A METHOD FOR THE MOLECULAR CLASSIFICATION OF POLLEN FOR FORENSIC PURPOSES

By

Rebecca Curulli

A thesis submitted in fulfilment of the requirements for the degree of

Master of Forensic Science (Professional Practice)

in

The School of Medical, Molecular and Forensic Sciences

Murdoch University

Brendan Chapman

2022

Declaration

I, Rebecca Curulli,

declare that this thesis does not contain any material submitted previously for the award of any other degree or diploma at any university or other tertiary institution. Furthermore, to the best of my knowledge, it does not contain any material previously published or written by another individual, except where due reference has been made in the text. Finally, I declare that all reported experimentations performed in this research were carried out by myself, except that any contribution by others, with whom I have worked is explicitly acknowledged.

Acknowledgements

I would like to thank my supervisor, Brendan Chapman for his patience and continued support in not only this thesis project, but also the entirety of the Masters course. Thank you for always pushing us academically, and allowing us the opportunity to learn what we need to be great forensic investigators.

To the amazing and incredible students I have met while completing my masters degree; I admire your intelligence, passion and dedication to your chosen areas within forensic science, and while I know life will separate us for a while, I cannot wait to cross paths with everyone again within our chosen fields. I am proud to have made future colleagues and lifelong friends within you all.

Additionally, I would like to thank my family for their unwavering support and encouragement throughout my years at university. To my parents, Vince and Rhonda, for raising me to be independent and headstrong, and my brothers Jacob and Jordan, for your reassurance and guidance whenever I needed it.

Lastly, to my husband Rhys; your unconditional love and support over the last five years at university completing my bachelor and masters degree has been unwavering. You have listened to all of my ideas, my thoughts, read my papers, even the things you never understood. You have pushed me to accomplish goals and dreams you always knew I could achieve, and been there to cheer me every step of the way. With all my heart, thank you.

Table of Contents

Title Page i

Declaration ii

Acknowledgements iii

Part One

Literature Review.....1-37

Part Two

Manuscript.....39-54

Blank Page – not numbered

Part One

Literature Review

A method for the molecular classification of pollen for forensic purposes

ABSTRACT

Forensic palynology and the identification of pollen grains using DNA barcoding is an underutilised field in criminal investigations. Successful identification of pollen grains is used to establish links between an offender or body and location-related crimes. The use of palynology in criminal investigations has been limited due to slow speed of identification, limited expert specialization and poor taxonomic resolution. To fill the gap, DNA barcoding overcomes these limitations by offering taxonomic resolution down to species in a faster, more accessible way. Despite this, there are technical applications that need to be addressed before DNA barcoding can be implemented as a laboratory standard for forensic palynology. This study aims to address these gaps, and highlight the need to make DNA barcoding more accessible in the future of forensic palynology.

TABLE OF CONTENTS

DECLARATION ii

ACKNOWLEDGEMENTS iii

ABSTRACT 1

TABLE OF CONTENTS 2

List of figures 4

List of tables 5

List of abbreviations 6

List of definitions 7

PART ONE: Examination of the literature 8

1. Introduction 8

1.1 History of Forensic Palynology 9

1.2 Pollen in Forensics 12

1.3. DNA Barcoding 13

1.4 Challenges in DNA Barcoding 16

2. Discussion 19

2.1 DNA Barcoding Markers 19

2.2 Qiagen Extraction Kit 23

2.3 Pollen in Forensics: Applications 26

2.4 Gaps in the Literature 29

Experimental Design 34

Conclusion 36

PART TWO: Exploratory Study 39

ABSTRACT 40

Keywords 40

Research Aims 41

Introduction 42

Methodology 46

Collection of Eucalyptus pods 46

DNA Extraction and Quantification 47

Data analysis 18

Conclusion 53

REFERENCES 55

List of figures

Figure 1: Map representation of Australian in 2013 showing distribution of Eucalyptus native forest [26].

List of tables

Table 1 – Published primer information for chloroplast DNA for trnL loci with reference to location on Eucalyptus globulus genome

Table 2: Summary of published studies using DNA barcoding methods for pollen species identification

Table 3: Qubit 3.0 Fluorometer measurements of DNA yield from Eucalyptus species extracted using Qiagen DNeasy Pro Mini Kit

List of abbreviations

Abbreviation	Meaning
HTS	High Thoroughput Sequencing
ITS2	Internal transcribed spacer 2
Mat K	Maturase K
NGS	Next Generation Sequencing
PCR	Polymerase Chain Reaction
psbA-trnH	Intergenic spacer region trnH-psbA
rbcL	Ribulose biphosphate carboxylase
SEM	Scanning Electron Microscopy
SNP	Single Nucleotide Polymorphism
STR	Short Tandem Repeat
TLM	Transmitted Light Microscopt
trnL	Chloroplast intron region of tRNA gene

List of definitions

Word	Definition
Angiosperm	Plant group that produces flowers, seeds within a carpel including most trees, grasses, shrubs and herbaceous plants
Assemblage	An accumulation or grouping
Bryophytes	Non-vascular land plants comprising liverworts, hornworts and mosses
Electropherogram	Plot results from an analysis of automated electrophoresis sequencing
Endemic	In relation to plant species; native, restricted to an area
Exine	The hard outer wall of a pollen grain
Geolocation	The particular location in reference to tracking of an animal or plant species
Locus	A specific or fixed position on a chromosome where a gene or genetic marker is located. Plural = loci.
Palynomorph	Organic microfossil extracted from sedimentary rocks composed of pollen and spores
Phenotype	An observable trait or characteristic of an organism
Phylogeography	The study of spatial arrangement of genealogical lineages in closely related species
Short Tandem Repeat	Short, repeated nucleotide sequences
Single Nucleotide Polymorphism	A DNA sequence variation that occurs within a single nucleotide which alters the genome sequence
Taxonomy	A branch of science used to classify organisms
Gymnosperm	Seed producing plants that encompass conifers, cycads, Ginkgo and gnetophytes.
Cryptogram	A species of plant or plant-like organism that reproduces by spores without the need for flowers or seeds. Non-seed bearing plants.

PART ONE – Examination of the literature

Introduction

1. Introduction

Palynology refers to the studies of the origin and structure of pollen grains [1] encompassing many fields across a geographical and archaeological context [2-3]. It has also been used in the fields of paleoecology, palaeontology and archaeology [4]. Advances in recent plant genomics have allowed it to move into the field of forensics, using pollen as a DNA profiling tool within criminal investigations [3]. There is an estimated range of over 300,000-400,000 plant species worldwide [5] and studying the grains and spores, or pollen assemblage in each area allows for unique pollen fingerprints in different geolocations [1, 6]. Pollen provides great potential within forensic palynology to help establish a link between a person, object, place, or time in criminal investigations [1, 4]. The spores can be transferred by physical contact with the plant onto a person or surface, animal transport, or airborne spores by wind and insect transport sometimes as far as several hundred meters [1]. Palynological evidence can help establish supportive links and relationships between suspects and crime scenes [7], making them a high interest in forensic discipline and crime scene investigations [3]. One way of establishing a link is by documenting known pollen assemblage within the area. Charting yearly pollen analysis allows for spatial and temporal tracing of plant species [7], contributing to a record known as a pollen calendar, which can narrow down time or season of death by comparing the pollen fingerprint to where the body was found [8]. Pollen can also be used in forensic archaeology, prevailing for many thousands or even millions of years surviving in honey, charcoal, or evidence of regular fires and even residing in the intestine for up to 21 days [2,8].

The application of forensic palynology within investigations has been limited within studies, propelled by the need for expertly trained scientists within a specialised field who can accurately identify pollen grains by traditional microscopic methods [2]. An alternative to this is to create the potential to classify pollen taxonomically through its molecular characteristics or genetic structure known as DNA barcoding. Offering DNA barcoding as a standard to forensic laboratories would change forensic palynology, allowing it to progress as a science and expanding the field in new and challenging ways. This thesis aims to review the history of forensic palynology, the use of pollen in forensics, DNA barcoding of pollen, and its challenges and potential for future research.

1.1 History of forensic palynology

Forensic palynology has been used as a tool in criminal investigations since the 1950s with more traditional methods involving transmitted light microscopy (TLM) and scanning electron microscopy (SEM) to identify pollen by phenotype using morphology [2]. The images are captured by light microscopy to identify unique features of the pollen grain, such as size, shape, colour, and outer structure of the grain wall known as the exine [6]. They are compared by an expert against a sourced pollen image library for the identification of plant families [7]. Identification by microscopy has low taxonomic resolution, is time-consuming, labour intensive, and difficult to perform continuous routine analysis [2, 9]. As pollen is present in every environment worldwide, it can be invaluable in investigations, however; the field of forensic palynology has its limitations. The number of experts who can accurately identify pollen to a species level is small and dwindling [10]. Traditional methods based on morphological criteria can be difficult to apply due to the subjective bias of the expert [5], and therefore it is recommended that

palynologists are dubious of their identifications, particularly in unfamiliar geographical locations [8]. There are multiple biomes within the world; each with their own rich pollen assemblage.

Under a microscope some of the morphological characteristics of the exine are variable at higher taxonomic levels [10], and differences between pollen of similar species can be subtle leading to classification errors [6]. In an attempt to combat this, there has been experimentation with automated databases such as the POLLEN73S program for pollen grain image identification, but the field lacks information, sequencing, and technologies required for reasonable accuracy. For example, pollen grain types like *Manihot esculenta* have primitive characteristics with a predominantly rounded shape and accentuated texture – and within the POLLEN73S program yield a low precision rate for identification [6].

