	3	3	3	3	1	3	3	3	1	3	3	1	3	3
MB3_2_2796.1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
MB3_2_2796.1	2	2	2	2	2	2	2	2	2	4	2	2	2	2	2	2	2	2	4	4	2	4	2	2	2	2	2
MB3_2_2796.2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
MB3_2_2796.2	1	1	3	3	3	3	3	3	3	1	3	1	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3
MB3_2_31	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	2	1	1	1	1
MB3_2_31	1	1	1	2	1	1	1	1	2	2	1	1	1	1	2	2	2	2	2	2	1	1	1	1	1	1	1
MB3_2_10.3	3	3	3	3	3	3	1	3	3	3	3	3	1	3	3	1	1	3	1	1	3	1	3	3	3	3	1
MB3_2_10.3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3
MB3_2_102.1S	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2	2
MB3_2_102.1S	2	2	2	2	2	2	2	3	2	3	2	2	2	2	3	2	2	3	3	2	2	2	3	2	2	2	2

MB3_2_103	3	3	3	3	3	3	3	3	3	5	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
MB3_2_103	4	4	4	3	4	3	3	3	3	5	3	3	3	3	3	4	4	3	4	3	3	3	3	3	3	3	3
MB3_2_168.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MB3_2_168.1	4	4	1	4	4	4	4	4	1	4	4	4	4	4	4	1	1	4	4	4	1	4	4	4	4	4	4
MB3_2_168.2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
MB3_2_168.2	1	1	3	1	1	1	1	1	3	1	1	1	1	1	1	3	3	1	1	1	3	1	1	1	1	1	1
MB3_2_168.3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	1	3
MB3_2_168.3	3	3	1	3	1	3	3	3	1	3	3	1	3	3	3	3	1	3	3	3	3	3	3	3	3	3	3
MB3_2_168.4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	3	1
MB3_2_168.4	1	1	3	1	3	1	3	1	3	1	1	3	1	1	1	1	3	1	1	1	1	1	1	3	1	3	1
MB3_2_22	1	1	4	4	4	4	4	4	4	4	4	4	1	1	1	1	1	4	1	1	1	4	1	4	4	4	4
MB3_2_22	1	1	4	1	4	1	1	4	4	4	4	4	1	4	4	1	4	4	4	1	1	1	4	4	4	4	4
MB3_2_2433	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
MB3_2_2433	4	4	4	4	4	2	4	4	4	4	4	2	4	2	4	4	4	4	4	4	4	4	4	2	4	4	2	4
MB3_2_262.1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
MB3_2_262.1	4	2	2	2	2	4	2	2	2	2	2	2	4	2	2	4	2	2	2	2	2	2	2	4	2	2	2
MB3_2_262.2	4	4	2	4	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
MB3_2_262.2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4
MB3_2_262.3	1	1	4	1	4	1	4	1	1	1	1	1	1	1	1	1	1	4	1	1	1	1	4	1	1	1	1
MB3_2_262.3	4	4	1	4	1	4	1	4	4	4	4	4	1	4	4	4	4	1	4	4	4	4	1	1	4	1	4
MB3_2_265.1S	3	3	3	3	3	3	1	1	3	3	1	3	3	3	3	1	3	1	1	3	3	3	1	1	1	3	1
MB3_2_265.1S	3	3	3	3	3	1	1	1	3	3	3	3	1	1	3	3	1	1	1	1	1	3	3	1	1	3	1
MB3_2_265.2S	3	3	3	3	3	3	4	3	3	3	3	3	3	3	3	4	3	3	4	3	3	3	3	3	3	3	3
MB3_2_265.2S	3	3	3	3	3	3	4	3	3	3	3	3	4	3	3	3	4	3	4	4	4	3	3	3	3	3	4
MB3_2_265.3S	4	4	4	4	4	4	1	1	4	4	1	4	4	4	4	1	4	1	1	4	4	4	1	1	1	4	1
MB3_2_265.3S	4	4	4	4	4	1	1	1	4	4	4	4	1	1	4	4	1	1	1	1	1	4	4	1	1	4	1
MB3_2_30.4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4	4	2	4	4	4	4	4	4	4	4	2	4
MB3_2_30.4	2	2	2	2	2	4	2	2	4	2	2	2	2	2	2	2	4	2	2	2	2	4	4	2	2	4	2
MB3_2_30.5	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
MB3_2_30.5	3	3	3	3	4	3	3	3	3	3	3	3	4	3	3	3	3	3	3	3	3	3	3	3	3	3	3
MB3_2_30.6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MB3_2_30.6	1	1	1	1	3	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MB3_2_30.7	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
MB3_2_30.7	4	4	4	4	1	4	4	4	4	4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4

