**A molecular assessment of the identity of regenerating mallees on the
Tropicana Mine Access Rd, in relation to the
DRF *Eucalyptus articulata* (Myrtaceae). II.**

A report from

BGPA SCIENCE

to

ANGLOGOLD ASHANTI AUSTRALIA LTD

Prepared by S.L. Krauss, J. Anthony, M. Barrett and L. Sinclair

23rd November 2009

Report # 47 from the Genetics Laboratory, BGPA Science, November 2009.

Executive Summary

The objective of this research was to apply the molecular tools of DNA sequence analysis and DNA fingerprinting to identify whether the DRF *Eucalyptus articulata* occurs among samples from along the proposed Tropicana Gold Project Mine Access Road Pinjin Option road route. Molecular testing was applied, as confident identification from morphology was difficult due to the burnt and only just regenerating nature of the eucalypt specimens of interest. Leaf samples were collected by Mattiske Consulting staff, and included 6 specimens of the DRF *E. articulata*, as well as various samples of known and questioned identity, from along a 60km stretch of the proposed access road route. DNA sequence variation was assessed for these samples, and compared to eucalypt sequences obtained from GenBank, for three regions - the Internal Transcribed Spacer (ITS) and External Transcribed Spacer (ETS) of nuclear ribosomal DNA (nrDNA), and the *matK* region from the chloroplast genome. Genetic variation was also assessed with the multi-locus DNA fingerprinting technique Amplified Fragment Length Polymorphism (AFLP). Using the Neighbor-Joining procedure with bootstrapping, ITS, ETS and matK sequence data all generated trees with significant support for a cluster of *E. articulata* individuals that ultimately excluded all samples of questioned identity. Ordination of AFLP data revealed a cluster of the known *E. articulata* samples that was strongly differentiated from all other samples. Collectively, these results conclusively indicate that *E. articulata* was not amongst the questioned samples.

Introduction

Molecular tools such as DNA sequencing and DNA fingerprinting potentially enable the unambiguous identification of specimens that are otherwise taxonomically ambiguous. This DNA barcoding approach is of particular utility when the quality and/or quantity of material is poor or fragmentary, and confident, positive identification is difficult from morphological features alone. These molecular tools facilitate the taxonomic identification of samples through the detection of genetic polymorphisms that are consistent, or fixed, among samples within taxa, but differ among taxa – a barcoding gap. DNA barcoding, and specifically the identification of suitably universal markers for DNA barcoding, has been a significant area of intense activity recently (Cowans et al 2006; Lahaye et al. 2008; Hollingsworth 2008; Valentini et al. 2008; Hollingsworth et al. 2009). While a universal marker(s), such as the mitochondrial cytochrome oxidase I gene (COI) for animals, remains somewhat elusive for plants (Hollingsworth et al. 2009), DNA sequencing and DNA fingerprinting are powerful molecular forensic tools of general utility for more narrowly defined objectives such as the comparison of unknown specimens to a known taxon.

In August 2008, AngloGold Ashanti commissioned the conservation genetics laboratory at Kings Park and Botanic Garden to undertake a molecular forensic analysis of the identity of plant specimens in relation to the DRF *Eucalyptus articulata* Brooker and Hopper, otherwise known as the Ponton Creek Mallee (Fig. 1). *Eucalyptus articulata* is a low straggly mallee to 3m, with smooth coppery or bronze bark throughout, and forming a lignotuber. It is known only from three populations of approx 120 plants near Mulga Rock, north-east of Kalgoorlie, where it occurs on red sand dunes with arkose rubble. *Eucalyptus articulata* belongs to the *Eucalyptus* subgenus *Symphyomyrtus* section *Bisectae* subsection *Glandulosae* because the cotyledons are bisected, buds have an operculum scar and the branchlets have oil glands in the pith. Within this subsection, *E. articulata* belongs to a small subgroup of four species, series *Loxophlebae*, with *E. loxophleba* (with 4 subspecies), *E. blaxellii*, and *E. semota* (Brooker et al. 2006; Hines & Byrne 2001).

The objective of this research was to apply the molecular tools of DNA sequence analysis and DNA fingerprinting to identify whether the DRF *Eucalyptus articulata* occurs amongst samples collected from along the proposed Tropicana Gold Project Mine Access Road Pinjin Option road route. Molecular tools were applied as the geographic area had been recently extensively burnt, and confident taxonomic identification of newly regenerating mallee eucalypts was often difficult from morphology alone. Specifically, for each sampled specimen, we asked whether it was taxonomically distinct from *E. articulata*. This objective is distinguished from that of positive taxonomic identification of the unknown specimens, which we did not pursue, and for which the limited sampling was not specifically designed.

Fig. 1: *Eucalyptus articulata* habit. Photo from Brooker et al (2006).


Materials and Methods

Sampling

With the exception of *E. loxophleba*, all material was sampled by Mattiske Consulting staff in August 2008 or January 2009 (*E. trivalva 1-5*), and these samples of fresh leaf material were delivered to Kings Park, where they were stored briefly at 4°C prior to the extraction of DNA. Three *Eucalyptus loxophleba* subsp. *supralaevis* PERTH herbarium specimens, and three fresh specimens from York, were also sampled and included in analyses, as *E. articulata* is a member of the *E. loxophleba* complex (Hines and Byrne 2001). Samples included 6 specimens identified as *E. articulata*, 5 specimens identified as *E. trivalva*, specimens with taxonomy identified, specimens with taxonomy tentatively determined and/or undetermined, and 2 specimens provided to us with identification unrevealed as a control (Table 1; Fig. 2). Voucher specimens of samples have been collected and deposited at PERTH.

Table 1: Collection details of samples supplied to Kings Park for DNA analysis. ? indicates a tentative identification made by Mattiske staff from limited morphological material from burnt mallees.

Label	Species	EASTING	NORTHING	NOTES
EA1	*E. articulata*	528412	6665115	roadside
EA2	*E. articulata*	528412	6665115	5 m from EA1
EA3	*E. articulata*	526957	6665458	?4.5 m
EA4	*E. articulata*	526931	6665471	DEC tag # 67123
EA5	*E. articulata*	526875	6665490	~60m W of EA4 on track
EA6	*E. articulata*	526817	6665492	DEC rare flora # 9.1.20
EC1	*E. concinna*	498212	6673091	for comparison, on Pinjin Rd
EC @ EA2	*E. concinna*	528414	6665118	
2	*? E. oleosa*	520452	6680957	red sand
2a	*? E. rosaceae*	515846	6681057	in mulga
2b	*E. pimpiniana*	514913	6681064	open heath
3a	*? E. transcontinentalis*	522330	6682365	
4a	*? E. oleosa*	548756	6703806	
4b	*?E.*	547463	6703663	burnt mallee, yellow-red sand
4c	*?E. platycorys*	544126	6703272	burnt mallee
4d	*?E. rigidula*	544001	6703244	mallee
4e	*? E. platycorys*	541746	6703662	burnt mallee
4f(1)	*?E. ceratocorys*	538399	6702131	burnt mallee
4f(2)	*?E. rigidula*	538399	6702131	burnt mallee
4g	*? E. mannensis*	536667	6700014	unburnt, yellow sand
4g(2)	*E. youngiana*	536667	6700015	
4h	*?E. gongylocarpa*	535270	3396960	burnt, yellow sand
4i	*?E.*	531517	6690669	burnt mallee, yellow sand
4i(a)	*?E. concinna*	531517	6690669	burnt mallee
4i(b)	*?E. trivalva*	531517	6690669	burnt mallee
4j	*E. trivalva*	531010	6689813	
Etriv1	*E. trivalva*	647940	6762196	
Etriv2	*E. trivalva*	647940	6762196	
Etriv3	*E. trivalva*	647940	6762196	
Etriv4	*E. trivalva*	647940	6762196	
Etriv5	*E. trivalva*	647940	6762196	
Elox1	*E. loxophleba*			PERTH05546699
Elox2	*E. loxophleba*			PERTH07122187
Elox3	*E. loxophleba*			PERTH07111800
Elox4	*E. loxophleba*	490964	6477963	K. Dixon collection (York)
Elox5	*E. loxophleba*	490964	6477963	K. Dixon collection (York)
P11	*X*	681399	6714500	ID not revealed
P12	*Y*	692800	6716000	ID not revealed

Fig. 2: Sample locations of material supplied to Kings Park for genetic analysis, from along the proposed Tropicana Gold Project Mine Access Road Pinjin Option road route. See Table 1 for sample labels. Scale: 10000 map units = 10km. Axes show eastings and northings.
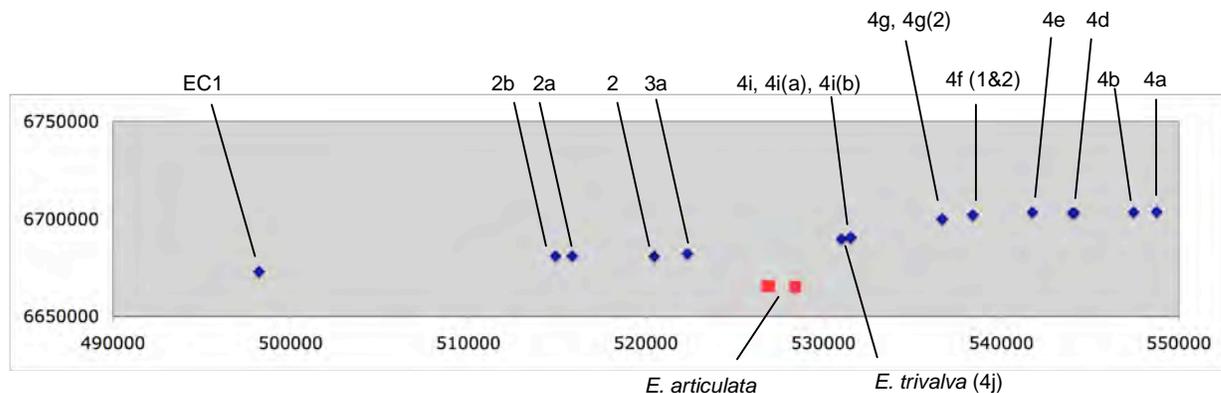
DNA extraction

DNA was extracted from fresh leaves (or leaves from herbarium specimens for *E. loxophleba*) using the protocol described in Glaubitz *et al* (2001) with a wash buffer described by Wagner *et al* (1987). The extracted DNA was visualised on agarose gel and quantified using a Nanodrop® spectrophotometer (Appendix 1).

nrDNA Sequencing (ETS and ITS regions)

Sequences were generated for two regions of nuclear ribosomal DNA (nrDNA) – the internal transcribed spacer region (ITS) and the external transcribed spacer region (ETS). ITS comprised the entire region between the nuclear ribosomal 26S and 18S genes, and has been widely used in reconstructing plant phylogeny (Baldwin et al. 1995). ETS has been demonstrated to have potentially 30% more informative sites than ITS (Baldwin and Markos, 1998). Sequence primers used are shown in table 2.

cpDNA Sequencing (matK)

DNA from the chloroplast region *matK* gene was amplified using the primer pairs 909F/2520R and 2516F/2518R (Table 2) (O'Brien et al. 2000). *MatK* is a functionally significant gene approximately 1600 bp in length, and located within the *trnK* intron. *MatK* has been observed to have a rapid substitution rate compared to other chloroplast coding regions (Lahaye et al 2008; Hollingsworth 2008). However, this gene has a reputation for being one of the more difficult chloroplast regions to routinely amplify and sequence, especially across divergent lineages (Hollingsworth 2008).