Within palynology newer technology offers the potential for more accurate taxonomic resolution for pollen grain identification, opening the field up to forensic scientists of a qualified nature to take on the role of forensic palynologist.

1.2 Pollen in forensics

Pollen grains will often present as a mixture, and isolating enough individual grains for genus-level identification can be challenging depending on sample size [8-9]. Pollen can be found at different types of crime scenes and objects, passed through direct contact or as trace evidence, or recovered from a body. Due to the relatively small size, pollen grains can settle into hard to trace items, such as the interior of undetonated explosive devices,

paper documents and computers [9]. Grains can also be collected from ropes, soil, clothing, drugs, filters, animal and human fur and even stomach contents [2]. Pollen can be recovered as evidence using sterile swabs, adhesive tapes, vacuuming, pinch method or washing directly with sterile water off the object or body. They are collected into clean, tamper-evident bags or containers and sealed for analysis [7]. It is important to preserve the integrity of the evidence within forensic investigations.

Common at every crime scene, pollen is often found in dust and dirt [3], carried into homes on clothes and shoes or entering via open windows and doors [11]. The grains can be used in a multitude of ways within forensic investigations, and due to its durability, it can be utilised even decades after sample collection making it ideal for archeological forensics and cold cases [9]. Still, the most valuable tool it offers is being used to trace direct contact and establish a link between an offender to a crime scene or object [2, 4, 7].

Pollen grains retain a useful ability to adhere within the folds of the skin or the clothes and soles of shoes [2], ideal considering that environmental pollen constitutes 94% of the pollen count on a body [8]. When a body is located in the outer environment, consultation of a pollen calendar, if one has been conducted and recorded, can suggest the season that the body was placed within the grave. The pollen assemblage recovered from multiple sources can be useful in narrowing a search to locate a clandestine grave, scene or body [7-8], any of which could yield a secondary scene producing more forensic evidence. Studies have shown that different soil samples collected within a localised region even up to fifteen meters display similar pollen assemblages, but there are still discernable differences even in ranges up to one kilometre of the same geographical

location [4]. A set of bushes will have its distinctive pollen assemblage that can be traceable, showing the uniqueness of each individual area [4]. This displays the importance of knowing pollen distributions in each topographical environment as a representative for comparison. This is also useful knowledge for environmental preservation and ecological forensics, it will also provide a reference point if any pollen is located within clandestine burial grounds and gravesites, particularly in archeological forensics.

When analysing pollen as a forensic investigator, its important to consider factors that could affect the spatial dispersion of the grain. Depending on the flowering parent, the dispersal mechanism will vary between species, as well as pollinators, insects and weather patterns spreading airborne grains [2, 4]. The location of the particular flora, and the surrounding assemblage needs to be contemplated. Is there a larger selection of non-flowering trees that create canopies and limit wind dispersal? Are the pollen yields from shrubs that are close to the ground that can be dispersed by animal traffic or potentially a hiking and running trail? Other factors include height of release, wind strength and direction, weight and shape of pollen grain, aerodynamics of the grain, vulnerability to destruction, and usual method of pollination for that species [7].

Pollen is made of microscopic particles called microgametophytes that house sperm cells that are produced in the anther of seed-producing plants and transported to the stigma of related female plants for reproduction [7]. The pollen grain can be transferred between two different plants known as cross-pollination, or from the anther to the stigma of the same plant in self-pollination. The anatomy of the pollen grain consists of two cells, a generative (productive) cell, and a vegetative (non-productive) cell within a

two-layered wall [12]. The innermost layer is known as the intine and consists of cellulose and protein, the outer layer or exine is mostly sporopollenin and protects the pollen from physical and chemical stressors [7]. Pollen grains contain haploid DNA, which is protected by the exine making it resistant to chemical and environmental changes and each grain size ranges from 5 to 200 micrometres in diameters [9].

Extracting plant DNA is more difficult compared to human DNA and usually requires a mechanical method with a chemical component to break down the tough exine wall. This can be done by disruption of the pollen exine with pulverisation techniques using bead beating, chemicals like tissueLyster, mortar and pestle, or liquid nitrogen freezing [9, 13]. There are a several kits commercially available for the process such as Qiagen DNeasy® Plant Kit, Omega Bio-tek E.Z.N.A® Plant DS Mini Kit and NucleoSpin® Food Kit. The standard kit function uses solid-phase nucleic acid purification by spin column using centrifugal force, which then yields an efficient DNA purification [14]. DNA can be recovered from a single pollen grain [2, 9] so considering forensic analysis and destruction of samples, if the recovery is small as long as care is taken to keep the exine intact, the grains can be preserved. A secondary alternative is to take multiple high-resolution photographs of the pollen grain before DNA analysis to allow for digital preservation of the sample.

1.3 DNA Barcoding

DNA barcoding (hereafter barcoding) is not a relatively new concept but is now only recently being explored within palynology and how that can be applied to forensic investigations. The technique aims to characterise pollen using a short DNA sequence or

genetic marker from a universal standard in the genome [15]. Barcoding has been well researched for its applications including ecological forensics, conservation assessments, potential to discover unnamed land plants in tropical biomes, endangered species assessments and monitoring of commercial food and herbal supplements [9, 16]. Within forensic investigations, by making barcoding readily available to forensic laboratories, it would increase taxonomic resolution of identification of pollen species [2, 10] and reduce subjective bias [5]. A secondary advantage of barcoding over traditional microscopy techniques is that it would allow for taxonomic identification of areas of the organism that aren't displayed in the morphological characteristics down to genus or species level [2, 10]. Barcoding can be used across multiple taxonomic groups, allowing species classification without the need for multiple experts, making it more universal and easily accessible. Unlike microscopy, which requires multiples of the same samples for comparison, barcoding can be done with a small sample of tissue from sterile, juvenile or fragmented material where morphological identification has proven difficult [17] but it is limited within forensic investigations in its destruction of the sample.

For barcoding to be successful three things are required [2], the first being high throughput sequencing (HTS) or (next generation sequencing) NGS sequencing. Both with advantages, for DNA barcoding NGS is the preferred method of choice. Originally used, Sanger sequencing provided single sequencing pattern or electropherograms for each sequence, compared to the newer technology of NGS which allows for the sequencing of millions of DNA fragments from thousands of DNA templates in parallel [18]. Sanger also requires a higher concentration of template DNA (100-500ng), but NGS is ideal as it more efficient in comparison due to its simplicity, cost-effectiveness, ability to process more

fragments per run and faster throughout making it ideal for DNA barcoding [18]. The success of DNA barcoding relies on a well-chosen PCR sequencing strategy and good purification and yield of chloroplast DNA [19].

Secondly, a set of genetic markers for replication and sequencing is needed, preferably a set of loci that has good universality, discriminatory power and can distinguish between species [9-10]. In forensics, the standard method of genetic analysis utilises high-throughput sequencing and PCR analysis using short tandem repeats (STRs) or single nucleotide polymorphisms (SNPs) for identification of human individuals, in which the same can be applied to plants [10]. Both chloroplast and nuclear DNA can be amplified from pollen [9], and the literature shows extensive research into the gene region of the mitochondrial, plastid and nuclear genomes of plants. Research has revealed five chloroplast regions that have been widely accepted as DNA markers, those of which are known as matK, rbcL, psbA-trnH, trnL and ITS2 [2, 9, 16-17]. This first consensus for a standard set of genetic markers was published at the end of 2009 [17]. Since then multiple studies [5, 17-18,20-21,24,30,35,40] have attempted to analyse further markers, or find the best loci combination and have had isolated success.

Taberlet 1991 [20] published standard primers covering a large section of known chloroplasts – primarily ordered along tobacco *Nicotiana tabaccum* genome as that was the best characterized chloroplast genome to date [19].

In terms of the third component, a database that contains both elements with DNA reference to compare them to such as the Australian Plant census, or online databases like GenBank, BOLD, NCBI is needed [9-10, 21-22]. This is currently the most limiting

aspect of DNA barcoding as there is no universal database that encompasses all herbal literature worldwide. Reference databases such as Barcode Of Life Datasystems (BOLD: http://boldsystems.org) and the International Nucleotide Sequences Database Collaboration (NCBI: http://www.ncbi.nln.nih.gov/genbank/), which encompasses GenBank, only contains a small portion of barcode sequences for plant diversity worldwide. This is due to time-consuming methods of data generation, analysis, storage and availability of processing available plant species. Compiling the worldwide herbal literature, as well as genomic sequences into one unified database is a task that has not yet been undertaken but has been considered and mentioned within the literature [1, 9-10].

The databases are limited in their sequences as application and upload of sequence data is supplied by each individual scientist, and therefore relies on the responsibility of that data to be solely accurate with no programs available for verification [22]. This creates a risk for misidentification in sequences, and those programs that do have authentication factors available have a much lower number of sequences accessible. Even with these three components, DNA barcoding has not yet been incorporated into forensic palynology as a standard due to the many limitations within the field, which recent studies [1, 9-10] are working to overcome.