MB3_2_38.1	1	1	4	1	1	4	1	4	4	1	4	4	4	4	4	4	4	1	4	4	1	1	4	4	1	1	4
MB3_2_38.1	1	1	1	1	4	1	4	4	1	1	4	4	4	4	1	1	4	4	4	1	1	1	1	1	4	4	4
MB3_2_38.2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
MB3_2_38.2	2	2	2	2	2	1	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
MB3_2_49	1	1	1	5	1	1	1	1	1	1	1	1	1	1	3	1	1	3	3	1	1	1	1	1	1	1	1
MB3_2_49	1	1	1	5	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3	1	1	3
MB3_2_60S	2	2	4	4	4	4	4	2	4	2	4	2	4	2	4	4	4	4	2	2	2	4	4	4	2	4	4
MB3_2_60S	4	4	2	4	2	2	2	4	2	2	2	2	4	4	4	2	4	2	4	4	2	4	2	2	4	2	2
MB3_2_77.2S	4	1	4	1	4	1	1	4	1	1	1	1	4	4	1	1	1	1	1	4	1	4	4	1	1	1	1
MB3_2_77.2S	1	4	1	4	1	4	1	1	4	4	4	4	1	1	4	4	4	4	4	1	4	1	1	4	4	4	4
MB3_2_84.1	3	1	3	1	3	3	3	1	1	1	3	1	3	1	3	3	1	3	1	3	1	3	3	3	1	1	1
MB3_2_84.1	1	3	3	1	3	3	1	3	1	3	3	3	3	3	3	1	1	3	3	3	3	3	1	3	1	1	3
MB3_2_84.2	4	4	2	4	2	2	2	4	4	4	2	4	2	4	2	2	4	2	4	2	4	2	2	2	4	4	4
MB3_2_84.2	2	2	2	4	2	2	4	2	4	2	2	2	2	2	2	4	4	2	2	2	2	2	4	2	4	4	2
MB3_2_84.3	4	2	4	2	4	4	4	2	2	4	4	2	4	2	4	4	2	4	4	4	4	4	4	4	2	2	2
MB3_2_84.3	2	4	4	2	4	4	2	4	2	2	4	4	4	4	4	2	2	4	2	4	2	4	2	4	2	2	4
MB3_2_942(3).3	3	3	3	3	3	3	1	3	3	3	3	3	3	3	1	1	3	3	1	1	3	1	3	3	3	3	1
MB3_2_942(3).3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	1	3	3	1	3	3	3	3	3	3	1
MB3_2_97.1S	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
MB3_2_97.1S	2	2	2	2	2	2	2	1	2	1	1	2	2	2	2	1	1	2	1	2	2	1	2	2	2	2	2
MB3_2_97.2S	2	2	4	2	4	2	2	2	2	2	2	2	4	2	2	2	2	2	2	2	2	2	2	2	2	2	4
MB3_2_97.2S	2	2	4	2	4	4	4	4	4	2	2	4	4	2	4	2	2	4	2	2	2	2	2	4	4	2	4
MB3_2_97.3S	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MB3_2_97.3S	3	3	1	1	1	1	1	1	1	3	1	3	1	3	1	3	3	3	3	3	3	1	3	3	1	1	1
MB3_2_97.4S	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
MB3_2_97.4S	4	4	4	4	4	3	4	4	3	4	4	4	4	4	3	4	4	4	4	3	4	4	4	4	3	3	4
MB3_2_97.5S	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
MB3_2_97.5S	2	2	2	2	2	2	2	2	2	3	3	2	2	2	2	3	3	3	3	2	2	3	3	2	2	2	2
MB3_2_97.6S	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MB3_2_97.6S	1	1	1	1	1	3	1	1	3	1	1	1	1	1	3	1	1	1	1	3	1	1	1	1	3	3	1
MB3_2_110	4	4	4	4	4	4	5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	2	4	4
MB3_2_110	4	4	4	4	4	4	5	4	4	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	2	4	4
MB3_2_13.1	4	4	4	4	4	4	5	4	4	4	4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4	1
MB3_2_13.1	4	4	4	4	4	4	5	4	4	4	1	4	4	4	1	4	4	1	4	4	4	4	1	4	4	4	4