Table 2: Sequence Primers: All sequences are given 5' to 3'. F, forward primer; R, reverse primer.

PRIMER		Sequence	Reference
ITS-4	R	TCCTCCGCTTATTGATATGC	White et al. (1990)
ITS-5	F	GGAAGTAAAAGTCGTAACAAGG	White et al. (1990)
ETS-18S	R	GAGCCATTCGCAGTTTCACAG	Wilson et al. (2007)
ETS-Kunzea	F	CGTGCTGGTGCACCGAA	Wilson et al. (2007)
matK 909F	F	GGGGTTGCTAACTCAACGG	O'Brien et al. (2000)
matK 2520R	R	GATCCTTCCTGGTTGAAACCAC	Gadek et al. (1996)
matK 2516F	F	TATGCACTTGCTCATGATCA	Gadek et al. (1996)
matK 2518R	R	AACTAGTCGGATGGAGTAG	Gadek et al. (1996)

Sequences were amplified by polymerase chain reaction (PCR), and the products visualized on a 2% agarose electrophoresis gel (stained with SYBERsafe dye). Amplification products were cleaned up using the Argencourt® AMPURE PCR cleanup kits according to the manufacturers protocol. A standard cycle sequence reaction was performed using the Beckman-Coulter cycle sequence dye terminator reaction kit (see Protocol in Appendix 2). DNA sequences were visualised on a CEQ8800 Genetic Analysis System (Beckman Coulter) according to the manufacturer's specifications. Sequences were analysed using the CEQ8800 default sequence analysis parameters. Forward and reverse electropherogram profiles were combined and edited manually in CodonCode Aligner v. 2.0.4 (CodonCode Corporation). Edited sequences were then manually aligned in Se-Al v. 2.0a11 (Rambaut, 1996) along with additional eucalypt ITS and ETS sequences obtained from GENBANK (Appendix 3). Where ambiguities in the alignment occurred, the alignment chosen was the one generating the fewest potentially informative characters. As we are here concerned with phenetic similarity rather than phylogenetic relationships, overall sequence similarity among samples was assessed by a Neighbor-Joining (NJ) tree from aligned sequences using PAUP* 4.0b10 for Macintosh (Swofford 2002). Support for clusters was estimated by using the bootstrap option in PAUP*. Bootstrap values >50% were interpreted as support for clusters.

Amplified Fragment Length Polymorphism (AFLP)

The DNA fingerprinting technique Amplified Fragment Length Polymorphism (AFLP) (Vos et al. 1995, Mueller & Wolfenbarger 1999) uses stringent PCR protocols for highly reproducible multi-locus DNA fingerprints. The AFLP procedure is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA. Oligonucleotide adapters are ligated to the ends of restriction fragments, which then serve as priming sites for an initial PCR amplification. A second selective PCR amplification uses primers of complementary sequence to the ligated adapter plus additional arbitrary

nucleotides to amplify fragments for subsequent electrophoresis. Detailed protocols are listed in appendix 4. Six primer pair combinations were used (Table 2). *Pst*I and *Mse*I primer sequences are from Muluvi *et al.* (1999). Selective PCRs were amplified separately prior to visualizing with a Beckman CEQ8800 capillary machine, with internal size standard 400. Fragments were scored by eye using Beckman software for the unambiguous presence (1) or absence (0) of fragments between 60 and 400 base pairs in length. Replicate samples were run on each 96-well plate to ensure consistent scoring of reliable markers. Of the six primer pairs tested, two (m-CAG/p-AC and m-CAG/p-CG) generated strong, reliable and reproducible fingerprints. Genetic dissimilarities among all pairwise combinations of samples was estimated as the sum of alternate marker states (presence/absence) across all markers for each of two primer pairs, using Genalex 6 (Peakall & Smouse 2006). Pairwise genetic dissimilarities were visualized by ordination using principal co-ordinates analysis in Genalex 6.

Table 2. AFLP primer pairs assessed for polymorphism, reliability, and reproducibility.

m-CTC/p-AC	m-CTC/p-CA	m-CTC/p-CG
m-CAG/p-AC	m-CAG/p-CA	m-CAG/p-CG

Results

Sequencing (ITS)

We assessed ITS sequences for 58 samples, including 21 taxa obtained from GENBANK (Appendix 4). ITS failed to amplify sufficiently for all herbarium sourced *E. loxophleba* samples, presumably due to poor quality DNA from herbarium samples. One fresh collection, sourced from the York area, successfully amplified. The final matrix contained 640 characters, with 80 (12.5%) variable characters. The NJ tree generated a strongly supported cluster (bootstrap support 65%) containing all *E. articulata* samples, all known *E. trivalva* samples (including 4j), *E. loxophleba*, *E. wandoo*, and the unknown samples 4i and 4i(b) (Fig. 3). Further strong support (bootstrap support 80%) was found for this cluster with the addition of *E. dundasii*. This cluster agrees with previous ITS results identifying this as the monophyletic section *Bisectae I* within the subg. *Symphyomyrtus* (Steane et al. 2002). A single base polymorphism at 398 bases (G instead of A) was diagnostic for this cluster, for the samples analysed. All other samples, with known and uncertain identification, were excluded from the *E. articulata* cluster (Fig. 3).

Sequencing (ETS)

We assessed ETS sequences for 41 samples, including 8 taxa obtained from GENBANK (Appendix 4). ETS failed to amplify sufficiently for 1 of 3 *E. loxophleba* samples, presumably due to poor quality DNA from herbarium

samples. The final matrix contained 467 characters, with 58 (12.4%) variable characters. The NJ tree generated a strongly supported cluster (bootstrap support 59%) containing all *E. articulata* samples and the 2 *E. loxophleba* samples (Fig. 4). A single base polymorphism at 362 bases (T instead of C) was diagnostic for this cluster, for the samples analysed. This result is in agreement with the known close relationship between *E. articulata* and *E. loxophleba* (Hines and Byrne 2001). The *E. loxophleba* samples were identical, and significantly (bootstrap support 72%) clustered within the *E. articulata* cluster. All other samples, with known and uncertain identification, were excluded from the *E. articulata* cluster (Fig. 4).

Sequencing (matK)

We assessed *matK* sequences for 24 samples, including 5 of the 6 *E. articulata* samples. Due to poor and difficult amplification for many samples (as has been noted previously for *matK* (Hollingsworth 2008), the total number of samples assessed for *matK* sequence variation was less than for ITS and ETS. The final matrix contained 1697 characters, with 32 (1.9%) variable characters. The NJ tree generated a strongly supported cluster (bootstrap support 70%) containing all 5 *E. articulata* samples, which were themselves identical (Fig. 5). A single base polymorphism at 203 bases (G instead of T) was diagnostic for this cluster, for the samples analysed. All other samples, with known and uncertain identification, were excluded from the *E. articulata* cluster (Fig 5). In addition, a single base state difference at both 982 and 1661 bases differentiated all *E. trivalva* samples from all *E. articulata* samples (982 - *E. trivalva* (T), *E. articulata* (A); 1661 - *E. trivalva* (T), *E. articulata* (C)).

Amplified Fragment Length Polymorphism (AFLP)

For AFLP primer pair 1 (m-CAG/p-AC), we assessed 31 samples for presence/absence variation across 152 markers, of which 150 (99%) were polymorphic. For AFLP primer pair 1 (m-CAG/p-CG), we assessed 31 samples for presence/absence variation across 131 markers, of which 129 (98%) were polymorphic. Ordination of data from each primer pair showed clustering of all *E. articulata* samples, and marked differentiation (non-overlap) between the *E. articulata* cluster and all other samples (Fig 6, 7). Plotting the first and third principal co-ordinate axes of the m-CAG/p-CG data matrix revealed a cluster of all known *E. trivalva* samples with samples 4i and 4i(b) (Fig. 8), suggesting that these 2 samples are *E. trivalva*, in agreement with the ITS and *MatK* sequence data (Figs. 3, 5). Fixed marker differences between *E. articulata* and all other samples, or *E. articulata/E. loxophleba* and all other samples, were found at numerous AFLP markers (Table 3).