1.4 Challenges in DNA Barcoding

The first challenge with DNA barcoding in plants have been searching for a marker comparable to the mitochondrial CO1 gene in animal groups [9, 17]. There has been some disagreement amongst the community upon which marker is singularly ideal for PCR analysis. An ideal barcode should provide maximal taxonomic recovery amongst species,

be retrievable with a single primer pair, and have good discriminatory power [9, 17]. However, using a single locus for chloroplast regions has shown poor taxonomic resolution leading to various studies attempting multi-locus approaches [9, 10, 16-17]. The currently debate amongst the scientific community regards the best selected core loci of choice for plant barcoding, primer universality and boosting discriminatory power to achieve better results [5]. To combat the limitations of the single locus approach, a secondary marker can be added, however there is a lack of agreement on the best choice of marker combinations [17]. Adding a second loci improves discrimination power but each added loci increases costs of the project [16], as well as creating issues with a common set of nuclear markers including primer site mutations and recombination [21].

There is a lack of comparative data in the literature encompassing all of the markers with a broad taxonomic sample to allow for comparison, given the large diversity between species of land plants [9, 21]. When comparing the sample, there is also a lack of comprehensive databases available and without automated programs available sequences often need to be matched by eye [19]. Furthermore, within herbal literature there is a need for standardised and more accurately defined taxonomy and common names for universality [16].

Within the laboratory, there are many methods of DNA extraction that could have variability of extraction efficiency. Contamination in any scientific field is a probable issue, and due to the small nature of pollen, cross contamination could potentially skew results [9]. Alternatively, if pollen is treated with a compound to eliminate non-pollen plant

material and avoid contamination issues, if the researcher isn't careful they could eliminate the ptDNA.

There are also issues that may arise related specifically to forensic investigations and DNA barcoding. These include a lack of consolidated geographical data on plant distribution and pollination times, technical challenges with recovery of a small number of pollen grains, and that the sequencing process destroys the sample and therefore cannot be retained as evidence in an investigation [9]. When evidence is gathered in a forensic investigation, it is assumed all information regarding that evidence must be accurate and truthful. The use of pollen as a biomarker to identify geolocation, spatial distance or place a person/body at the scene of the crime requires knowledge of where the pollen originated. This includes species, growth, dispersal rate, amount of pollen produce, seasonal bloom (when it actively produces pollen), durability of the grain, and uniqueness of the species [7, 9]. Development of location charts or seasonal pollen calendars would help to assist forensic investigations and provide a reference point for future studies, but due to the expansive nature of seed-producing plants it's a sizeable task that hasn't yet been undertaken. DNA barcoding shows great potential to allow forensic palynology to become a more accurate, efficient and useful tool in solving crime.

Discussion

Forensic palynology is an underutilised field in the world of forensics, partly due to the outdated and time-consuming methods that are required for analysis. Not every crime scene will return pollen, and while there are other sources of evidence that could yield verification of an opinionated higher value such as blood, fingerprints or human DNA, pollen can still be a valuable foundation of information. The structure of pollen is resistant to many external elements, degradation and may remain preserved for decades[2] making it useful evidence in archeological forensics, cold cases and scenes where other evidence types are limited. By proposing PCR analysis as a laboratory standard for forensic identification of pollen, it could produce a more efficient and accurate method of investigation over traditional palynological techniques. Microscopy has been the dominant approach for forensic palynology for decades, and this method has proven to be time-consuming, labour-intensive, without routine analysis and at risk of bias due to inefficient expert analysis [2, 5, 8, 9-10]. Genetic exploration progressed to the point where PCR analysis using STRs or SNPs was applied for identification between humans [10], and now DNA barcoding using PCR analysis in plants can also be applied to forensics. By exploring the use of common markers in plant genomes, researchers have been able to identify chloroplast regions providing identification down to plant species – a useful instrument in many scientific fields [2, 9, 16-17].

2.1 DNA Barcoding Markers

The choice of DNA marker under the current scientific literature presents a selection of problems. The selected marker needs to dictate the scope of taxonomic

recovery, have excellent discriminatory power, and take into considerations the sequence needed[10]. This presents a limitation when considered for use within a molecular or genetic laboratory where the sample being sequenced is unknown. For unidentified samples, there is a greater need for a universal or core loci to allow for easy processing and taxonomic resolution down to species. There is still uncertainty amongst the scientific community regarding the best core loci for plant barcoding, and which choice will produce the best primer universality and discriminatory power[5]. There have been several studies [17- 19, 21, 23] regarding the best choice of DNA marker, each with their own advantages and limitations. A review of the literature has shown several studies in recent decades employing the use of these plant markers with attempts to highlight their limitations and advantages, as well as reviews discussing their use within botany and forensic fields.

One study [17] pooled data across labs including sequence data from 907 samples representative of 445 angiosperms, 38 gymnosperm and 67 cryptogam species (flowering seed plants, open seed plants and spore-like reproducers). For these samples, seven candidate loci were evaluated to determine which was the best for routine sequencing of land plants. Direct universality assessment using a single primer pair for each locus in angiosperms showed results of 90%-98% with PCR and sequencing success for 6/7 regions of atpF-atpH spacer, matK, rbcL, rpoB, rpoC1, and trnH-psbA . For the remaining seventh loci psbK-psbI, the results showed 77%. This study recommended rbcL and matK as standard barcode pair for land plants due to its assessment of recovery, sequence quality and species discrimination.

A secondary study [23] used plant barcodes ITS2 and trnL P6 loop to identify botanical components of honey to establish geographical provenance, providing insight into honeybee diet and foraging preferences in Australia. The study detected the nominal primary source for top 5 most abundant taxa for 85% of samples, yielding a highly successful result of metabarcoding.

It has been accepted within the community that rbcL and matK are a common pairing of choice due to the straightforward recovery of rbcL and high discriminatory power of matK in angiosperms (flowering land plants) [16], however it has been noted that matK is difficult to amplify by PCR analysis using existing primer sets [21]. rbcL is also the best characterised of the DNA markers, but not the most variable [17]. psbA-trnH demonstrates high species discrimination and good amplification across land plants when used with a single pair of primers – achieving 93% discrimination down to species for angiosperms in one study, however showed problems locating bidirectional sequences [17]. ITS2 is the nuclear internal transcribed spacer, and is the comparable DNA barcode for fungi [18], which offers good discriminatory power especially at genus level [19].

One of the most common chloroplast markers argued for use is the trnL (UAA) intron and intergenic spacer between trnF, which has been widely used in phylogeography to study genealogical lineages since the early 1990s [24]. The trnL intron is 254-767 base pairs long (513 bp in total), with a P6 loop consisting of 133 base pairs (along 10-143) [24]. The P6 loop is a short chloroplast region with high amplification success, considered a 'mini barcode' useful to scientists with conserved priming sites flanking the variable loop region [19, 21]. It is useful in its ability to amplify even highly degraded DNA making it an ideal choice for use in forensic science [24]. The trnL intron

region is relatively simple to sequence with a robust set of primers (seen below in Table 1) first displayed in Taberlet 1991 [20].

The primers being used in this study have been chosen as they are highly conserved among land plants from angiosperms to bryophytes [24]. They are c and d from Taberlet 1991, Primer c Forward (5'-CGAAATCGGTAGACGCTACG-3') and Primer d Reverse (5'-GGGGATAGAGGGACTTGAAC-3'), also known below as ucp-c and ucp-d.
For a preliminary study, trnL is well suited as a locus of choice as the primers are universal enough to amplify a wide taxonomic range of plant species and due to the relatively small region size meaning there is no need for internal primers for sequencing [24].

There are also a high number of trnL intron sequences available in databases, allowing for easier taxonomic identification down to genus or species [24]. Table 1 below represents the published primer information for trnL loci in DNA barcoding with reference to the studies that have attempted it. It offers the primer name, as well as the location on the trnL gene. The average length of each primer is around 20 base pairs and all are capable of providing taxonomic resolution down to species. Majority of the primers come from the original publication in Taberlet 1991 [20], but some studies have attempted creation of new primer sets in an attempt to refine the DNA barcoding process. The location on the *Eucalypt globulus* genome was included for relevance as this is the genus chosen for this small-scale study.

Table 1
Published primer information for chloroplast DNA for trnL loci with reference to location on Eucalptus globulus genome

Primer Name	Gene	Fwd./Rvs	Length	Sequence (5' --> 3')	Reference(s)	Eucalypt	Resolution
Ucp-c	trnL exon1	F	20	CGA AAT CGG TAG ACG CTA CG	Taberlet 1991	51460	species
Ucp-b/trnL-r	trnl exon1	R	20	TCTACCGATTTCGCCATATC	Taberlet 1991	51472	species
Ucp-e	trnl exon2	F	20	GGTTCAAGTCCCTCTATCCC	Taberlet 1991	51472	species
Ucp-d	trnl exon2	R	20	GGGGATAGAGGGACTTGAAC	Taberlet 1991	52309	species
ccSSR-16 R	trnL	R	19	CCTGGCCCAACCCTAGACA	Chung and Staub, 2006	108239	species
IRB20F	trnL	F	26	GGGCTATTAGCTCAGTGGTAGAGCGC	Dhingra and Folta, 2005	107466	species
IRB19R	trnL	R	26	ATCGAAAGTTGGATCTCCATTCCATC	Dhingra and Folta, 2005	107708	species
ccSSR-16 F	trnL	F	22	TACGAGATCACCCCTTTCATTC	Chung and Staub, 2006	107884	species
IRB21F	trnL	F	26	GAGGTCTCTGGTTCAAGTCCAGGAT	Dhingra and Folta, 2005	108459	species
G	trnL		16	GGGCAATCCTGAGCCA	Taberlet 2006	-	species
H	trnL		22	CCATTGAGTCTCTGCACCTATC	Taberlet, 2006	-	species
cpDNAtrnLG_F	trnl G		20	GGGCAATCCTGAGCCAAATC	Santos and Pereira 2016	-	species
cpDNAtrnLH_R	trnL H		21	CATCGAGTCTCTGCACCTATC	Santos and Pereira 2016		species

2.2 Qiagen Extraction Kit

Ideally in a forensic laboratory, the scientific process will need to be able to be routine, simple and without bias to maintain quality control and chain of command. By narrowing down choice of DNA marker and simplifying primer pair, the process is simplified to allow the forensic scientists to isolate the DNA, run the sample, and obtain identification down to species. It is important to select a kit that will provide an effective extraction method to ensure the best yield of DNA possible. A study by Sovova 2018 [25], analysed Qiagen DNeasy® Plant Mini Kit and NucleoSpin® Food Kit with multiple methods

using fruit jam to determine the optical extraction method. It was found that the Qiagen DNeasy ® Plant Mini Kit proved to be the most suitable method, yielding well-amplifiable DNA for all samples analysed.