MB3_2_13.2	4	4	4	4	4	4	5	4	4	4	4	4	4	2	4	4	4	4	4	4	4	4	4	4	4	4	2
MB3_2_13.2	4	4	4	4	4	4	5	4	4	4	2	4	4	4	2	4	4	2	4	4	4	4	2	4	4	4	4
MB3_2_13.3	4	4	1	4	4	4	5	1	4	1	1	1	1	1	1	4	4	4	4	1	1	1	1	1	1	1	1
MB3_2_13.3	4	4	4	4	1	1	5	4	1	1	1	1	1	1	1	4	4	1	4	1	1	4	1	1	4	1	1
MB3_2_13.4	4	4	4	4	4	4	5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4
MB3_2_13.4	4	4	4	4	4	4	5	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
MB3_2_13.5	3	3	3	3	3	3	5	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	3
MB3_2_13.5	3	3	3	3	3	3	5	3	3	3	3	4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
MB3_2_130.1	3	3	1	3	3	3	5	3	1	3	3	3	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3
MB3_2_130.1	3	3	3	3	1	3	5	3	3	1	3	3	3	3	3	3	3	1	1	3	3	3	3	3	1	3	3
MB3_2_130.2	1	1	3	1	1	1	5	1	3	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1
MB3_2_130.2	1	1	1	1	3	1	5	1	1	3	1	1	1	1	1	1	1	3	3	1	1	1	1	1	3	1	1
MB3_2_184.1	1	1	1	3	1	1	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MB3_2_184.1	3	3	1	3	1	1	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MB3_2_184.2	4	4	4	4	4	4	5	1	4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
MB3_2_184.2	4	4	4	4	4	1	5	1	1	4	4	1	1	1	1	1	4	4	4	4	4	4	1	4	4	4	1
MB3_2_191	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	5	3	3	3	3	3	3	3	3	3
MB3_2_191	3	3	3	3	3	3	1	3	3	3	3	3	3	3	1	3	3	5	1	1	3	3	3	1	1	1	3
MB3_2_455.1	3	3	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	5	3	1	3	3	3
MB3_2_455.1	3	3	1	1	1	3	3	1	3	3	3	1	3	3	3	3	3	3	3	1	1	5	3	3	3	1	1
MB3_2_48.1	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	5	3	3	3	3	3
MB3_2_48.1	3	3	3	3	1	3	3	1	3	3	3	1	3	3	3	3	3	3	3	3	3	5	3	3	3	3	3
MB3_2_48.2	4	4	3	3	3	4	4	3	3	3	3	3	3	3	3	4	4	4	4	3	3	5	4	3	3	3	3
MB3_2_48.2	4	4	3	3	3	3	4	4	4	3	4	3	3	3	4	3	4	3	3	4	3	5	4	4	4	3	3
MB3_2_827S	1	1	1	1	1	3	3	3	1	1	3	3	5	3	3	1	3	3	1	1	3	1	3	3	3	3	3
MB3_2_827S	1	1	1	3	1	3	1	3	3	3	1	3	5	3	3	1	3	1	1	1	1	3	1	1	3	1	1
MB3_2_88.1S	4	3	3	4	3	3	3	4	3	3	3	3	3	3	3	4	4	3	3	3	3	5	4	3	3	3	3
MB3_2_88.1S	3	4	4	4	4	3	3	4	4	3	4	4	3	3	3	4	4	3	4	4	3	5	3	4	4	4	3
MB3_2_88.2S	3	3	3	3	3	1	3	3	3	1	1	3	3	1	3	3	3	3	3	3	1	5	3	3	3	3	1
MB3_2_88.2S	1	1	3	3	3	1	3	3	3	3	3	1	3	3	3	3	3	1	3	3	1	5	3	3	3	3	1
MB3_2_104.1	3	3	3	1	3	1	3	1	3	3	3	3	3	3	3	1	3	1	3	3	3	5	3	3	3	5	3
MB3_2_104.1	3	3	1	3	1	3	3	1	3	3	3	3	3	3	3	1	1	3	1	3	1	5	1	1	1	5	3
MB3_2_104.2	3	3	3	2	2	2	3	2	3	3	3	3	3	3	3	2	2	3	2	3	2	5	3	3	3	5	3
MB3_2_104.2	3	3	2	3	3	3	3	2	3	3	3	3	3	3	3	2	3	2	3	3	3	5	2	2	2	5	3