Fig. 3: Neighbor-joining tree of ITS sequence data. Bootstrap support values above 50% are indicated. Note significant cluster (indicated) containing all *E. articulata* samples, all *E. trivalva* samples, 4i and 4i(b), and *E. wandoo*.
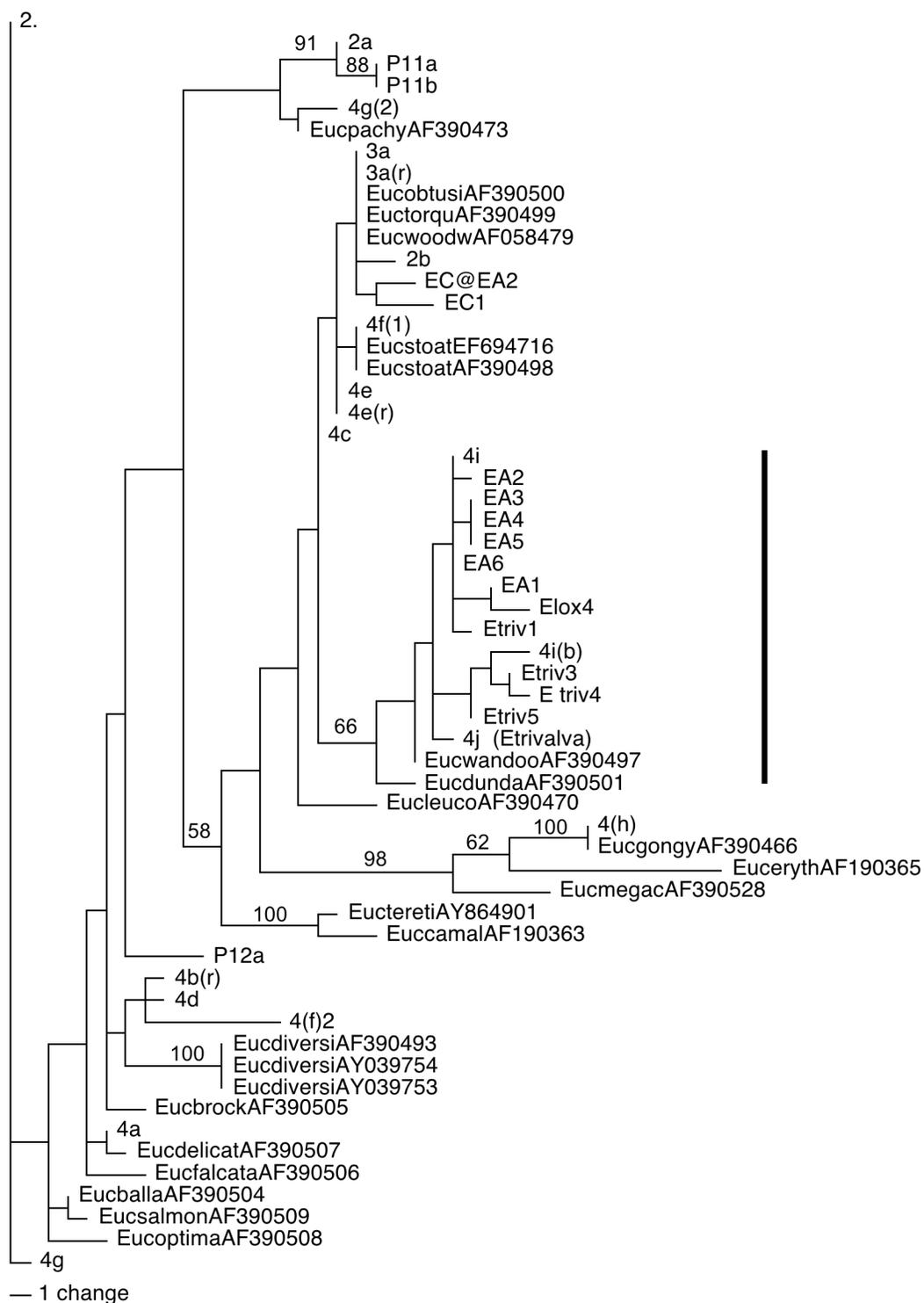

2.

91 ┌ 2a
88 ┤ P11a
   └ P11b
┌ 4g(2)
└ EucpachyAF390473
3a
3a(r)
EucobtusiAF390500
EuctorquAF390499
EucwoodwAF058479
2b
EC@EA2
EC1
4f(1)
EucstoatEF694716
EucstoatAF390498
4e
4e(r)
4c

4i
EA2
EA3
EA4
EA5
EA6
EA1
Elox4
Etriv1
4i(b)
Etriv3
E triv4
Etriv5
66 4j (Etrivalva)
EucwandooAF390497
EucdundaAF390501
EucleucoAF390470

100 ┌ 4(h)
62 ┤ EucgongyAF390466
98   EucerythAF190365
EucmegacAF390528

100 ┌ EucteretiAY864901
    └ EuccamalAF190363
58
P12a
4b(r)
4d
4(f)2
100 ┌ EucdiversiAF390493
    ├ EucdiversiAY039754
    └ EucdiversiAY039753
EucbrockAF390505
4a
EucdelicatAF390507
EucfalcataAF390506
EucballaAF390504
EucsalmonAF390509
EucoptimaAF390508
4g

— 1 change

Molecular Forensic Identification of *Eucalyptus articulata*

Fig. 4: Neighbor-joining tree of ETS sequence data. Bootstrap support values above 50% are indicated. Note significant cluster (indicated) containing all *E. articulata* samples, and *E. loxophleba* samples.
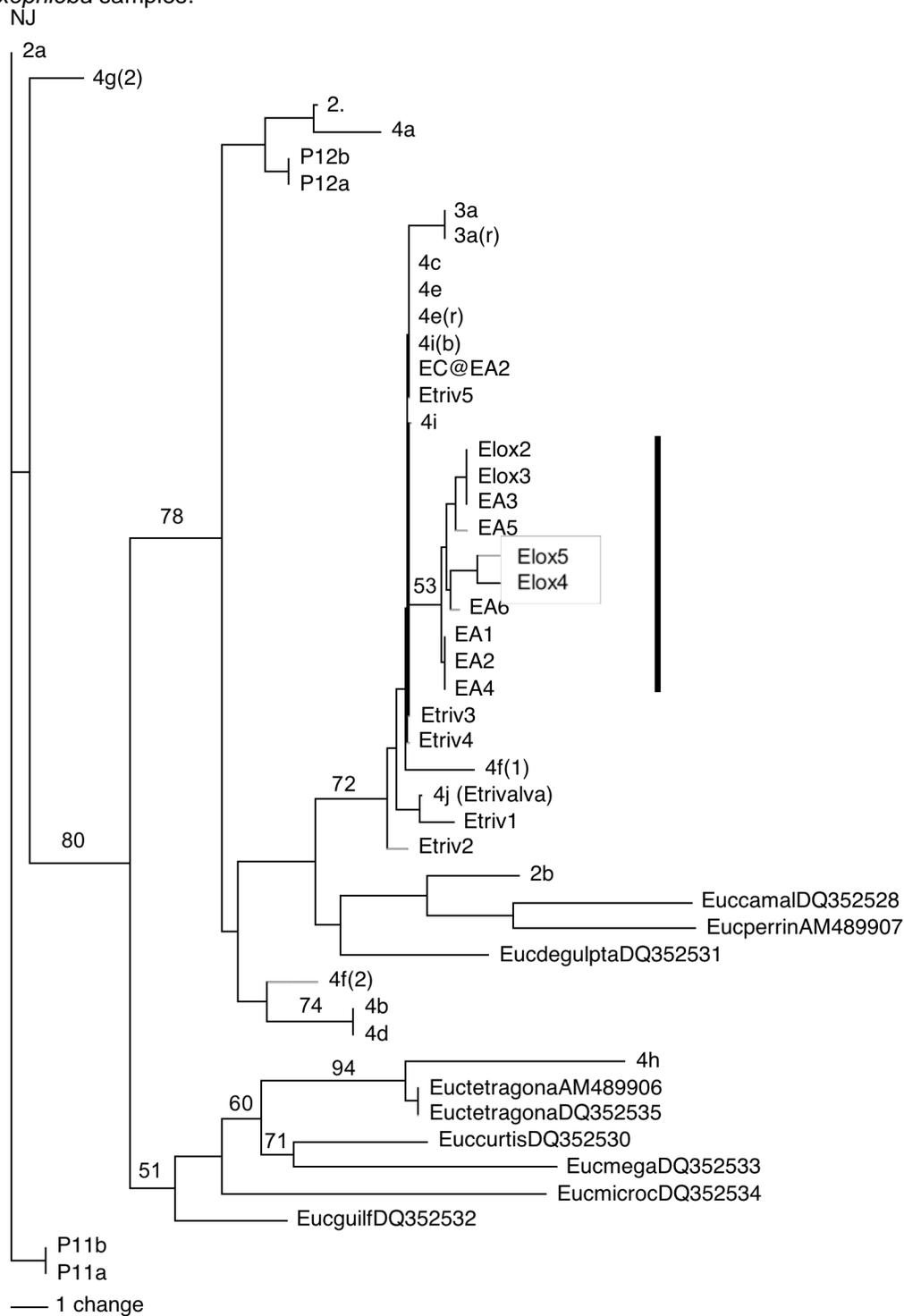
Fig. 5: Neighbor-joining tree of *matK* sequence data. Bootstrap support values above 50% are indicated. Note significant cluster (indicated) containing all and only *E. articulata* samples.