In Pipan 2018 [14], six different kits including the Qiagen DNeasy Plant Pro Kit were performance tested on samples of apple tree leaves to determine purification in DNA extraction, with the highest sample purity obtained with Qiagen (although with the lowest DNA concentration). Despite the low DNA concentration, further PCR amplification performed well for analysis with low levels of contamination with the Qiagen kit. Although the analysis is limited by the factor that selection of primer is dependent of DNA marker of choice, which is why it is important to either identify a universal marker, or for a forensic laboratory to utilise a standard combination of loci however more research is needed in this area.

The literature is limited in areas of studies referencing the use of Qiagen kits with raw samples of pollen. Majority of the studies within the literature that utilised the kit on pollen were samples previously extracted from honey.

When choosing a marker it is also important to keep in mind the geolocation the sample originated from. Australia is abundant with over 23,000 native plant species[22] of unique flora and fauna. Of those, botanists have identified around 800 species encompassing three genera *Eucalyptus, Corymbia* and *Angophora*, covering 101 million hectares of Australian land equaling 77% of the total native forest area [26]. Due to the richness of Eucalypt within the Australian environment, it was chosen for this experiment as an equal representative of the native bush land. It was important to select a native

bush flora as opposed to a commercial suburban plant due to the nature of forensic investigations. As majority of the Australian population lives on the outer coast of the country, it makes the Australian bush the ideal location and proximity to bury a body or dig a gravesite. *Eucalptus* grows both wildly amongst the southwest and is planted in both suburbia and commercial areas for aesthetic, preservation and conservation. Flowers and blooms are popular amongst florists and prized by apiarists for its flavor profile in honey. Not all of the *Eucalptus* species produce flowers, but those that do produce bright colourful caps in a variety of colours. The trees range from smaller shrubs up to over 200 meters in height[26], allowing a large surface area of pollen dispersal by insects or wind. These reasons have made *Eucalptus* an ideal choice for DNA barcoding and forensic analysis in this study.

Below is a figure 1 of the country Australia showing distribution of the natural Eucalypt native forest in 2013. The Eucalptus is the most common forest type recorded in Australia, covering 77% of the total native forest area with up to 800 native species[26].

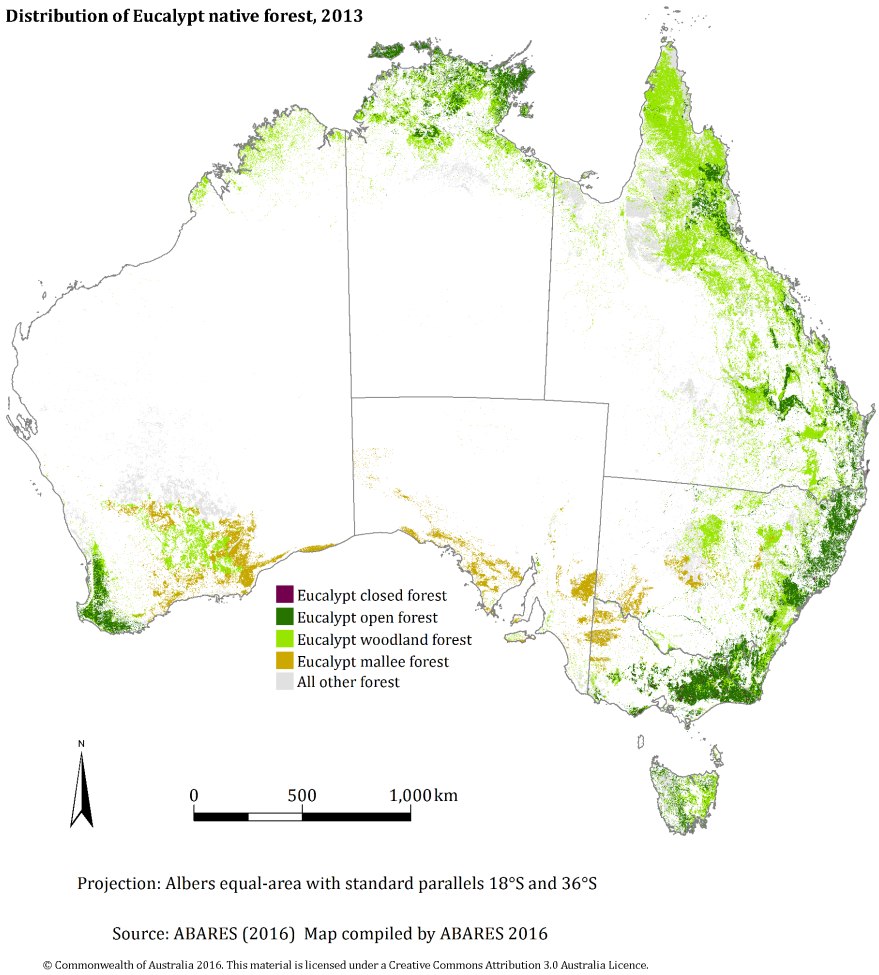

Figure 1: Map representation of Australian in 2013 showing distribution of Eucalptus native forest [26].

2.3 Pollen in forensics: applications

Pollen is useful as evidence due to its hardy exterior that is resistant to change, natural ability to spread and abundance in nature. It is found in commonplaces, in both indoor and outdoor areas, flowering houseplants are bought and planted willingly, and due to the microscopic nature it is easily trapped in clothes, hair and filters [2, 9-11, 16, 27].

One study[27] investigates the importance of spatial distribution by analysing pollen distribution from a bouquet of flowers in a room over time. It was found that the pollen persisted in increasing quantities during the time the flowers remained in the room, and

25-32% of the original pollen was present in the room even twenty days after the flowers removal. This provides an opportunity for unique pollen assemblages to accumulate in a domestic dwelling and transfer onto a suspect or clothing through direct contact, which could provide useful information in forensic investigations. The pollen was recovered in high amounts within 0.8 meters of the source showing strong decay trends despite no airflow or winds. Where there were areas of movement, the pollen count was slightly lower indicating that the surface had been disturbed which also signifies that the pollen transferred contact between the person within the room and the surface. A bouquet within a room in a domestic dwelling will have a distinct assemblage that will reflect the history of flowers and plants within that room. This offers the potential for a person entering the room to be within direct contact of grain transfer onto clothing or objects, and serves as evidence of contact with that room. This knowledge is valuable to forensic investigations as there is little information on pollen movement and transference in domestic dwellings.

The study of fossilised pollen, or palynomorphs, has various applications in archaeology, paleoecology, anthropology and forensic palynology as well. Newer DNA barcoding methods such as NGS allow for handling of delicate samples that may be partially destroyed or too delicate to process by other means. Traditional microscopy based methods do not have the same resolution or taxonomic power that barcoding can offer to improve the efficiency of pollen grain identification [10].

When discovering an older clandestine grave, it can be useful knowledge to know what season the body was buried there. Forensic palynology was applied to mass graves

uncovered from the Bosnian War, both at primary sites where the bodies where initially buried and secondary sites where they had been moved to and buried at a later date. The soil composition was analysed and pollen species discovered which helped investigators match the bodies to the primary burial site [10].

Pollen analysis on the body can help to provide correspondence between geolocation and time of year. A pollen calendar is a useful consulting tool in estimating time or season of death in forensic investigations. A palynological study[28] in Northern Italy covered analysis on 28 corpses over the course of a year, with pollen being sampled from eyebrows, hairline, facial skin and nasal cavities. The recovered grains were well preserved and analysed against pollen calendar records showing pollen types and concentrations by month in the areas the bodies were located. The variables considered where date of death, the region and locality of death and results showed good correspondence between the pollen types recorded by aerobiological monitoring stations at the time of death, and those retrieved from the corpses.

Pollen can also be used to place people at locations, and not just for determining clandestine gravesites. A murder case[29] in the English Midlands showed the cast of a footprint as forensic evidence to a murder of a young woman at a nearby field and stream. A sample was taken from the soil where the cast was taken for palynological analysis as well as samples from nineteen surrounding fields for exclusionary purposes. Forensic analysis from the cast of a footprint taken in soil provided information such as shoe size, gait, tread and brand, but analysis of the interface between cast and soil can also provide a source of information placing the offender in that location. Palynology

samples were stained with safranin, mounted and analysed using SEM. Pollen analysis of the soil depicted the murderer had been standing in a nearby stream and together with fiber analysis, physical/chemical characteristics of the soil, forensic reconstruction of the crime taken place provided valuable insight and they were apprehended. In this case, the pollen grains found were in a good state of preservation with many grains degraded and crumpled except for six remaining samples. In this instance, DNA barcoding would provide the gap needed to analyse broken and deteriorated pollen grains, and perhaps provided information that was missing due to degraded grains.