MB3_2_106	4	4	4	4	4	5	5	4	4	3	4	4	4	4	4	3	4	4	3	4	4	5	4	4	3	4	4
MB3_2_106	4	4	4	4	4	5	5	3	4	4	4	4	4	4	4	4	3	4	4	4	4	5	4	4	3	3	3
MB3_2_19.1	4	4	4	4	4	4	4	2	4	4	2	4	4	4	4	4	4	4	2	2	4	5	4	5	4	4	4
MB3_2_19.1	4	4	4	4	4	4	2	4	4	4	4	4	4	4	4	2	2	4	4	4	4	5	4	5	4	4	2
MB3_2_19.3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	5	2	5	2	2	2
MB3_2_19.3	2	2	2	2	2	2	2	4	2	2	2	2	2	4	2	2	2	2	2	2	2	5	2	5	2	4	2
MB3_2_238	4	4	4	4	4	5	4	4	2	4	2	2	2	2	2	4	2	2	4	4	4	5	2	4	2	4	2
MB3_2_238	4	4	2	4	2	5	4	2	4	2	4	4	2	2	2	2	4	4	4	4	4	5	4	4	2	2	2
MB3_2_8	3	3	3	5	1	3	1	3	3	1	1	3	3	3	3	1	1	3	3	1	3	1	1	3	1	5	3
MB3_2_8	3	3	1	5	3	3	1	3	3	1	1	1	3	3	1	1	1	1	1	3	1	1	3	3	3	5	1
MB3_2_172.1S	1	3	3	5	1	1	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	5	1	1	1	1	1
MB3_2_172.1S	3	1	1	5	3	1	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	5	1	1	1	1	1
MB3_2_179.1S	4	4	4	4	4	2	5	4	4	5	4	4	2	2	2	4	4	4	4	2	4	5	4	2	4	4	4
MB3_2_179.1S	2	2	4	2	4	4	5	4	4	5	4	4	2	4	4	4	4	4	4	2	4	5	2	4	4	2	2
MB3_2_2000.1	4	4	4	5	4	4	4	4	4	4	1	4	4	4	4	5	4	4	4	4	4	5	1	4	4	4	1
MB3_2_2000.1	4	4	4	5	4	4	4	4	4	4	4	4	4	1	1	5	4	1	4	4	4	5	4	4	4	4	4
MB3_2_257.1	4	4	4	4	3	4	5	3	4	4	3	4	4	4	4	3	4	4	4	4	4	5	4	4	4	5	4
MB3_2_257.1	4	4	3	4	4	4	5	4	4	4	4	4	4	4	4	4	4	4	3	4	4	5	4	4	4	5	4
MB3_2_788.1S	4	4	4	4	4	4	5	4	4	4	4	4	4	4	4	4	4	5	4	4	4	5	4	4	4	4	4
MB3_2_788.1S	3	3	3	4	3	3	5	3	3	3	3	3	3	3	3	3	3	5	3	3	3	5	3	3	3	3	3
MB3_2_788.3S	2	2	2	2	2	2	5	2	2	2	2	2	2	2	2	2	2	5	2	2	2	5	2	2	2	2	2
MB3_2_788.3S	4	4	4	2	4	4	5	4	4	4	4	4	4	4	4	4	4	5	4	4	4	5	4	4	4	4	4
MB3_2_12.1	1	1	1	1	1	1	5	1	1	5	1	3	1	1	1	5	1	1	1	1	1	5	1	1	1	1	1
MB3_2_12.1	1	1	3	1	3	1	5	1	1	5	1	1	1	1	1	5	1	1	1	1	1	5	1	1	1	1	1
MB3_2_2000.2	4	4	4	5	4	5	4	4	4	4	4	4	4	4	4	5	4	4	4	4	4	5	2	4	4	4	2
MB3_2_2000.2	4	4	4	5	4	5	4	4	4	4	2	4	4	2	2	5	4	2	4	4	4	5	4	4	4	4	4
MB3_2_257.2	4	4	4	4	2	4	5	2	4	5	2	4	4	4	4	2	4	4	4	4	4	5	4	4	4	5	4
MB3_2_257.2	4	4	2	4	4	4	5	4	4	5	4	4	4	4	4	4	4	4	2	4	4	5	4	4	4	5	4
MB3_2_81	3	3	3	3	3	3	3	5	3	5	3	3	3	3	3	3	3	3	3	3	3	5	3	3	5	3	3
MB3_2_81	3	3	3	3	3	3	3	5	3	5	1	3	3	1	3	3	3	3	3	1	1	5	3	3	5	3	1
MB3_2_177	1	1	4	1	4	1	5	1	1	4	5	1	1	1	1	1	4	1	5	4	1	5	5	1	1	1	1
MB3_2_177	1	1	1	1	1	1	5	1	1	4	5	1	1	1	1	1	4	1	5	1	1	5	5	1	1	4	1
MB3_2_259.1	2	2	2	2	2	2	5	2	2	5	2	5	4	2	2	2	4	2	2	2	2	5	2	5	2	4	2
MB3_2_259.1	2	2	4	2	4	2	5	4	4	5	4	5	4	4	4	4	4	2	4	2	2	5	2	5	2	2	4