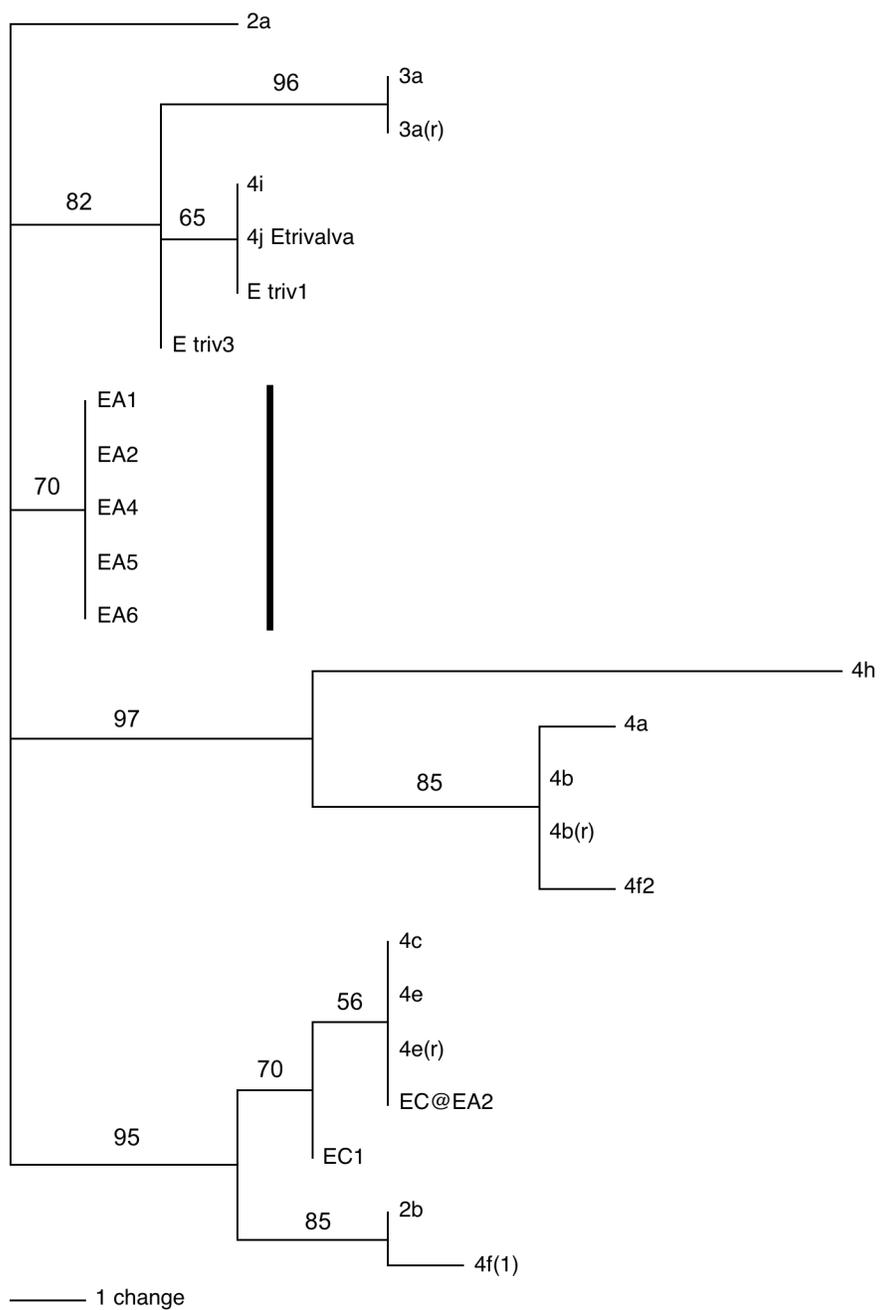Molecular Forensic Identification of *Eucalyptus articulata*

Fig. 6: Ordination showing genetic similarities among eucalypt samples for AFLP products from primer pairs m-CAG/p-CG. Plotted are the first two principal co-ordinates, which account for 48% of the total variation. Note "EA1" to EA6" are *Eucalyptus articulata* samples that are strongly differentiated from all other samples. See also Table 1 for sample codes.
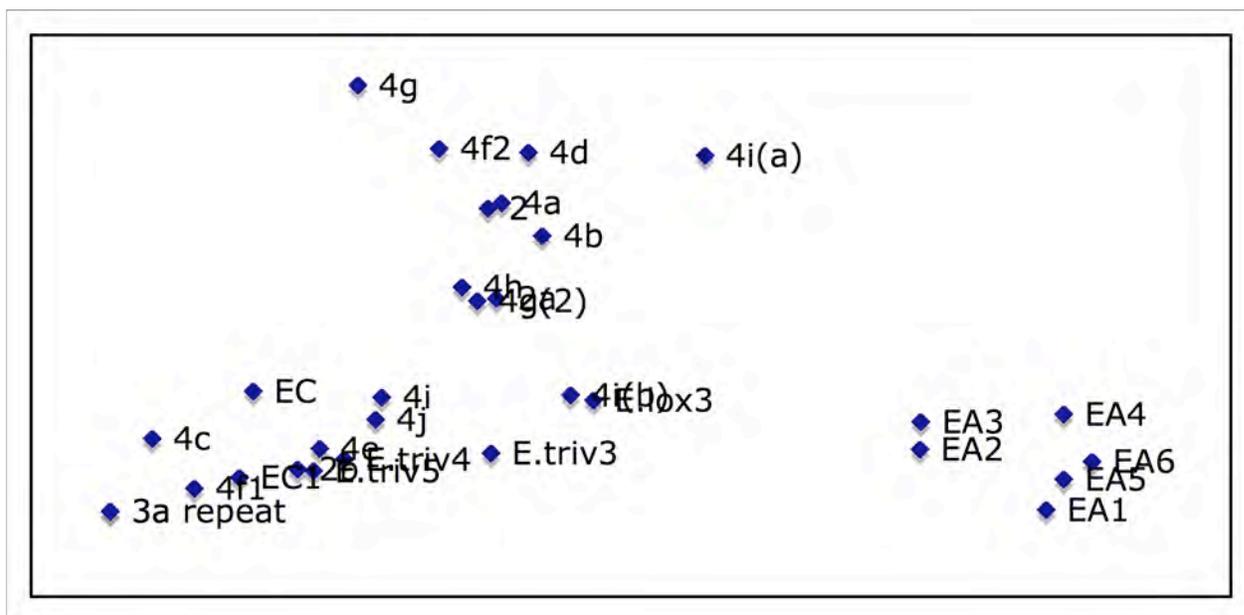

Fig. 7: Ordination showing genetic similarities among eucalypt samples for AFLP products from primer pairs m-CAG/p-AC. Plotted are the first two principal co-ordinates, which account for 46% of the total variation. Note "EA1" to EA6" are *Eucalyptus articulata* samples that are strongly differentiated from all other samples. See also Table 1 for sample codes.

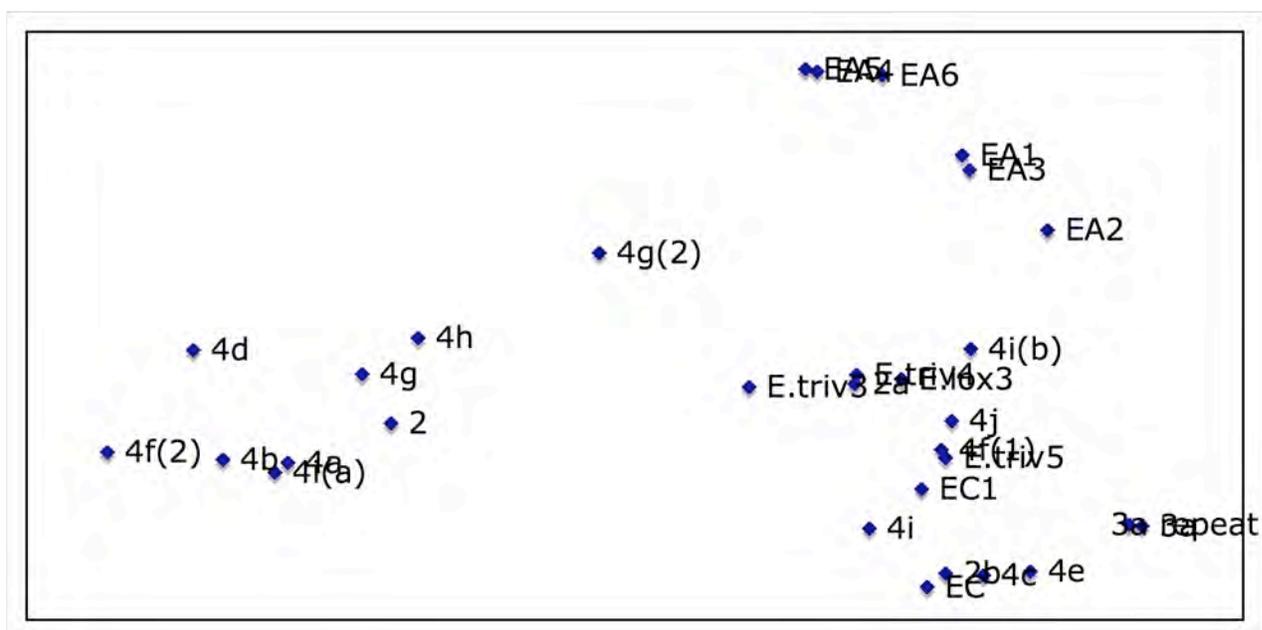Molecular Forensic Identification of *Eucalyptus articulata*

Fig. 8: Ordination showing genetic similarities among eucalypt samples for AFLP products from primer pairs m-CAG/p-CG. Plotted are the first and third principal co-ordinates, which account for 46% of the total variation. Note strong overlap of "4i", "4ib", and "4j" with known *E. trivalva* samples ("E.triv3-5"). See also Table 1 for sample codes.
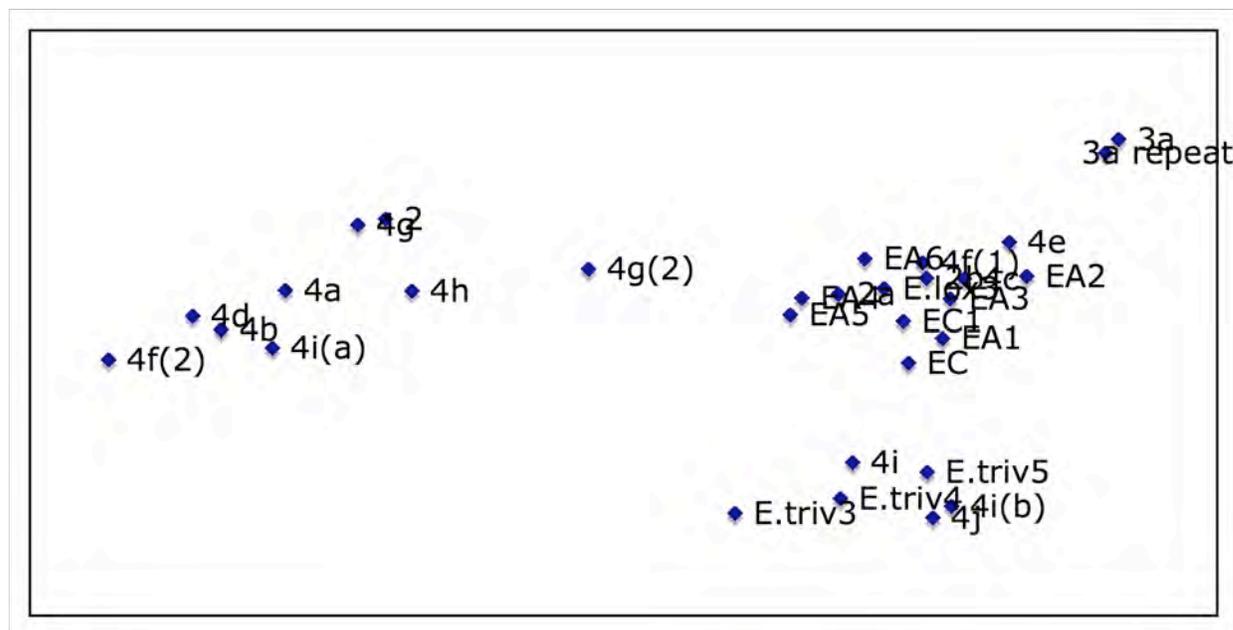