2.4 Gaps in the literature

Potentially the biggest gap within the literature is the lack of consolidated data across all areas of DNA barcoding. There are multiple forensic cases [4, 7, 8, 11] available in the literature where pollen and palynology have been used but few real life applications where DNA barcoding has been employed. While DNA barcoding can identify multiple taxonomic groups, and is considered the fastest way to differentiate between species of pollen[2], it is still limited by the uncertainty of a chosen DNA marker and lack of universal database. There are published studies that have made attempts to use DNA barcoding with loci of choice for pollen species identification, but there remains no consolidated data with all DNA markers and a universal plant source for comparison.

Table 2 below represents a consolidation of published studies using DNA barcoding methods for pollen analysis. The literature covers mostly honey composition and plant-pollinator reactions, however more recent studies are turning to DNA barcoding as a source for identification of wild and native flora. It is worth mentioning

that there are few, if any studies, that are currently undertaking DNA barcoding in a forensic capacity but there are many scientific reviews [2, 9-10, 16, 22] detailing its important and urging a change within the field.

Table 2

Summary of published studies using DNA barcoding methods for pollen species identification

Application	Loci	Sequencing Platform(s)	Author(s)	Reference
Identification of wild plants	Plastid rbcL	Cloning and NGS	Bafeel et al, 2012	4
Honey composition	Nuclear ITS2 and Plastid trnL P6 loop	Cloning and NGS	Milla et al, 2021	23
Identification of local flora	Plastid rbcL and matK	Cloning and Sanger sequencing	Burgess et al, 2011	30
Identification of native flora	Plastid rbcL and matK	Database and herbarium	de Vere et al, 2012	31
Honey composition	Plastid rbcL and trnH-psbA	Cloning and Sanger sequencing	Bruni et al, 2015	32
Honey composition	Plastid trnL-trnF, Nuclear ITS2, and plastid psbA-trnH and matK	Cloning and NGS	Kamo et al, 2017	33
Airborne Allergen monitoring	Plastid trnL	Ion torrent	Kraaijeveld et al, 2014	34
Plant-pollinator interactions	Nuclear ITS2	Roche GS junior	Keller et al, 2014	35
Honey composition	Plastid trnL	Roche GS20	Valentini et al, 2010	36
Plant-pollinator interactions	Nuclear ITS2	Illumina MiSeq	Richardson et al, 2015	37
Plant-pollinator interactions	Plastid rbcL and trnH-psbA	Cloning and Sanger sequencing	Galimberti et al, 2014	38
Honey composition	Plastid rbcL	Roche/454 GS FLX	Hawkins et al, 2015	39
Plant-pollinator interactions	Nuclear ITS2	Illumina MiSeq	Sickel et al, 2015	40

There are no worldwide herbarium collections combining reference data for all plant species, and to do so would be a vast and time-consuming task. Not all the land-plants on Earth are documented, particularly those in tropical biomes[16], and there are also rare and endangered species that now only exists in herbarium collections. The second part of this issue – would a database be created – is how that data set is portrayed. Ideally the data would exist together as referenced photographs of the plant, geographical locations found, morphological pollen grain attributes and identifying features, and ptDNA coding sequences (with access to BLAST for searches). Attempts have been discussed and potentially started within the botany and scientific community, but thus far no database as such exists. A database was attempted within traditional palynological techniques called POLLEN73S[6]. Its design was to create a baseline for pollen grain classification with an annotated image dataset for 73 pollen types in 2523 images from the Brazilian Savanna. The concept achieved good results, with 95.7% and 94% precision identification rate in an automated classification system, but with limitations to the chosen area.

Furthermore, choice of DNA marker presents difficulties within DNA barcoding that requires more research in the area. There are many studies that have successfully used DNA barcoding with pollen (consolidated in Table 1), productively using a range of different loci across the plant genome. However, not all of those markers achieve a hundred percent effective identification rate, and it is shown that choice of marker has an affect on successful taxonomic resolution of the sample analysed. Within forensic palynology, when every sample isolated and analysed as trace-evidence will more than

likely be an unknown, it is important for the DNA marker of choice to have a high degree of universality across a broad range of taxonomic groups [10].

A secondary issue is whether or not DNA barcoding can be used in a standard laboratory under simulated forensic case-settings, and if it is applicable to be introduced to forensic laboratories as they do with standard human DNA analysis. As with forensic evidence non-destructive analysis is essential. It is important within forensic investigations that evidence is preserved, however DNA barcoding destroys the sample due to the structure of the exine and the process of DNA extraction. When evidence is sampled, the pollen species is unknown and this leaves the forensic palynologist to identify multiple of the same species through microscopy. The palynologist can then use DNA barcoding on a single sample for identification purposes with multiples of the same species still remaining. There has been research into non-destructive DNA analysis on single pollen grains with success, allowing the grain to remain intact for secondary analysis. In this study[1], pollen grains were isolated and gently tweezed from their exine into capillary tubes. Ethanol was added and heated, allowed the DNA from the pollen grain to transfer into the capillary tube for PCR analysis. Digital microscopy showed the grains remained intact and identifiable. This method allows for analysis of genetic material without destroying the integrity of the pollen grain, something that could prove useful in the future of forensic palynology.

The trnL DNA marker has been chosen for its useful ability to amplify even highly degraded DNA – making it an ideal choice for forensic use as often the DNA can be degraded or older. It is also not a highly researched marker so potentially it can yield new information to the field. This literature review has attempted to consolidate the data

available regarding DNA barcoding and pollen analysis studies (see table 2 above). Other studies are available regarding plant species, or traditional palynology methods but those have been excluded from the table to allow for more related data.

To fill the gaps in the literature, this project started as an interesting point on why pollen isn't classified by its molecular characteristics as opposed to its morphological ones. From there, research into the literature led into topics such as traditional palynology techniques, DNA barcoding, metabarcoding and what potential they have for forensic investigations. There are gaps in the literature that a small introductory project like this one cannot fill due to time-constraints, budget, knowledge and lack of resources and contacts. The largest gap is whether, given time, selected loci, databases and data, can DNA barcoding become a laboratory standard for forensic palynology in investigations? This is the main question that this research aims to answer.

The experimental design is intended to fill the gap where a standard laboratory, using a standard kit (in this case the Qiagen® DNeasy Plant Mini Kit) could attempt to extract a ptDNA sample using a chosen marker (trnL), primers, and produce a clean yield of DNA. As standard forensic laboratories carry the molecular and genetic equipment necessary for human DNA extraction, with enough knowledge, DNA barcoding of plant DNA could become routine analysis in those laboratories as well.

Experimental Design

For the purposes of this study, the pollen samples chosen was the genus *Eucalptus*, belonging to the family *Myrtaceae.* These were chosen for their growing abundance in the Australian regions and their large number of genus and species. They are also internationally recognised, and a popular flowering plant overseas. *Eucalptus* have long flowering times in Autumn and Summer, and their capped gums of flora and pollen can often be found covering the ground close to and far away from their original location, helped by pollinators, birds and the wind. They bloom in a variety of colours, boasting over 800 species endemic to Australia alone [26].

Three species where chosen and selected from private property in the Armadale area, Western Australia. The first species is *Eucalyptus camaldulensis*, also known as the "red river gum". It is the most widely distributed of all the *Eucalyptus* species, common amongst rivers and waterways. The second species chosen is *Eucalyptus erythocorys,* or the "red capped gum". This particular species is endemic to Western Australia, found in the Wheatbelt, Mid-west and Gascoyne regions but is also sold commercially due to its popular golden flowers. The third species is *Eucalyptus s ficifolia* or *Corymbia ficifolia,* commonly known as the "red flowering gum". It is widespread in the south-east and south west regions in the Stirling Range. These three *Eucalyptus* species represent a larger variety of a family and cover a large area of native bushland (see Fig 1) making them an ideal choice for forensic palynology studies.

To reflect the real-life situations of forensic technicians, the method of choice for recovery of the pollen sample will be tape-lift or swab technique. This was chosen to

reflect that pollen and DNA recovered from crime scenes is often degraded, torn, or limited in number. Even though samples were collected from *Eucalyptus camaldulensis, erthocorys and ficifolia* as whole flowers with several capped gums to ensure there was plenty of samples to work with, in forensics it is very unlikely there will be such a large quantity of raw sample available. The only exception is collection of comparison to a piece of evidence or offender. The tape lift and swab method could be employed for pollen often it is microscopic, and difficult to see on clothing or other forms of evidence. Unless it was deliberately sought out, more often it would be found as part of trace-evidence and this study employs a method to reflect that.

DNA will be isolated using the Qiagen® DNeasy Plant Mini Kit with bead-beating technology for DNA extraction. Instructions will process as the manufactures guides. This kit was chosen as it has been used successfully to isolate genomic plant DNA in several other studies [13-14, 25] and it offers a simplicity that is needed for a starting project. The only negative review found for the Qiagin kit is that the DNA concentration yield was slightly lower than normal[14], but all sample purity was high.