MB3_2_259.2	1	1	1	1	1	1	5	1	1	5	1	5	3	1	1	1	3	1	3	1	1	5	1	5	1	3	1
MB3_2_259.2	1	1	3	1	3	1	5	3	3	5	3	5	3	3	3	3	3	1	1	1	1	5	1	5	1	1	3
MB3_2_259.3	2	2	2	2	2	2	5	2	2	5	2	5	2	2	2	2	2	2	2	2	2	5	2	5	2	2	2
MB3_2_259.3	2	2	2	2	2	2	5	4	4	5	2	5	2	2	2	2	2	2	2	2	2	5	2	5	2	2	2
MB3_2_102.2S	4	4	4	4	4	4	5	4	4	5	4	5	4	5	4	3	4	4	4	3	3	5	4	4	4	5	4
MB3_2_102.2S	4	4	4	4	4	4	5	4	4	5	4	5	4	5	3	3	4	4	3	4	4	5	4	4	4	5	4
MB3_2_172.2S	1	3	3	5	1	1	5	1	1	1	1	1	5	1	1	1	1	1	1	5	5	5	1	1	1	1	1
MB3_2_172.2S	3	1	1	5	3	1	5	1	1	1	1	1	5	1	1	1	1	1	1	5	5	5	1	1	1	1	1
MB3_2_194.1	2	3	2	5	5	2	2	5	5	5	3	2	2	2	2	3	3	2	2	2	2	5	3	2	3	2	2
MB3_2_194.1	3	2	2	5	5	2	3	5	5	5	2	2	2	2	2	2	2	2	3	2	2	5	2	2	2	2	2
MB3_2_2000.3	4	4	1	5	1	5	5	1	4	1	1	1	1	1	1	5	4	4	5	1	1	5	1	1	1	1	1
MB3_2_2000.3	4	4	4	5	1	5	5	4	1	1	1	1	1	1	1	5	4	1	5	1	1	5	1	1	4	1	1
MB3_2_218.1	2	2	2	2	2	2	5	5	2	2	2	5	4	2	2	2	5	2	2	4	2	5	2	2	2	5	2
MB3_2_218.1	2	2	2	2	2	4	5	5	2	2	2	5	4	4	4	2	5	2	2	2	4	5	2	2	2	5	2
MB3_2_132.1	4	4	4	4	4	4	5	4	4	5	4	4	5	5	4	4	4	4	5	4	4	5	4	4	4	4	4
MB3_2_132.1	4	4	3	4	4	4	5	4	4	5	4	3	5	5	3	3	4	3	5	4	4	5	4	4	4	4	4
MB3_2_194.2	3	4	3	5	5	3	3	5	5	5	4	3	3	3	3	3	4	3	3	3	3	5	4	5	4	3	3
MB3_2_194.2	4	3	3	5	5	3	4	5	5	5	3	3	3	3	3	4	3	3	4	3	3	5	3	5	3	3	3
MB3_2_43	2	2	2	2	2	2	5	5	5	5	5	5	2	2	2	2	2	2	2	2	2	5	2	5	4	2	2
MB3_2_43	2	2	2	2	2	2	5	5	5	5	5	5	2	2	2	2	2	2	2	2	2	4	5	2	5	4	2	2
MB3_2_179.2S	4	4	4	4	4	4	5	4	4	5	4	5	2	5	2	4	5	4	4	5	4	5	4	4	5	5	4
MB3_2_179.2S	4	4	4	2	4	2	5	4	4	5	4	5	2	5	4	4	5	4	4	5	4	5	2	4	5	5	4
MB3_2_194.3	4	2	4	5	5	4	5	5	5	5	2	4	4	4	4	4	2	4	4	4	4	5	4	5	2	4	2
MB3_2_194.3	2	4	4	5	5	4	5	5	5	5	2	4	4	4	4	2	4	4	2	4	2	5	2	5	4	4	4
MB3_2_132.2	4	4	2	4	2	4	5	4	4	5	4	4	5	5	2	4	4	2	5	5	5	5	4	4	5	4	4
MB3_2_132.2	2	2	2	4	2	4	5	4	4	5	4	4	5	5	2	2	4	2	5	5	5	5	2	4	5	4	4
MB3_2_132.3	2	2	2	2	2	2	5	2	2	5	2	2	5	5	2	2	2	2	5	5	5	5	2	2	5	2	2
MB3_2_132.3	2	2	4	2	2	2	5	2	2	5	2	2	5	5	2	4	2	4	5	5	5	5	2	2	5	2	2
MB3_2_112	4	5	2	5	5	2	2	4	2	5	4	2	4	2	2	5	5	5	5	5	4	5	2	2	4	4	4
MB3_2_112	4	5	2	5	5	2	2	4	4	5	4	2	2	2	4	5	5	5	5	5	4	5	2	4	4	4	4
MB3_2_132.4	2	2	1	2	2	2	5	2	2	5	2	5	5	5	2	2	2	2	5	5	5	5	2	2	5	2	2
MB3_2_132.4	2	2	2	2	2	2	5	2	2	5	2	5	5	5	2	1	2	1	5	5	5	5	2	2	5	2	2
MB3_2_204	1	1	1	1	1	1	5	1	5	5	5	5	5	5	1	1	1	1	1	5	1	5	1	5	5	1	1
MB3_2_204	2	2	1	2	2	1	5	1	5	5	5	5	5	5	2	1	1	1	1	5	1	5	1	5	5	1	1