Table 3: Diagnostic AFLP markers for *E. articulata* from primer pairs m-cag/p-ac and m-cag/p-cg

PRIMER	MARKER	DIAGNOSTIC FEATURE
m-cag/p-ac	79	Absent in all *E. articulata*, present in all other samples
m-cag/p-ac	95	Present in all *E. trivalva*, 4i, 4i(b), absent in all other samples
m-cag/p-cg	101	Absent in all *E. articulata, E. trivalva, E.loxophleba*, 4i & 4i(b), present in all other samples.
m-cag/p-cg	263	Present in all *E. articulata & E. loxophleba*, absent in all other samples
m-cag/p-cg	272	Present in all *E. articulata*, absent in all *E. trivalva, E. loxophleba*, 4i & 4i(b)
m-cag/p-cg	277	Present in all *E. articulata & E. loxophleba*, absent in all other samples

Molecular Forensic Identification of *Eucalyptus articulata*

Conclusions

A strong feature of this molecular assessment is the congruence in the results regarding *E. articulata* from four independent molecular marker sets – nr DNA ITS sequence data, nr DNA ETS sequence data, cpDNA *matK* sequence data, and AFLP DNA fingerprinting data. All four data sets identified genetic clusters of the *E. articulata* samples that, with the partial exception of ITS sequencing, significantly excluded all other samples of unknown or uncertain taxonomy. For the ITS data, only identification uncertain samples 4i and 4i(b) clustered with *E. articulata*, along with known *E. trivalva* and *E. wandoo*. Ultimately, numerous diagnostic AFLP markers were identified that differentiated *E. articulata* samples from all others. Diagnostic sequence variation for *matK* and ETS was found that uniquely defined the *E. articulata* samples from all others, with the exception of *E. loxophleba*. One current limitation in our results is the absence of *E. loxophleba* for *matK*, and we are currently attempting to overcome amplification difficulties to address the ability of *matK* to differentiate between these close relatives. However, this does not affect the conclusions with regards the absence of *E. articulata* from the specimens of uncertain taxonomy.

From this genetic assessment, we confidently conclude that the taxonomically uncertain or unknown eucalypts sampled and provided to us for genetic assessment do not include *E. articulata*. This conclusion is supported by the limited morphological material provided for voucher specimens, and is supported by the tentative conclusions drawn in the field by Mattiske staff undertaking the sampling (pers comm).

While our objective was specifically focussed on confirming or rejecting the identity of eucalypt samples as *E. articulata*, the association of samples 4i and 4i(b) to known *E. trivalva* from ITS sequence data and AFLP data, and 4i to known *E. trivalva* from matK sequence data (sample 41b did not amplify here for matK) strongly suggests that we are able to conclude the identification of these samples as *E. trivalva*. This conclusion supports those made by Mattiske staff from limited morphological material in the field. In addition, we are confident that the P11 and P12 samples (taxonomy unrevealed to us) are not *E. articulata*. From the limited sampling, our best guess is that P11 is most likely *E. rosaceae* or close relative, and P12 is most likely *E. oleosa* or close relative – but these conclusions are tentative due to the limited sampling. More generally though, we are unable to confidently confirm nor deny the identifications made on other samples, and this was not our specific objective. If desired, then more detailed sampling of these other species is required for a confident DNA barcoding identification of these non-*E.articulata* species.

Whilst the genetic results are clear with regards the absence of *E. articulata* from the samples provided to us from burnt eucalypt mallees sampled along the proposed road route, this result does serve to reinforce the highly restricted nature of the distribution of the DRF *E. articulata*, and the apparently specific

substrate requirements (red sand dunes, sandy loams, arkose rubble) of this species.

More generally, debate continues on the most suitable combination of gene regions for universal DNA barcoding of plants (Hollingsworth 2008; Lahaye et al. 2008). To this debate, our results highlight the particular utility of *matK* for the specific objectives addressed here within *Eucalyptus*. This supports conclusions made elsewhere in the Myrtaceae, where O'Brien et al. (2000) found *matK* to be phylogentically informative at both generic and species level in the *Leptospermum* suballiance. In particular, *matK* appears superior to ITS or ETS due to its apparent lack of variation within the species across 1697 bases, yet it was significantly discriminated from all other samples, and particularly *E. trivalva*, which caused some ambiguity for ITS. That is, *matK* best exhibited a "barcoding gap" between inter versus intraspecific divergences (Meyer & Paulay 2005). This is despite generating far fewer informative characters than ITS or ETS (*matK* 32 of 1697 bases (1.6%) variable for 24 samples; *ITS* 80 of 640 (12.5%) 58 samples; ETS 58 of 467 (12.4%) 41 samples). While these results are biased in an absolute sense due to different sampling intensities (ie fewer taxa), and the absence of *E. loxophleba* from the final *matK* assessment, they do identify the utility of matK for barcoding in eucalypts.

The observed DNA barcoding utility of *matK* in this eucalypt study is in agreement with the conclusions of Lahaye et al (2008), and the recommendation of the International Barcode of Plants consortium, with regards the utility of *matK* (in addition to *rbcL*) as the most suitable universal barcoding gene for plants. Extending the DNA barcoding research commenced here to WA eucalypts more generally would contribute significantly to the practical issue of rapid identification of eucalypts of ambiguous identity (by establishing a DNA barcoding database), as well as contributing significantly to the broader issue of DNA barcoding in plants generally. However, this must proceed with the understanding that DNA barcoding presents significant challenges for delineating closely related species in taxonomically understudied groups (Meyer & Paulay 2005).

References

Baldwin BG et al (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals Missouri Botanic Garden* **82**, 247-277.

Baldwin BG and Markos S (1998) Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: congruence of ETS and ITS trees of *Calycadenta* (Compositae). *Molecular Phylogenetics and Evolution.* **10**, 449-463.

Brooker MIH et al (2006) Euclid: Eucalypts of Australia. CSIRO Publishing.

Cowans RS et al (2006) 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. *Taxon* **55**, 611-616.

Gadek PA, Wilson PG, Quinn CJ (1996) Phylogenetic reconstruction in Myrtaceae using *matK,* with particular reference to the position of *Psiloxylon* and *Heteropyxis. Australian Systematic Botany* **9**, 283-290.

Glaubitz JC, Emebiri LC and Moran GF (2001) Dinucleotide microsatellites from *Eucalyptus seeberi*: inheritance, diversity, and improved scoring of single-base differences. *Genome,* **44 (6)**, 1041.

Hines B & Byrne M (2001) Genetic differentiation between mallee and tree forms in the Eucalyptus loxophleba complex. *Heredity* **87**, 566-572.

Hollingsworth ML et al (2009) Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources* **9**, 439-457.

Hollingsworth PM (2008) DNA barcoding in biodiversity hotspots: progress and outstanding questions. *Heredity* **101**, 1-2.

Lahaye R et al. (2008) DNA barcoding the floras of biodiversity hotpsots. *PNAS* **105**, 2923-2928.

Meyer CP & Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* **3**, 2229-2238.

Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution* **14**, 389-394.

Muluvi GM, Sprent JI, Soranzo N, Provan J, Odee D, Folkard G (1999) Amplified fragment length polymorphism (AFLP) analysis of genetic variation in *Moringa oleifera* Lam. *Molecular Ecology* **8**, 463-470.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Steane DA, Nicolle D, McKinnon GE, Vaillancourt RE, Potts BM (2002) Higher-level relationships among the eucalypts are resolved by ITS-sequence data. *Australian Systematic Botany* **15**, 49-62.

Steele KP, Vilgalys R (1994) Phylogenetic analysis of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. *Systematic Botany* **19**,126-142.

Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Valentini A, Pompanon F, Taberlet P (2008) DNA barcoding for ecologists. *Trends in Ecology and Evolution* **24**, 110-117.

Vos P, Hogers R*, et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407-4414.

Wagner BD, Furnier GR, Saghai-Maroof, Williams SM, Danick BP and Allard RW (1987) Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proceedings National Academy Science USA* **84**, 2097-2100.

White TJ, Bruns T, Lee S and Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315-322. In: *PCR Protocols: A Guide to Methods and Applications,* eds. Innis MA, Gelfand DH, Sninsky JJ and White TJ. Academic Press, Inc., New York.

Wilson PG, Heslewood MM, Quinn CJ (2007) Re-evaluation of the genus *Babingtonia* (Myrtaceae) in eastern Australia and New Caladonia. *Australian Systematic Botany* **20**, 302-318.