The DNA marker chosen for this study is the trnL intron due to its simplicity in sequencing, routine recovery and primer publication[19]. The primers being used in this study have been chosen from Taberlet 1991, Primer c Forward (5'-CGAAATCGGTAGACGCTACG-3') and Primer d Reverse (5'-GGGGATAGAGGGACTTGAAC-3')[20]. For a full set of primers, see Table 1.

Once the PCR analysis has been performed, samples will be run on a gel electrophoresis. It is expected to see well-identified, clear DNA markers.

If time and access allows, it would be ideal to return nucleotide sequences for samples and run those against databases such as NCBI blast for comparison and taxonomic resolution.

Conclusion

DNA barcoding has the potential to allow forensic palynology to become a more valuable tool in criminal investigations by providing a faster, more efficient method of pollen grain analysis. Not only is it a valuable asset to modern day forensic cases, but it can also be used to provide insight into forensic archaeology cases such as clandestine grave locating, study of lifestyle from ancient remains[8] and mass war grave identifications[10]. Barcoding also possess the ability to extract quantifiable DNA from degraded, torn or damaged pollen grains that are otherwise unidentifiable by morphological means making it a more efficient method than traditional microscopy techniques.

Published studies are slowly filling the gaps in the literature, allowing for a more refined process and highlighting the greater need for a more consolidated data across a universal taxonomic range. There is still disagreement on a universal barcode, although with each published study there is more information added to the greater scientific works. Arguably, the greatest gap in the literature remains the lack of universal database with confirmed and verified sequences, as well as consolidation of the data but each additional paper adds to the literature and advances the field.

This study intends to close that gap a little further by experimenting on a small-scale study, simulating a forensic palynological case and using the power of DNA barcoding for palynological analysis. Although there are still limitations that need to be addressed, the method is closer than before to being feasible for routine analysis in forensics.

Blank Page – not numbered

Part Two

Manuscript

A method for the molecular classification of pollen for forensic purposes

ABSTRACT

The successful taxonomic identification of palynological evidence could be valuable to forensic investigations, and this can be done quickly and efficiently through DNA barcoding. There are various methodologies implemented to ascertain the best way to obtain the highest yield of good quality DNA from the pollen grain without contamination, however there is no universal agreement on an acceptable method across all platforms. This study examines the use of raw plant material from the species *Eucalyptus* and the mechanical method of pollen exine perforation for DNA extraction. The DNA yield was measured for further analysis and allowed for suggestions made regarding further studies in the field of DNA barcoding to assist within forensic investigations. This study demonstrated the need for further research into DNA barcoding as well as a consolidation of the literature, both of which will benefit the field of forensic palynology.

Keywords: Forensic, palynology, DNA barcoding, trnl intron, pollen.

PART TWO – EXPLORATORY STUDY

Hypothesis: Raw plant samples provide a large yield of plant DNA. Through sufficient means of exine perforation, DNA extraction will provide a suitable quantity of ptDNA for PCR analysis.

Research Aims

The following exploratory study aims to examine the following research questions;

- What is the amount of DNA that can be extracted from plants, and does this differ between classes and species?
- Can raw samples of plant material provide sufficient yield of DNA for use in PCR analysis?
- Can mechanical means of exine perforation be sufficient enough to yield a suitable quantity of DNA for PCR analysis?

Introduction

DNA barcoding can help the field of forensic palynology by making it readily accessible to a wide range of forensic laboratories, as well as increase taxonomic resolution of identification of plant species[1]. This is done through careful collection of evidence, PCR analysis and comparison of samples. Degree of success of the examination depends on the type of study conducted, pollen-collecting method, and the quantity of pollen used in analysis[9]. The method chosen for analysis can be based on user preference dependent on type of plant sample available, region selected for amplification, primers chosen and technology available. Depending on the steps chosen, they can affect the end result in the downstream process, so it is important in DNA barcoding and metabarcoding to select the right method for optimal results[41].

To date, there is no universally recognised method available for the independent extraction of plant DNA[42-43], but there are several commercial kits available for varying needs. The extraction procedure requires DNA of a reasonable quantity, quality and purity in order for successful PCR analysis. There are numerous inhibitors that can affect efficiency of DNA extraction and amplification, including polysaccharides, lipids and polyphenols, which need to be taken into consideration during the process[42]. Plant DNA can be extracted from many sources, including from plant material, tissues, honey, processed seeds, oil and food[44]. Within criminal investigations, it is important to understand the sources of plant DNA available, and the processes necessary for extraction for maximum DNA yield and choose an appropriate method on a case-by-case basis.

When obtaining pollen DNA, the initial step begins with extraction of the non-germinated pollen of the pollen grain by breaking through the exine using chemical or mechanical means[41]. There are factors that affect the effectiveness of rupture of the pollen exine including chosen method, as well as size and morphology of the grain. As strength and elasticity of the exine structure varies between species, this can affect how efficiently the exine ruptures, which in turn affects DNA yield[9,41]. Pollen grains share similar morphological features between genera and families[43] so various methods can be considered when there is multiple pollen species analysed, or where there is a pollen assemblage. The exine is composed of a biopolymer called sporopollenin that is highly resistant to physical, chemical and enzymatic degradation, being designed to protect the plant DNA inside for germination[41]. The composition of the sporopollenin is made of a smaller distribution of sub-units with the layout of these units allowing a geometric pattern for force distribution, this being just one factor that contributes to the variable strength and elasticity found in the exine between species. On the wall of the grain are structures known as pollen apertures that play a role in pollen tube development and vary in number, shape, size and margination, and these can also factor into exine rupture[41]. The large range in morphology of the exine composition between species is also a result of environmental pressures, differences in dispersal, and methods of reproduction[45].

Simel et al. [46] studied several chemical and mechanical techniques to determine the more efficient method of rupture the exine while yielding non-degraded DNA in the process. The study showed bead milling to be the most effective method. A second study[23] evaluates several chemical and mechanical methods in comparison of soil samples

to determine the most efficient method of extraction based on yield and molecular size of DNA recovered. Of the methods, results showed a combination of bead milling homogenisation and chloroform or phenol optimised the DNA extracted. To increase DNA yield in addition to mechanical disruption of the pollen exine, a lysis buffer can be utilised in the DNA extraction technique. Studies have also shown duration of incubation in lysis could play a role in final DNA yield, especially with pollen that has a thicker exine, or multiple apertures as opposed to thin-walled species[41].

As DNA barcoding is a relatively new concept within the field of palynology, there is currently no recommended standard within the literature as to the average amount of DNA required for PCR and sequencing analysis, and results have varied by study and by kit. One study[47] using the Qiagen DNeasy Plant Kit used samples of fresh leaf tissue, noting a significantly higher than usual yield (10.40ug gDNA/g) of extracted genomic plant DNA than extraction protocols while following kit instructions. The plant DNA for this study was extracted using mortar and pestle with assistance through liquid nitrogen. A second study[14] showed a comparison between six genomic DNA extraction methods using the Qiagen DNeasy Plant Pro Kit and Omega Bio-tek SP Plant DNA Kit with apple leaves homogenized with Tissue Lyser for exine perforation. DNA concentrations of the extracted samples were measured using Qubit 3.0 Fluorometer (Thermo Scientific) and ranged from 0.9ng/uL to over 1000 ng/uL. The cause of this was stated to potentially be a high amount of DNA template, potential PCR inhibitors, or contamination within the mixture. A further study[48] uses grapevine material for DNA extraction, ground in a mortar and pestle with assistance of liquid nitrogen, however the extraction kits are unknown (labelled only as commercial). Quantification showed a range of DNA yields from

29.5ng/ul to higher yields of 1586 ng/ul. The study suggests the differences between the yields may be attributed to sample amount used and extra precipitation steps. Reasons for differing concentrations within the process of DNA extraction could be due to multiple wash stages, the manual transference of samples between tubes, and the binding of particles to the substrate instead of elution. There has been shown to be a positive correlation between level of exine rupture and amount of DNA yielded, which could explain the variation between the quantification results[41].

In order for DNA extraction to be efficient, there needs to be effective disruption of the pollen exine, gametophytic cell walls and nuclear membranes without DNA degradation[41]. Once the DNA extraction has been successful, and the quantified DNA has been determined to be adequate, primers can be added and the solution amplified for analysis and comparison. The studies[47-48] continued with PCR analysis despite the resulted DNA quantifications with successful amplification, showing gel electrophoresis of clear high-molecular weight bands with little or no smearing[47], and acknowledged the samples were well suited to NGS sequencing.

The purpose of this study is to identify the gaps within the literature of the effectiveness of DNA barcoding and the uses it has within the field of forensics. The aims of this study was to utilise the use of raw plant material within the extraction process, as well as only mechanical means for exine perforation. To achieve this, DNA was extracted using a Qiagen DNeasy Plant Mini Kit from raw samples of three chosen species of Eucalyptus and quantified using the Qubit 3.0 Fluorometer to ascertain if PCR analysis was achievable.

Methodology

Collection of Eucalyptus pods

Three species of native *Eucalyptus camaldulensis, erythocorys* and *ficifolia* were collected as whole flowers with several capped gums from private property within the Armadale area of Western Australia in the season of Autumn, 2022. Each sample was collected directly from the tree, individually of each other on the same day, and left to dry completely before being placed inside a sterile labelled jar.

To confirm the species of *Eucalyptus*, the literature defining morphological characteristics was consulted, as well as the EUCLID database, located at (https://apps.lucidcentral.org/euclid/text/intro/index.html). The lucid key was used to eliminate and confirm the three species of *Eucalyptus camaldulensis, erythocorys* and *ficifolia.* The flora was collected from private property within Western Australia for the purposes of this study.