MB3_2_259.2	1	1	1	1	1	1	5	1	1	5	1	5	3	1	1	1	3	1	3	1	1	5	1	5	1	3	1
MB3_2_259.2	1	1	3	1	3	1	5	3	3	5	3	5	3	3	3	3	3	1	1	1	1	5	1	5	1	1	3
MB3_2_259.3	2	2	2	2	2	2	5	2	2	5	2	5	2	2	2	2	2	2	2	2	2	5	2	5	2	2	2
MB3_2_259.3	2	2	2	2	2	2	5	4	4	5	2	5	2	2	2	2	2	2	2	2	2	5	2	5	2	2	2
MB3_2_102.2S	4	4	4	4	4	4	5	4	4	5	4	5	4	5	4	3	4	4	4	3	3	5	4	4	4	5	4
MB3_2_102.2S	4	4	4	4	4	4	5	4	4	5	4	5	4	5	3	3	4	4	3	4	4	5	4	4	4	5	4
MB3_2_172.2S	1	3	3	5	1	1	5	1	1	1	1	1	5	1	1	1	1	1	1	5	5	5	1	1	1	1	1
MB3_2_172.2S	3	1	1	5	3	1	5	1	1	1	1	1	5	1	1	1	1	1	1	5	5	5	1	1	1	1	1
MB3_2_194.1	2	3	2	5	5	2	2	5	5	5	3	2	2	2	2	3	3	2	2	2	2	5	3	2	3	2	2
MB3_2_194.1	3	2	2	5	5	2	3	5	5	5	2	2	2	2	2	2	2	2	3	2	2	5	2	2	2	2	2
MB3_2_2000.3	4	4	1	5	1	5	5	1	4	1	1	1	1	1	1	5	4	4	5	1	1	5	1	1	1	1	1
MB3_2_2000.3	4	4	4	5	1	5	5	4	1	1	1	1	1	1	1	5	4	1	5	1	1	5	1	1	4	1	1
MB3_2_218.1	2	2	2	2	2	2	5	5	2	2	2	5	4	2	2	2	5	2	2	4	2	5	2	2	2	5	2
MB3_2_218.1	2	2	2	2	2	4	5	5	2	2	2	5	4	4	4	2	5	2	2	2	4	5	2	2	2	5	2
MB3_2_132.1	4	4	4	4	4	4	5	4	4	5	4	4	5	5	4	4	4	4	5	4	4	5	4	4	4	4	4
MB3_2_132.1	4	4	3	4	4	4	5	4	4	5	4	3	5	5	3	3	4	3	5	4	4	5	4	4	4	4	4
MB3_2_194.2	3	4	3	5	5	3	3	5	5	5	4	3	3	3	3	3	4	3	3	3	3	5	4	5	4	3	3
MB3_2_194.2	4	3	3	5	5	3	4	5	5	5	3	3	3	3	3	4	3	3	4	3	3	5	3	5	3	3	3
MB3_2_43	2	2	2	2	2	2	5	5	5	5	5	5	2	2	2	2	2	2	2	2	2	5	2	5	4	2	2
MB3_2_43	2	2	2	2	2	2	5	5	5	5	5	5	2	2	2	2	2	2	2	2	4	5	2	5	4	2	2
MB3_2_179.2S	4	4	4	4	4	4	5	4	4	5	4	5	2	5	2	4	5	4	4	5	4	5	4	4	5	5	4
MB3_2_179.2S	4	4	4	2	4	2	5	4	4	5	4	5	2	5	4	4	5	4	4	5	4	5	2	4	5	5	4
MB3_2_194.3	4	2	4	5	5	4	5	5	5	5	2	4	4	4	4	4	2	4	4	4	4	5	4	5	2	4	2
MB3_2_194.3	2	4	4	5	5	4	5	5	5	5	2	4	4	4	4	2	4	4	2	4	2	5	2	5	4	4	4
MB3_2_132.2	4	4	2	4	2	4	5	4	4	5	4	4	5	5	2	4	4	2	5	5	5	5	4	4	5	4	4
MB3_2_132.2	2	2	2	4	2	4	5	4	4	5	4	4	5	5	2	2	4	2	5	5	5	5	2	4	5	4	4
MB3_2_132.3	2	2	2	2	2	2	5	2	2	5	2	2	5	5	2	2	2	2	5	5	5	5	2	2	5	2	2
MB3_2_132.3	2	2	4	2	2	2	5	2	2	5	2	2	5	5	2	4	2	4	5	5	5	5	2	2	5	2	2
MB3_2_112	4	5	2	5	5	2	2	4	2	5	4	2	4	2	2	5	5	5	5	5	4	5	2	2	4	4	4
MB3_2_112	4	5	2	5	5	2	2	4	4	5	4	2	2	2	4	5	5	5	5	5	4	5	2	4	4	4	4
MB3_2_132.4	2	2	1	2	2	2	5	2	2	5	2	5	5	5	2	2	2	2	5	5	5	5	2	2	5	2	2
MB3_2_132.4	2	2	2	2	2	2	5	2	2	5	2	5	5	5	2	1	2	1	5	5	5	5	2	2	5	2	2
MB3_2_204	1	1	1	1	1	1	5	1	5	5	5	5	5	5	1	1	1	1	1	5	1	5	1	5	5	1	1
MB3_2_204	2	2	1	2	2	1	5	1	5	5	5	5	5	5	2	1	1	1	1	5	1	5	1	5	5	1	1