Appendix 1: DNA extraction

Eucalyptus DNA Extraction Protocol
(Glaubitz *et al.* 2001)

1. Add Na metabisulfite to proportion of extraction buffer and wash buffer and keep on ice.
2. Grind 1-2g leaf material in liquid nitrogen
3. Add 10ml cold extraction buffer (it will freeze in the mortar- allow to thaw before the next step).
4. Filter through muslin (in a funnel) into 10 ml tube, and keep on ice.
5. Mix and spin at 2000 rpm for 10 min to pellet.
6. Pour off supernatant; invert tubes to drain for 1 minute.
7. Add 1 ml wash buffer, resuspend pellet
8. Add 100 µl 20% SDS (final concentration 2%). Mix.
9. Add 100 µl 20% Triton X-100 (final concentration 2%). Mix.
10. Mix gently, incubate at room temperature for 5 minutes, with occasional mixing by inversion.
11. Add 200 µl 5M NaCl. Mix. (final concentration 0.7M)
12. Add 500 µl 8.6% CTAB/0.7M NaCl. Mix.
13. Incubate in 65^OC water bath for 20-30 minutes.
14. Add 2 volumes (3.6 ml) chloroform:isoamyl alcohol (24:1); place on the shaker for at least 2 minutes.
15. Centrifuge at 5000 rpm for 10 minutes.
16. Transfer supernatant (top layer) to new tube.
17. Add 2/3 volume isopropanol. Mix.
18. Hook out DNA with the tip of a sealed glass pipette. (If unhookable, centrifuge for 8 minutes at 2000 and remove supernatant.)
19. Place pipette in an eppendorf tube and wash in 1 ml of 50% isopropanol/0.3M Ammonium acetate (NH_4Oac) overnight (or for several hours) at 4^OC.
20. If not on a pipette re-centrifuge to pellet DNA and pour off supernatant.
21. Air dry DNA
22. Re-suspend in minimal volume of warm (65^OC) TE

Eucalyptus Extraction Buffer

	For 1 litre	For 500 ml
0.35M Sorbitol	64 g	32 g
100 mM Tris	12.1 g	6.05 g
100 mM Boric Acid	6.2 g	3.1 g
25 mM EDTA pH 8.0)	50 ml (0.5M pH 8.0)	25 ml (0.5 ml
1M NaCl	58.4 g	29.2 g
Adjust pH to 8 before adding PEG and PVP		
10% PEG 8000[1]	100 g	50 g
2% PVP 40 000[2]	20 g	10 g
0.5% BSA[3]	5 g	2.5 g

0.1% spermine	1 g	0.5 g	
0.1% spermidine[4]	1 g	0.5 g	
0.2% sodium metabisulphite[5]	2 g	1 g	

Notes:
[1]PEG may need to be placed on the stirrer to dissolve
[2]Make a paste of PVP in a little dH_2O before adding
[3]Buffer containing BSA must be soaked in bleach after use
[4]Spermidine is optional
[5]Add sodium metabisulphite on the day of use

Eucalyptus Wash Buffer (From Wagner et al 1987)

	500 ml	50 ml	100 ml
50 mM Tris	3.025 g	0.3 g	0.6 g
25 mM EDTA (use 0.5M, pH 8)	25 ml	2.5 ml	5 ml
0.35 M sorbitol	32 g	3.2 g	6.4 g
0.5% spermine	2.5 g	0.25 g	0.5 g
0.5% spermidine	2.5 g	0.25 g	0.5 g
0.2% sodium metabisulphite*	1 g	0.1 g	0.2 g

*Add on the day of use

Notes
- Always keep extraction and wash buffer in the fridge
- SDS (sodium dodecyl sulfate), to make 10 ml of 20% SDS, use 2 g and make up to 10 ml with dH_2O.
- Triton –X 100, to make 10 ml of 20%, use 2 ml and add 8 ml dH_2O.
- To make 100 ml of 50% isopropanol/0.3M ammonium acetate. Make up in the fume hood and store in the fridge. Make up 0.6M ammonium acetate and add 50 ml of this to 50 ml 100% isopropanol for a total of 100ml.
- To make 100 ml 8.6% CTAB/0.7M NaCl: Add 8.6 g CTAB and 4.09 g NaCl and make up to 100 ml with dH_2O.

References.

Glaubitz JC, Emebiri LC and Moran GF (2001) Dinucleotide microsatellites from *Eucalyptus seeberi*: inheritance, diversity, and improved scoring of single-base differences. Genome, 44 (6), 1041.

Wagner BD, Furnier GR, Saghai-Maroof, Williams SM, Danick BP and Allard RW (1987) Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. Proc. Natl. Acad. Sci. USA, 84, 2097-2100

Fig. 1.1 Agarose gel of demonstrating quantity and quality of DNA extracted from *Eucalyptus* species (lanes from left to right correspond to numbers 1- 19 in table). The final lane (right) is a DNA ladder.
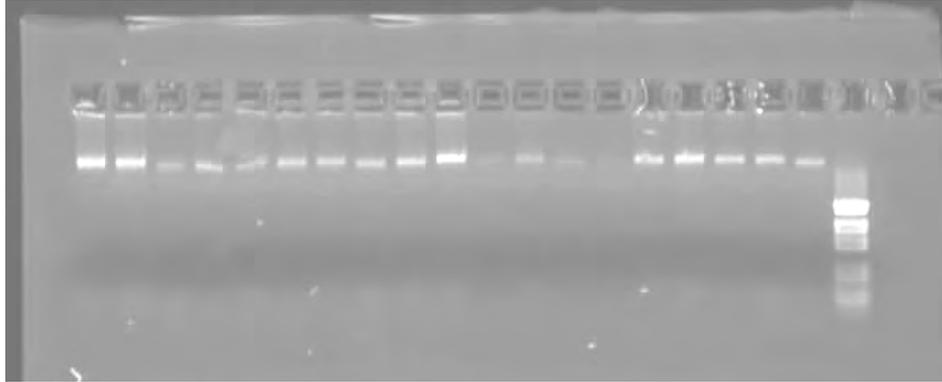

Table 1.1: Quantity of DNA extracted from *Eucalyptus* leaf material

Number/symbol		Species	DNA ng/µL
1.	Ec1	*E. concinna*	14.59
2.	4(f)1	Burnt mallee	32.02
3.	Ea5	*E. articulata*	19.99
4.	Ea6	*E. articulata*	26.51
5.	Ec	*E. concinna*	10.86
6.	Ea2	*E. articulata*	25.32
7.	Ea1	*E. articulata*	37.31
8.	Ea3	*E. articulata*	28.32
9.	2b	*E. pimpiniana*	25.67
10.	2	*? E. oleosa*	21.55
11.	4c	Burnt mallee	6.73*
12.	4b	Burnt mallee	5.19*
13.	4e	*? E. concinna,* burnt mallee	4.12*
14.	4d	mallee	0.48*
15.	4(f)2	Burnt mallee	11.72
16.	4g	*? E. mannensis*	19.56
17.	4i	Burnt mallee	18.69
18.	3a	*? E. transcontinentalis*	15.9
19.	4j	*E. trivalva*	5.73*
20.	Ea4	*E. articulata*	52.99
21.	2a	*? E. rosaceae*	38.62
22.	4a	*? E. oleosa*	59.39
23.	4i(b)	Burnt mallee	26.28
24.	4i(a)	Burnt mallee	241.27
25.	4h	*E. gongylocarpa,* burnt	14.21
26.	4(g)2	*E. youngiana*	9.04*
27.	E.lox1	*E. loxophleba* (herbarium)	2.01

28.	E.lox2	*E. loxophleba* (herbarium)	4.87
29.	E.lox3	*E. loxophleba* (herbarium)	2.17
30.	E.triv1	*E. trivalva*	36.77
31.	E.triv2	*E. trivalva*	20.07
32.	E.triv3	*E. trivalva*	20.35
33.	E.triv4	*E. trivalva*	44.2
34.	E.triv5	*E. trivalva*	57.45

* These specimens were re-extracted due to the low yield (data not shown).

Appendix 2: DNA sequencing protocol

Gene Amplification – PCR 1

Concentrations for ETS. Standard modifications are: changing $MgCl_2$ concentration to vary primer specificity, and slight changes in concentration of primers and Taq [add more or less H_2O as required to make to 45 µL]. You may also wish to double amounts to do a 100 µL total volume. Mix a stock batch for N samples (+10% of N) xf each reagent below [the 10% is an extra amount ensure you don't run out of stock master mix before reaching the last tube]

EXAMPLE STOCK MASTER MIX ONLY:
Stock:
Distilled H2O	22.9 µL
5X polymerisation buffer	10 µL
$MgCl_2$ (50 mg/µL)	2 µL
Fwd Primer (5 µM)	5 µL
Rev Primer (5 µM)	5 µL
Taq (5.0 U/µL)	0.1 µL

Vortex stock and spin briefly in centrifuge.
Add 45 µL of stock to each well of your PCR plate.
Add 5 µL of your sample DNA.
Vortex the plate to mix reagents and spin the plate in the lettuce centrifuge.
Run your relevant gene amplification PCR profile (usually obtain from a published procedure for the region you are using – slight changes to annealing temperature (48-55°C) may help amplify uncooperative samples).

EXAMPLE PCR ONLY:
Hold: 95°C for 90 sec
35 cycles of:
       95°C for 30 sec
       55°C for 60 sec
       72°C for 90 sec
Hold 72°C for 7 min
Hold indefinitely at 4°C

Check product on agarose gel. Check that products are around the expected size range, and that there is only a single band (sometimes weak secondary bands will make no difference, but if you see equally strong bands or common secondary bands you need to re-optimise the PCR conditions).

If you don't get any, or only partial amplification (and are sure your DNA is ok), try the following modifications to the PCR:

(1) Add DMSO and BSA to the PCR reaction if not already used. DMSO assists the primer to bind to the template and BSA helps bind PCR inhibitors.
(2) Lower the annealing temperature.
(3) Increase the $MgCl_2$ concentration.
(4) Try a "touchdown" PCR [often these don't work as reliably as standard PCR, but some templates seem to preferentially amplify with touchdown]

PCR 1 cleanup [using Agencourt AMPure Kit]

ASSUMING 46 µL PCR Volume [50 µL reaction minus 4 µL for agarose gel]

*Make new labelled tubes IF you are going to transfer the cleaned product to eppendorf tubes rather than leave them in the plate.

1. Transfer successfully amplified PCR products to a new plate so there are no gaps (not necessary if all samples successfully amplified).
2. Gently shake AMPure bottle to resuspend magnetic particles.
3. Add 82.8 µL AMPure to each well [ie 1.8 x reaction volume].
4. Seal tubes with lids* and vortex strongly for *at least* 30 s. Check all reagents are thoroughly mixed and homogenous before continuing. *Note: if a plate cover is used it MUST be able to seal isopropanol during heavy vortexing, and be VERY careful of contamination between wells – lids are much safer but will still leak if not completely closed.
5. Spin plate briefly in lettuce spinner. Incubate samples OFF THE MAGNET for minimum 3-5 mins to bind extension products.

Binding
6. Place plate on magnet for 5-10 mins (or until solution is very clear).
7. WHILE ON MAGNET, remove clear supernatant from wells by placing pipette tip on bottom of plate. Remove as much supernatant as possible. [Use a multichannel pipette, but don't touch the bead ring].