Sample Preparation (standards and blank preparation)

Appropriate laboratory protocol was followed, and powder-free Micro-touch (Nitrile) gloves were used and changed between each sample. The samples were removed from the jar and weighed individually on laboratory scales (A&D Company Ltd). The anthers and stamens were removed from the receptacle using sterilised tweezers. *Euc. Camaldulensis* yielded 0.405 grams total weight from two flower pods collected. The second sample, *Euc. ficifolia* produced 0.058 grams total weight from two pods. The third sample, *Euc. erythocorys* yielded a total weight of 0.904 grams from a singular flower pod, and as such, only one pod was used.

The plant anthers were finely ground using a stone mortar and pestle. Care was taken to avoid grinding the sample into a wet paste, and instead to attempt to perforate the exine. In between each use, the mortar and pestle was cleaned with 90% ethanol.

DNA Extraction and Quantification

Of the three Eucalyptus species, 100mg of raw sample for each was ground and used for the extraction. In total, ten samples were extracted; three replicates of each species and a control. DNA extraction was performed using Qiagen DNeasy Plant Mini Kit (Qiagen, Australia) with final elution volume of 100ul. The method was followed as per manufactures instructions with the exception of nitrogen freezing. The concentration of each sample was determined using a Qubit 3.0 fluorometer with the 1x dsDNA HS assay (Invitrogen). The method was followed as per Qiagen kit instructions, using 10uL of 10ng/uL, 10uL dsDNA standard combined with 10uL of DNA of each sample, and 170 working standard to final volume of 200uL each.

DATA ANALYSIS

Results are displayed in Table 3 below, showing the Qubit 3.0 Fluorometer measurements of DNA yield from Eucalyptus species extracted using the Qiagen DNeasy Pro Mini Kit. Three samples from each species of Eucalyptus (*camaldulensis, ficifolia and erythocorys*) were analysed, as well as a control, each with a starting mass of 100mg. There is a comparable trend within the same species of Eucalyptus resulting in similar DNA yields. The average for each species was calculated and tabulated at the end of the table (also in ng/ul). The individual results of each sample were tabulated to show any corresponding trends between species.

Sample no.	Eucalyptus Species	Starting Mass (mg)	Extraction Method	Post-Extraction Observations	DNA Con. (ng/uL)	Average Species (ng/uL)
1A	Euc. camaldulensis	100mg	Qiagen DNeasy Plant Minikit	clear	128	
1B	Euc. camaldulensis	100mg	Qiagen DNeasy Plant Minikit	clear	124	126
1C	Euc. camaldulensis	100mg	Qiagen DNeasy Plant Minikit	clear	126	
2A	Euc. ficifolia	100mg	Qiagen DNeasy Plant Minikit	clear	138	
2B	Euc. ficifolia	100mg	Qiagen DNeasy Plant Minikit	clear	134	135.3
2C	Euc. ficifolia	100mg	Qiagen DNeasy Plant Minikit	clear	134	
3A	Euc. erythocorys	100mg	Qiagen DNeasy Plant Minikit	clear	140	
3B	Euc. erythocorys	100mg	Qiagen DNeasy Plant Minikit	clear	138	140
3C	Euc. erythocorys	100mg	Qiagen DNeasy Plant Minikit	clear	142	
Control	-	-	Qiagen DNeasy Plant Minikit	clear	Sample out of range – too low	

Table 3: *Qubit 3.0 Fluorometer measurements of DNA yield from Eucalyptus species extracted using Qiagen DNeasy Pro Mini Kit.*

Obtaining a high DNA yield from plant material relies on factors such as successful exine perforation through appropriately chosen chemical and mechanical methods without DNA degradation. The samples were extracted using the Qiagen DNeasy Plant Mini Kit and the results quantified using the Qubit 3.0 Fluorometer, and showed an average concentration (ng/ul) of DNA obtained across all samples. According to the manufactures guide for the Qiagen DNeasy Plant Handbook[49] in determination of yield and purity at least 50ng of DNA is required when analysing palynological samples for successful PCR analysis.

This experiment used raw quantities of each sample of the species *Eucalyptus camaldulensis, erythocorys* and *ficifolia.* The pollen for the *Eucalyptus* family is located within the stamen, consisting of the anthers and filaments. The anthers were removed and ground using the mortar and pestle to puncture the exine, however no chemical assistance such as TissueLyser was used. An addition of a chemical method is often recommended to assist with perforation of the exine and facilitate DNA extraction but was not considered necessary given the large quantity of sample available. The choice was made to use a raw sample of *Eucalyptus* species, as evidence found within forensic investigations is unpredictable, and could range from raw samples, to single grains, as well as degraded or torn fragments.

The Qiagen DNeasy Plant Mini Kit uses silica-based spin columns to isolate the DNA and on average, yield 3-260ug of high quality DNA depending on binding of the membrane, and type of sample used[49]. Results for the *Eucalyptus* species varied between 124ng/ul to 142 ng/ul at the highest concentration. The variations between each family

species are slight, with *Euc. camaldulensis* between 124-128 ng/ul, *Euc. ficifolia* between 134-138 ng/ul, and *Euc. erythocorys* between 138-142 ng/ul. The similarity in results across species could be due to comparisons between morphology in exine and aperture, and how the sporopollenin network is crafted in the family[41]. In future research, observation of the morphological exterior of the pollen grain could provide an advantage into choosing the best method of exine rupture to maximise DNA yield and purity.

Given that the Qiagen Kit used recommends samples of a 50ng/ul minimum for PCR analysis, and all the samples of *Eucalyptus* recorded were above that, this establishes that there is an appropriate yield of DNA to continue with PCR analysis. There was no expected quantification for the results, and the literature demonstrates the variance between figures and species. As similar studies[47-48] showed a variation between samples and species and continued with successful PCR amplification, it can be assumed that provided the DNA is of good quantity, yield and purity, the end result will be optimal.

Without amplification of the samples, it is unknown whether the samples are pure, or whether there is contamination across the species. When working with raw samples collected from the environment, there is a risk of unintended capture of non-pollen plant material, or non-plant taxa such as fungal spores or insect fragments that could unintentionally be extracted in the process[41]. As these could inhibit the experimental procedure, without further processing there is no way to know if the samples are pure.

It is important to note that this experiment was performed under laboratory conditions and does not reflect the true conditions of a forensic scene. The samples were gathered in raw form in abundance, and there was a large amount of pollen sample per species for analysis. The type of species was also known beforehand, unlike in an investigation if the evidence was unknown and needed to be compared to a sample for analysis. The samples where gathered with attempts to be clean and free of contaminants (to the best of ability), collected directly from the geographical source and stored until used. These laboratory conditions can still provide valuable insight into palynological samples and their use within forensic investigations. In order for DNA barcoding to be successful in forensic palynology, studies need to continue to fill the gaps within the literature.

Recommendations for further research for this study include amplification of the samples through PCR analysis and gel electrophoresis. The chosen DNA marker for this study was the trnL intron, and is still recommended due to the routine recovery, simplicity in sequencing and primer publication. It also has the ability to amplify even highly degraded DNA, which can be a source of evidence in forensic investigations. Suggested primers are from Taberlet 1991[20]: Primer c Forward (5'-CGAAATCGGTAGACGCTACG-3') and Primer d Reverse (5' GGGGATAGAGGGACTTGAAC-3'

For sample comparison, database searches can be compared against online databases such as NCBI blast, Genbank or BOLD however there is a lack of consolidative herbarium data across the literature due to funding and sources[1,9-10].

The results demonstrate that it is possible to extract DNA of a sufficient yield and quantity in order for further analysis, using a raw sample collected from the environment. It also demonstrates that use of a mortar and pestle without further chemical assistance on the anthers and stamen of a flowering bud is sufficient enough force to perforate the pollen exine of species *Eucalyptus camaldulensis, erythocorys* and *ficifolia.*

Limitations of this study include the gaps in the literature surrounding DNA barcoding in terms of selection of DNA marker(s), primer selection and universality, and lack of established database. This study highlights the need for further research in the field of forensic palynology, and the need for a universally recognised method available for plant DNA extraction, sequencing and amplification. Further published studies would emphasise the need for more consolidated data across the literature and allow the field to expand. Directions for future studies in reference to forensic palynology include research into efficiency of maximising yield of DNA without destruction of the pollen grain or evidence, improving upon already learned selection of loci and primer knowledge, building of herbal literature and databases and further studies accumulating knowledge upon seasonal pollen calendars and assemblage.

Conclusion

This study demonstrates the suitability of raw plant sample processed with mechanical means for exine perforation without chemical assistance, providing a suitable DNA yield, which illustrates the capacity for further PCR analysis. This study was performed with the intent to be used in future reference for potential forensic palynological purposes. This research also highlights the need for further studies to be

performed to continue to identify gaps within the literature, as well as provide further data for analysis. Additionally, further PCR analysis would allow for application of the trnL intron, an insufficiently researched loci in DNA barcoding with pollen grains.

The opportunity to continue the study would provide research into the palynological field and may assist forensic investigators by allowing DNA barcoding to provide insight into further taxonomic identification as opposed to morphological characteristics identified through microscopy alone. Therefore, continued research into DNA barcoding could provide forensic experts with a valuable source of information.