MB3_2_188.1	4	4	4	4	4	2	5	2	2	5	5	5	2	5	2	2	2	5	2	2	5	5	5	5	5	5	2
MB3_2_188.1	4	4	4	2	4	2	5	2	4	5	5	5	2	5	2	2	2	5	2	2	5	5	5	5	5	5	4
MB3_2_188.2	1	1	1	1	1	3	5	1	3	5	5	5	3	5	3	3	1	5	3	1	5	5	5	5	5	5	1
MB3_2_188.2	1	1	1	1	1	3	5	1	1	5	5	5	3	5	3	3	1	5	3	3	5	5	5	5	5	5	1
MB3_2_64.1	3	3	3	3	3	5	5	5	3	5	3	3	5	3	3	5	3	3	5	5	3	5	3	3	5	5	5
MB3_2_64.1	3	3	3	3	3	5	5	5	3	5	3	3	5	1	3	5	1	3	5	5	3	5	3	3	5	5	5
MB3_2_64.2	3	3	3	3	3	5	5	5	3	5	3	3	5	3	3	5	3	3	5	5	3	5	3	3	5	5	5
MB3_2_64.2	2	3	3	3	3	5	5	5	2	5	2	3	5	2	2	5	2	3	5	5	2	5	3	3	5	5	5
MB3_2_194.4	1	4	1	5	5	1	5	5	5	5	4	1	1	5	1	5	5	1	1	1	1	5	4	5	1	5	4
MB3_2_194.4	1	1	1	5	5	1	5	5	5	5	4	1	1	5	1	5	5	1	4	1	1	5	1	5	4	5	1

Appendix 4. Table of all performed 'no missing data matches' and their outcomes.

Evidence Sample	Comparison Sample	# Loci				Overall match (%)
		Total	2 allele match	1 allele match	0 allele match	
LD210072*	LD210072*	114	99 (86.8%)	15 (13.2%)	0 (0%)	93.4†
LD210071	LD210070‡	100	97 (97%)	3 (3%)	0 (0%)	98.5§
LD210071	LD210072	100	55 (55%)	40 (40%)	5 (5%)	75
LD210071	LD210073	89	64 (71.9%)	19 (21.3%)	6 (6.7%)	82.6
LD210072	LD210071	100	55 (55%)	40 (40%)	5 (5%)	75
LD210072	LD210073	89	48 (53.9%)	33 (37.1%)	8 (9%)	72.5
LD210071	JS79	96	57 (59.4%)	33 (34.4%)	6 (6.3%)	76.6
LD210071	JS81	94	53 (56.4%)	38 (40.4%)	3 (3.2%)	76.6
LD210071	JS83	61	38 (62.3%)	19 (31.1%)	4 (6.6%)	77.9
LD210071	JS84	91	51 (56%)	29 (31.9%)	11 (12.1%)	72
LD210071	JS85	94	55 (58.5%)	36 (38.3%)	3 (3.2%)	77.7
LD210071	JS86	76	52 (68.4%)	16 (21.1%)	8 (10.5%)	78.9
LD210071	JS87	95	56 (58.9%)	34 (35.8%)	5 (5.3%)	76.8
LD210071	JS88	89	56 (62.9%)	26 (29.2%)	7 (7.9%)	77.5
LD210071	JS89	91	50 (54.9%)	27 (29.7%)	14 (15.4%)	69.8
LD210071	JS90	90	56 (62.2%)	28 (31.1%)	6 (6.7%)	77.8
LD210071	JS91	100	48 (48%)	44 (44%)	8 (8%)	70
LD210071	JS92	93	55 (59.1%)	32 (34.4%)	6 (6.5%)	76.3
LD210071	JS93	97	55 (56.7%)	34 (35.1%)	8 (8.2%)	74.2
LD210071	JS94	95	49 (51.6%)	42 (44.2%)	4 (4.2%)	73.7
LD210071	JS95	92	53 (57.6%)	32 (34.8%)	7 (7.6%)	75
LD210071	JS96	92	54 (58.7%)	33 (35.9%)	5 (5.4%)	76.6
LD210071	JS97	94	62 (66%)	29 (30.9%)	3 (3.2%)	81.4
LD210071	JS98	56	34 (60.7%)	20 (35.7%)	2 (3.6%)	78.6
LD210071	JS99	97	64 (66%)	30 (30.9%)	3 (3.1%)	81.4
LD210071	JT01	88	57 (64.8%)	24 (27.3%)	7 (8%)	78.4
LD210071	JT02	90	54 (60%)	29 (32.2%)	7 (7.8%)	76.1
LD210071	JT03	87	46 (52.9%)	36 (41.4%)	5 (5.7%)	73.6
LD210071	JT04	98	57 (58.2%)	31 (31.6%)	10 (10.2%)	74
LD210072	JS79	96	85 (88.5%)	11 (11.5%)	0 (0%)	94.3§
LD210072	JS81	95	49 (51.6%)	40 (42.1%)	6 (6.3%)	72.6
LD210072	JS83	62	29 (46.8%)	28 (45.2%)	5 (8.1%)	69.4
LD210072	JS84	92	50 (54.3%)	35 (38%)	7 (7.6%)	73.4
LD210072	JS85	95	60 (63.2%)	30 (31.6%)	5 (5.3%)	78.9
LD210072	JS86	76	40 (52.6%)	34 (44.7%)	2 (2.6%)	75
LD210072	JS87	96	50 (52.1%)	40 (41.7%)	6 (6.3%)	72.9
LD210072	JS88	90	55 (61.1%)	32 (35.6%)	3 (3.3%)	78.9
LD210072	JS89	92	46 (50%)	39 (42.4%)	7 (7.6%)	71.2