First ethanol wash.
8. WHILE ON MAGNET, Add 200 µL of 70% ethanol to each well. [Mixing and resuspension is NOT necessary.]
9. Incubate at room temp for 30 seconds.
10. WHILE ON MAGNET, remove clear supernatant from wells by placing pipette tip on bottom of plate. Remove as much supernatant as possible.

Second ethanol wash.
11. WHILE ON MAGNET, Add 200 µL of 70% ethanol to each well.
12. Incubate at room temp for 30 seconds.
13. WHILE ON MAGNET, remove clear supernatant from wells by placing pipette tip on bottom of plate. Remove as much supernatant as possible.
14. Place plate on bench (off the magnet) to air-dry for 10-20 minutes (sometimes longer). Plate must be completely dry. [**Note**: the plate can be sealed and stored dry at -20^oC indefinitely here.]

15. Add [20-40 µL] 30 µL elution buffer (TE, DI H$_2$O or Tris-Acetate) to each well. **Note:** 30 µL seems to be a good balance, but particularly weak bands on the agarose gel should be concentrated by adding only 20 µL (the difference between getting acceptable vs too weak and noisy sequences). Extremely strong bands could be diluted by using 40 µL, but it is safer to dilute them later after measuring DNA concentration.
16. Vortex strongly for 30 seconds [critical if using less than 40 µL to elute].
17. Place plate on magnet for 2-5 mins until solution is clear [**Note**: with 30 µL of liquid you can't actually see the ring of magnetic beads from the side, only from the top]. [NOTE: the plate can be sealed and stored at 4°C or –20°C indefinitely here.]
18. OPTIONAL: Transfer liquid to labelled eppendorf tube. [**Note**: beads carried over do not inhibit further reactions – but they CAN affect quantification]

Cycle Sequence PCR

Note: CEQ dyes are light sensitive and degrade over time (within a day or two in formamide). Keep any dye-containing reagents or mixes on ice. Plan to do the dye cleanup of the cycle sequence reaction within 18-24 hours of starting the cycle sequence thermocycling. Samples can be left dry indefinitely at –20°C AFTER the dye removal step if necessary.

Note: DNA concentration is an important variable for sequencing reactions. It is best, at least initially, to quantify your *cleaned* PCR products on the nanodrop spectrophotometer. Ideally you want concentrations in the order of:
[6]10-40[60] ng/µL for PCR products 400-800 bps long
[10]15-65[80] ng/µL for PCR products 900-1200 bps long
[15]20-80[120] ng/µL for PCR products 1300-2000 bps long
Concentrations lower than the lower bracketed limit will definitely NOT give good product, with variable success up to the lower non-bracketed limit. Concentrations higher than the upper-bracketed limit could cause potential problems, especially in larger size PCR products, and should ideally be diluted, at least until you have a feel for how your sequences work.

NOTE: Beckman-Coulter (CEQ) process uses a different dye chemistry to ABI systems used elsewhere.
Note: we usually use a 1/4 reaction, and it works reliably in all cases so far.
NOTE: For a sequence reaction you must add only one primer to get readable sequence. A short (< c. 5-600 bps) region is usually sequenced in two reactions using the same primers for the PCR. Longer regions need internal primers as well to get all of the sequence
The cycle sequence reaction requires the following reagents per reaction [for a 1/4 reaction, 10 µL total volume]:

5X Cycle Sequence Buffer 1.0 µL

Primer (5 uM)   0.64 µL [ie 3.2 pmoles of primer in 10 µL reaction]
Dye Terminator Mix   2.0 µL
DNA   [20]40-80[200] ng SEE BELOW
Distilled H_2O to 10 µL total volume

There are many ways to achieve this. Typically we mix as many reagents as practical into one or many "stock mixes", aliquot them to each well then add DNA and any other missing reagents as required.

PROTOCOL
How much DNA do I add?
You need a total volume 4 µL of DNA + distilled H_2O for the reaction
Initially, you should aim for 40-80 ng of DNA, depending on size of the amplified PCR product. We suggest aiming for [sequence reaction will almost certainly be too weak below the bracketed DNA amounts]:
0.5-0.7 kbases use [20]40 ng DNA (double-stranded)
0.8-1 kbases use [25]50-55 ng DNA (double-stranded)
1.1-1.5 kb use [30]65 ng DNA (double-stranded).
1.5-2 kb use [40]80 ng DNA (double-stranded).
Once you have a feel for how strong your PCR products typically are (in multiple reactions from material from different extractions), you can probably simplify set up considerably by using 4 µL of cleaned PCR product *regardless of concentration* as template for your sequence reaction, especially if you tweak the final elution volume at the end of PCR cleanup based on strength of agarose gel bands.

1. Stock reagent mix

Mix a stock batch for N samples (+10% of N) x each reagent below [the 10% of N is an extra amount to make sure you don't run out before reaching the last tube]. Remember you typically need 1 forward reaction and 1 reverse reaction for each sample. Stock mixes could be made for EACH primer (as below), or be made for all samples (excluding primers in stock mix, total vol. = 5.36 µL to each well), and adding 0.64 µL of 5 µM primer to each well. Which one is more efficient depends on the number of samples and the number of different primers used for each sample.
Amounts per reaction [for a 1/4 reaction, 10 µL total volume]:
Distilled H20   2.36 µL
5X Cycle Sequence Buffer 1.0 µL
Primer (5 uM)   0.64 µL [ie 3.2 pmoles of primer in 10 µL reaction]
Dye Terminator Mix   2.0 µL
2. Vortex the stock mix and centrifuge briefly.
3. Aliquot the relevant amount to each PCR well (eg 6 µL if you include the primer with the stock mix, 5.36 if you add the primers separately).
4. Add Primer (if necessary): ONLY IF you have not included primers in stock mixes, add 0.64 µL of 5 µM primer to each well.

5. Add DNA: If you have used the AMPure kit AND have not transferred the final eluant to a new tube (ie still contains magnetic beads), place plate on magnet for 2-5 mins until solution is clear. Transfer the appropriate amount of DNA [eg. 4 µL] to the sequence reaction. [**Note**: beads carried over do not inhibit further reactions – but they CAN affect quantification].
6. Seal plate with lids and vortex the plate to mix reagents and spin the plate in the lettuce centrifuge.
7. Run the CEQ cycle sequence PCR profile. This is usually the same for all templates, however occasionally it might need to be changed for difficult primer/templates [never yet in our experience].

This protocol can be found on PCR1 and PCR2 as: ***MattCEQcycseq***
96^oC for 20 sec
30 cycles of:
  96^oC for 20 sec
  50^oC for 20 sec
  60^oC for 4 min
Hold 4^oC indefinitely

Dye removal [Using Agencourt CleanSEQ protocol]
Note: The protocol takes about 2 hours for a plate, after which EITHER:
(1) If you add formamide at the last step the samples MUST be run on the Beckman immediately (within a few hours) as formamide degrades the dye, OR
(2) DO NOT add formamide after the last 20 min (no more) drying stage and store the dried product at $–20^oC$ indefinitely until you are ready to run them.
Note: the following assumes a 10 µL cycle sequence reaction volume.

Before starting, you need:
1. Enough CleanSEQ bead suspension for N x 10 µL (N=no. samples to clean)
2. Make some FRESH 73% Isopropanol (need 455 uL per 10 µL reaction) Dilute as to table below (volumes in µL) **NOTE**: 73% Isopropanol can be kept only 1-3 days in $–20^oC$ freezer; best to make fresh each day.

No. Samples	100% isopropanol	H_2O	Total Vol. (455 µL per sample x 1.1 dead volume)
1	365	135	501
8	2923	1081	3640
16	5846	2162	7280
24	8769	3243	10920
32	11692	4324	14560
40	14615	5405	18200

Molecular Forensic Identification of *Eucalyptus articulata*

48	17538	6487	21840
56	20460	7568	25480
64	23383	8649	29120
72	26306	9730	32760
80	29229	10811	36400
88	32152	11892	40040
96	35075	12973	43680

3. Make STOP solution (5.5 µL for every 10 µL reaction) as in the table below. Calculate Total vol needed (in µL) is: TV = 5.5 µL x N reactions x 1.1 dead volume.

No. samples	H$_2$O	3M Sodium Acetate pH 5.2	100 mM Na2EDTA pH 8	20 mg/µL glycogen (in kit)	Total (5.5 µL per sample x 1.1 dead volume)
1	3.0	1.2	1.2	0.6	6.05
8	24.2	9.7	9.7	4.8	48.4
16	48.4	19.4	19.4	9.7	96.8
24	72.6	29.0	29.0	14.5	145.2
32	96.8	38.7	38.7	19.4	193.6
40	121	48.4	48.4	24.2	242.0
48	145.2	58.1	58.1	29.0	290.4
56	169.4	67.8	67.8	33.9	338.8
64	193.6	77.4	77.4	38.7	387.2
72	217.8	87.1	87.1	43.6	435.6
80	242	96.8	96.8	48.4	484.0
88	266.2	106.5	106.5	53.2	532.4
96	290.4	116.2	116.2	58.1	580.8

(Check you have at least 5.5 µL for each sample)

PROTOCOL [*ASSUMING 10 µL sequence reaction*]
NOTE: retain lids for mixing and close them tight when requested to ensure no isopropanol is escaping during vortexing. Alternatively, can use multichannel pipette to pipette-mix 10-15 times to homogenise solutions when required.
Add STOP solution.
1. Add 5.5 µL dilute STOP solution to each sample well (in plate).

Prepare & add CleanSEQ
2. Shake CleanSEQ vigorously to resuspend magnetic beads (should be homogenous and uniform in colour).
3. Add 10 µL of CleanSEQ to each well.
Isopropanol precipitation
4. Add 55 µL of FRESH 73% Isopropanol to each well

Molecular Forensic Identification of *Eucalyptus articulata*

5. Seal tubes and vortex strongly for *at least* 30 s. Check all reagents are thoroughly mixed and homogenous before continuing.
6. Incubate samples OFF THE MAGNET AND IN THE DARK for minimum 10 mins to bind extension products.