REFERENCES

1. Kelley L, Rose E, McCullough B, Martinez M, Baudelet M. Non-destructive DNA analysis of single pollen grains. Forensic Chemistry. 2020;20:100275.
2. Alotaibi S, Sayed S, Alosaimi M, Alharthi R, Banjar A, Abdulqader N et al. Pollen molecular biology: Applications in the forensic palynology and future prospects: A review. Saudi Journal of Biological Sciences. 2020;27(5):1185-1190.
3. Boi M. The Importance of Palynology in Forensic Investigations. Forensic Science & Addiction Research. 2018;3(5).
4. Abdulrahaman A, Al Sahli A, Okoli J. The use of soil palynomorphs in forensics. Journal of Applied Sciences and Environmental Management. 2018;22(1):85.
5. Bafeel S, Arif I, Bakir M, Al Homaidan A, Al Farhan A, Khan H. DNA barcoding of arid wild plants using rbcL gene sequences. Genetics and Molecular Research. 2012;11(3):1934-1941.
6. Astolfi G, Gonçalves A, Menezes G, Borges F, Astolfi A, Matsubara E et al. POLLEN73S: An image dataset for pollen grains classification. Ecological Informatics. 2020;60:101165.
7. Ezegbogu M. Identifying the scene of a crime through pollen analysis. Science & Justice. 2021;61(3):205-213.
8. Arguelles P, Reinhard K, Shin D. Forensic Palynological Analysis of Intestinal Contents of a Korean Mummy. The Anatomical Record. 2015;298(6):1182-1190.
9. Bell K, de Vere N, Keller A, Richardson R, Gous A, Burgess K et al. Pollen DNA barcoding: current applications and future prospects. Genome. 2016;59(9):629-640.
10. Bell K, Burgess K, Okamoto K, Aranda R, Brosi B. Review and future prospects for DNA barcoding methods in forensic palynology. Forensic Science International: Genetics. 2016;21:110-116.
11. Bastl K, Berger U, Kmenta M, Weber M. Is there an advantage to staying indoors for pollen allergy sufferers? Composition and quantitative aspects of the indoor pollen spectrum. Building and Environment. 2017;123:78-87.
12. Barrett S, Hough J. Sexual dimorphism in flowering plants. Journal of Experimental Botany. 2012;64(1):67-82.

13. Ryan A, O'Hern C, Elkins K. Evaluation of Two New Methods for DNA Extraction of "Legal High" Plant Species. Journal of Forensic Sciences. 2020;65(5):1704-1708.
14. Pipan B, Zupančič M, Blatnik E, Dolničar P, Meglič V. Comparison of six genomic DNA extraction methods for molecular downstream applications of apple tree. Cogent Food & Agriculture. 2018;4(1).
15. BORISENKO A, SONES J, HEBERT P. The front-end logistics of DNA barcoding: challenges and prospects. Molecular Ecology Resources. 2009;9:27-34.
16. Kress W. Plant DNA barcodes: Applications today and in the future. Journal of Systematics and Evolution. 2017;55(4):291-307.
17. Hollingsworth P, Forrest L, Spouge J, Hajibabaei M, Ratnasingham S, van der Bank M et al. A DNA barcode for land plants. Proceedings of the National Academy of Sciences. 2009;106(31):12794-12797.
18. Shokralla S, Gibson J, Nikbakht H, Janzen D, Hallwachs W, Hajibabaei M. Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. Molecular Ecology Resources. 2014;14(5):892-901.
19. Heinze B. A database of PCR primers for the chloroplast genomes of higher plants. Plant Methods. 2007;3(1):4.
20. Taberlet P, Gielly L, Pautou G, Bouvet J. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology. 1991;17(5):1105-1109.
21. Hollingsworth P, Graham S, Little D. Choosing and Using a Plant DNA Barcode. PLoS ONE. 2011;6(5):e19254.
22. Dormontt E, van Dijk K, Bell K, Biffin E, Breed M, Byrne M et al. Advancing DNA Barcoding and Metabarcoding Applications for Plants Requires Systematic Analysis of Herbarium Collections—An Australian Perspective. Frontiers in Ecology and Evolution. 2018;6.
23. Milla L, Sniderman K, Lines R, Mousavi-Derazmahalleh M, Encinas-Viso F. Pollen DNA metabarcoding identifies regional provenance and high plant diversity in Australian honey. Ecology and Evolution. 2021;11(13):8683-8698.

24. Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A et al. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research. 2007;35(3):e14-e14.
25. Sovová T, Křížová B, Ovesná J. Determining the optimal method for DNA isolation from fruit jams. Czech Journal of Food Sciences. 2018;36(No. 2):126-132.
26. Department of Agriculture, Water and the Environment. Australia's State of the Forest Report 2018. Canberra: Australian Government; 2018.
27. Morgan R, Allen E, King T, Bull P. The spatial and temporal distribution of pollen in a room: Forensic implications. Science & Justice. 2014;54(1):49-56.
28. Montali E, Mercuri A, Trevisan Grandi G, Accorsi C. Towards a “crime pollen calendar”—Pollen analysis on corpses throughout one year. Forensic Science International. 2006;163(3):211-223.
29. Bull P, Parker A, Morgan R. The forensic analysis of soils and sediment taken from the cast of a footprint. Forensic Science International. 2006;162(1-3):6-12.
30. Burgess K, Fazekas A, Kesanakurti P, Graham S, Husband B, Newmaster S et al. Discriminating plant species in a local temperate flora using the rbcL+matK DNA barcode. Methods in Ecology and Evolution. 2011;2(4):333-340.
31. de Vere N, Rich T, Ford C, Trinder S, Long C, Moore C et al. DNA Barcoding the Native Flowering Plants and Conifers of Wales. PLoS ONE. 2012;7(6):e37945.
32. Bruni I, Galimberti A, Caridi L, Scaccabarozzi D, De Mattia F, Casiraghi M et al. A DNA barcoding approach to identify plant species in multiflower honey. Food Chemistry. 2015;170:308-315.
33. Kamo T, Kusumoto Y, Tokuoka Y, Okubo S, Hayakawa H, Yoshiyama M et al. A DNA barcoding method for identifying and quantifying the composition of pollen species collected by European honeybees, Apis mellifera (Hymenoptera: Apidae). Applied Entomology and Zoology. 2018;53(3):353-361.
34. Kraaijeveld K, de Weger L, Ventayol García M, Buermans H, Frank J, Hiemstra P et al. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Molecular Ecology Resources. 2014;15(1):8-16.
35. Keller A, Danner N, Grimmer G, Ankenbrand M, von der Ohe K, von der Ohe W et al. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biology. 2014;17(2):558-566.

36. Valentini A, Miquel C, Taberlet P. DNA Barcoding for Honey Biodiversity. Diversity. 2010;2(4):610-617.
37. Richardson R, Lin C, Sponsler D, Quijia J, Goodell K, Johnson R. Application of ITS2 Metabarcoding to Determine the Provenance of Pollen Collected by Honey Bees in an Agroecosystem. Applications in Plant Sciences. 2015;3(1):1400066.
38. Galimberti A, De Mattia F, Bruni I, Scaccabarozzi D, Sandionigi A, Barbuto M et al. A DNA Barcoding Approach to Characterize Pollen Collected by Honeybees. PLoS ONE. 2014;9(10):e109363.
39. Hawkins J, de Vere N, Griffith A, Ford C, Allainguillaume J, Hegarty M et al. Using DNA Metabarcoding to Identify the Floral Composition of Honey: A New Tool for Investigating Honey Bee Foraging Preferences. PLOS ONE. 2015;10(8):e0134735.
40. Sickel W, Ankenbrand M, Grimmer G, Holzschuh A, Härtel S, Lanzen J et al. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC Ecology. 2015;15(1).
41. Swenson S, Gemeinholzer B. Testing the effect of pollen exine rupture on metabarcoding with Illumina sequencing. PLOS ONE. 2021;16(2):e0245611.
42. Jasbeer K, Son R, Mohamad Ghazali F, Cheah Y. Real-time PCR evaluation of seven DNA extraction methods for the purpose of GMO analysis. International Food Research Journal. 2009;16:329-341.
43. Leontidou K, Vernesi C, De Groeve J, Cristofolini F, Vokou D, Cristofori A. Taxonomic identification of airborne pollen from complex environmental samples by DNA metabarcoding: a methodological study for optimizing protocols. 2017;.
44. Kaňuková Š, Mrkvová M, Mihálik D, Kraic J. Procedures for DNA Extraction from Opium Poppy (Papaver somniferum L.) and Poppy Seed-Containing Products. Foods. 2020;9(10):1429.
45. Wang R, Dobritsa A. Exine and Aperture Patterns on the Pollen Surface: Their Formation and Roles in Plant Reproduction. Annual Plant Reviews online. 2018;:589-628.
46. Simel E, Saidak L, Tuskan G. Method of Extracting Genomic DNA from Non-Germinated Gymnosperm and Angiosperm Pollen. BioTechniques. 1997;22(3):390-394.

47. Rana MM, Aycan M, Takamatsu T, Kaneko K, Mitsui T, Itoh K. Optimized nuclear pellet method for extracting next-generation sequencing quality genomic DNA from fresh leaf tissue. Methods and Protocols. 2019;2(2):54.
48. Akkurt M. Comparison between modified DNA extraction protocols and commercial isolation kits in grapevine (vitis vinifera L.). Genetics and Molecular Research. 2012;11(3):2343–51.
49. Qiagen. . 7th ed. DNeasy® Plant Handbook 2020 p. 54.