LD210072	JS90	91	42 (46.2%)	46 (50.5%)	3 (3.3%)	71.4
LD210072	JS91	101	50 (49.5%)	47 (46.5%)	4 (4%)	72.8
LD210072	JS92	93	49 (52.7%)	37 (39.8%)	7 (7.5%)	72.6
LD210072	JS93	97	48 (49.5%)	39 (40.2%)	10 (10.3%)	69.6
LD210072	JS94	95	62 (65.3%)	30 (31.6%)	3 (3.2%)	81.1
LD210072	JS95	92	48 (52.2%)	36 (39.1%)	8 (8.7%)	71.7
LD210072	JS96	92	48 (52.2%)	39 (42.4%)	5 (5.4%)	73.4
LD210072	JS97	95	54 (56.8%)	36 (37.9%)	5 (5.3%)	75.8
LD210072	JS98	56	35 (62.5%)	20 (35.7%)	1 (1.8%)	80.4
LD210072	JS99	98	48 (49%)	47 (48%)	3 (3.1%)	73
LD210072	JT01	89	51 (57.3%)	35 (39.3%)	3 (3.4%)	77
LD210072	JT02	91	50 (54.9%)	31 (34.1%)	10 (11%)	72
LD210072	JT03	88	50 (56.8%)	32 (36.4%)	6 (6.8%)	75
LD210072	JT04	99	46 (46.5%)	47 (47.5%)	6 (6.1%)	70.2

* Repeated extraction results compared before inconsistent loci removed

† Removed from further comparisons to avoid pseudoreplication

‡ Known match

§ Concluded match

Appendix 5. Frequency tables for 100,000 simulations of percentages matches between reference samples and A) LD210071, B) LD210072.

A

Bin Mid Point	Counts	Percentage Total	Cumulative Frequency
63	1	0.001	0.001
64	5	0.005	0.006
65	47	0.047	0.053
66	146	0.146	0.199
67	436	0.436	0.635
68	1117	1.117	1.752
69	2726	2.726	4.478
70	5173	5.173	9.651
71	8701	8.701	18.352
72	12408	12.408	30.76
73	15258	15.258	46.018
74	15960	15.96	61.978
75	14413	14.413	76.391
76	10665	10.665	87.056
77	6701	6.701	93.757
78	3724	3.724	97.481
79	1644	1.644	99.125
80	602	0.602	99.727
81	195	0.195	99.922
82	52	0.052	99.974
83	20	0.02	99.994
84	5	0.005	99.999
85	1	0.001	100

B

Bin Mid Point	Counts	Percentage Total	Cumulative Frequency
61	6	0.006	0.006
62	36	0.036	0.042
63	106	0.106	0.148
64	395	0.395	0.543
65	1099	1.099	1.642
66	2604	2.604	4.246
67	5266	5.266	9.512
68	8888	8.888	18.4
69	12924	12.924	31.324
70	15689	15.689	47.013
71	16302	16.302	63.315
72	14212	14.212	77.527
73	10653	10.653	88.18
74	6409	6.409	94.589
75	3245	3.245	97.834
76	1429	1.429	99.263
77	516	0.516	99.779
78	174	0.174	99.953
79	39	0.039	99.992
80	7	0.007	99.999
81	1	0.001	100

Appendix 6. Details of individuals and loci analysed with Sanger sequencing

Individuals
LD210070
LD210071
LD210072
LD210073
JS79*
JS80
JS89
JS93
Negative Control

*Sample sequenced twice at each locus to check for consistency in results

Loci	Notes
MB3_2_60	100% match
MB3_2_77	Insufficient coverage
MB3_2_88	Insufficient coverage
MB3_2_97	Insufficient coverage
MB3_2_102	100% match
MB3_2_172.1	Possible paralogous loci
MB3_2_172.2	Possible paralogous loci
MB3_2_179	Indication of allele dropout in Sanger
MB3_2_265.2	100% match
MB3_2_265.3	100% match
MB3_2_788.2	Insufficient coverage
MB3_2_788.3	100% match
MB3_2_827	Indication of allele dropout in CRoPS

Thesis End