Binding
7. Place plate on magnet for 3-5 mins or until solution is clear.
8. WHILE ON MAGNET, remove clear supernatant from wells by placing pipette tip on bottom of plate. Remove as much supernatant as possible. [Should be able to do this with a multichannel pipette, but don't disturb magnetic suspension]

First Isopropanol wash.
9. WHILE ON MAGNET, Add 200 µL of 73% isopropanol to each well. [Mixing and resuspension is NOT necessary.]
10. Incubate at room temp for 3 mins.
11. WHILE ON MAGNET, remove clear supernatant from wells by placing pipette tip on bottom of plate. Remove as much supernatant as possible.

Second Isopropanol wash.
12. WHILE ON MAGNET, Add 200 µL of 73% isopropanol to each well. [Mixing and resuspension is NOT necessary.]
13. Incubate at room temp for 3 mins.
14. WHILE ON MAGNET, remove clear supernatant from wells by placing pipette tip on bottom of plate. Remove as much supernatant as possible.

IF RUNNING SEQUENCES IMMEDIATELY, START THAWING DEIONISED FORMAMIDE.

Air drying
15. OFF THE MAGNET Let samples air-dry for 10-20 mins at room temp. DO NOT OVERDRY! (longer than 30 mins lowers signal)

[Long-term storage]
16. Stop here if you are not running immediately: seal plate with lids and keep DRY at -20°C indefinitely before resuming protocol when ready. Otherwise proceed to (14).

Re-suspension
17. OFF THE MAGNET, add 35 0L de-ionised formamide to each well. **NOTE: multiple re-thawings of deionised formamide may prevent full elution.**
18. Seal and vortex until beads are homogenously resuspended in deionised formamide (about 30 secs).

Binding
19. Separate on magnet for 3-5 mins or until solution is clear.
20. Transfer 30 µL to CEQ plate for running. **NOTE**: need to leave 5-10 µL behind to prevent bead transfer. Some bead transfer is OK, but must take care to minimise it. If too much is sucked up, re-dispense back into the well, wait 1-3 mins and re-transfer.
21. Centrifuge plate if necessary to remove air bubbles.
22. Add a drop of mineral oil to samples.

Molecular Forensic Identification of *Eucalyptus articulata*

Setup and loading sequencer

Prepare buffer tray, wetting tray and check sample plate has not bubbles and you have added mineral oil. Add gel cartridge and check gel quantity on sequencer. Open sample plate menu. Set run conditions to LFR-1. Select all samples and check analysis parameters and choose DefaultSequenceAnalysisParameters. Run the sample plate.

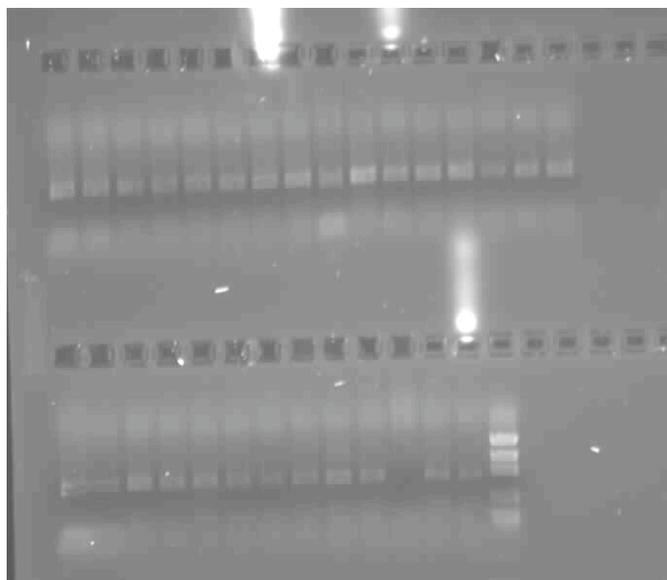Fig. 2.1: Agarose gel demonstrating amplification of the ITS region following PCR.
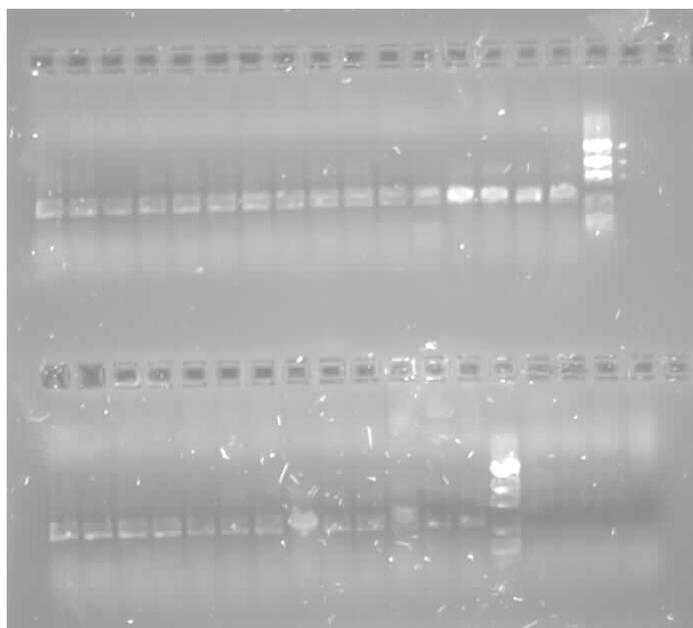

Fig. 2.2: Agarose gel demonstrating amplification of the ETS region following PCR.

Molecular Forensic Identification of *Eucalyptus articulata*

Appendix 3: Eucalypt species for which ITS and ETS sequence data was obtained from GenBank

ETS
Eucalyptus degulpta (DQ352531)
Eucalyptus camaldulensis (DQ352528)
Eucalyptus curtisii (DQ352530)
Eucalyptus perriniana (AM489907)
Eucalyptus megacarpa (DQ352533)
Eucalyptus guilfoylei (DQ352532)
Eucalyptus microcorys (DQ352534)
Eucalyptus tetragona (AM489906)
Eucalyptus tetragona (DQ352535)

ITS
Eucalyptus balladoniensis (AF390504)
Eucalyptus salmonophloia (AF390509)
Eucalyptus falcata (AF390506)
Eucalyptus optima (AF390508)
Eucalyptus delicata (AF390507)
Eucalyptus brockwayi (AF390505)
Eucalyptus diversicolor (AF390493)
Eucalyptus diversicolor (AY039754)
Eucalyptus diversicolor (AY039753)
Eucalyptus pachyphylla (AF390473)
Eucalyptus tereticornis (AY864901)
Eucalyptus leucophloia (AF390470)
Eucalyptus camaldulensis (AF190363)
Eucalyptus stoatei (EF694716)
Eucalyptus obtusiflora (AF390500)
Eucalyptus torquata (AF390499)
Eucalyptus stoatei (AF390498)
Eucalyptus woodwardii (AF058479)
Eucalyptus gongylocarpa (AF390466)
Eucalyptus erythrocorys (AF190365)
Eucalyptus megacarpa (AF390528)
Eucalyptus wandoo (AF390497)
Eucalyptus dundasii (AF390501)

Appendix 4: Amplified Fragment Length Polymorphism (AFLP) Protocol:

AFLP involved three steps, restriction-digestion, pre-selective PCR amplification, and selective PCR amplification (Vos et al. 1995; Mueller & Wolfenbarger 1999). Restriction of genomic DNA was done at 37^oC for 2 hr in a 20µl volume containing approximately 250ng of DNA, 2.5U of *Mse*1 and 5.2U *Pst*1, 2.0µl NE buffer 2 (supplied with *Mse*1 enzyme), 2.0µl 0.1% BSA, and DNA-free water. Next, 5µl of a solution containing 4.0µl *Mse*1/*Pst*1-adapter solution, 0.5µl T4 ligase, 0.5µl ligation buffer (supplied with T4 ligase) was added to the samples and further incubated at 20^oC overnight, then diluted 1/20 in DNA-free water. Pre-selective PCR was performed in a 20µl total volume containing 4.0µl 5X PCR buffer containing dNTPs, 1.2µl $MgCl_2$ (25mM), 0.5µl each of *Pst*1 and *Mse*1 primers (5µM), 0.825U *Taq* DNA polymerase (Fisher Biotech), 4.0µl restricted/diluted DNA template and DNA-free water. The PCR was performed in a PerkinElmer Applied Biosystems 9700 thermal cycler for 20 cycles each at 94^oC for 30s, 56^oC for 2 min, 72^oC for 2 min. A final extension step at 72^oC for 5 min was performed. PCR products were diluted 1/30 with DNA-free water for subsequent, selective amplification. Selective PCR was done in a 10µl total volume containing; 2.0µl 5X PCR buffer containing dNTPs, 0.6µl $MgCl_2$ (25mM), 0.25µl fluorescently-labelled *Pst*1 primer (1µM), 0.5µl *Mse*1 primer (5µM) (Sigma GenoSys), 0.25U *Taq* DNA polymerase, 2.5µl of diluted pre-selective PCR product, and DNA-free water. The selective PCR cycle consisted of a touchdown cycle for 13 cycles at 94^oC for 30s, $65-53^o$C for 30s, 72^oC for 1 min, followed by 25 cycles at 94^oC for 30s, 56^oC for 2 min, 72^oC for 2 min, and a final extension at 72^oC for 2 min. *Pst*I and *Mse*I primers sequences are from Muluvi *et al.* (1999). Six primer pair combinations were used (Table 2). Selective PCRs were amplified separately prior to visualizing with a Beckman CEQ8800 capillary machine, with internal size standard 400. Fragments were scored for the presence (1) or absence (0) of peaks unambiguously between 60 and 400 base pairs. Replicate samples were run on each 96-well plate for consistent scoring of bands. Of the six primer pairs used, two (m-CAG/p-AC and m-CAG/p-CG) generated strong, reliable and reproducible fingerprints.

Figure 4.1. Partial AFLP DNA fingerprints generated by the primer pair m-CAC/p-AC: From top to bottom- two burnt mallee's (4i(a) & 4i(b)), two *Eucalyptus articulata* Ea2 & Ea1*,* two *E. trivalva* (E.triv4 & E.triv3) and one *E. loxophleba* (E.lox3). The peaks in black are the generated fingerprints and the red peaks are the size standard.