

## DETECTION OF HYBRIDS IN COMMERCIALY PROPAGATED *EUCALYPTUS* USING 5S rDNA SEQUENCE

E.T. Steenkamp<sup>1\*</sup>, M.A. van der Nest<sup>2</sup>, M.J. Wingfield<sup>2</sup> & B.D. Wingfield<sup>2</sup>

<sup>1</sup> Department of Biology, University of York, PO Box 373, York YO10 5YW, United Kingdom.

<sup>2</sup> Forestry Molecular Biology Co-operative Programme (FMBC), Forestry and Agricultural Biotechnology Institute (FABI), Department of Genetics, University of Pretoria, Pretoria, 0002, South Africa.

\*Corresponding author: E.T. Steenkamp, Phone: +44-1904 328633, e-mail: es21@york.ac.uk

Received October 21, 2002; accepted October 23, 2003

### ABSTRACT

In South Africa, *Eucalyptus grandis* has been used extensively in commercial plantations. *E. grandis*, together with two other important South African forestry species (*E. saligna* and *E. urophylla*) form part of the *Eucalyptus* section *Transversaria*. Identification of these species and their hybrids is complicated by the presence of overlapping morphological features. Application of molecular diagnostics would, therefore, greatly simplify routine verification of species and hybrid identity. Our aim was to assess the value of the 5S ribosomal DNA repeat, for developing a PCR-based diagnostic method for differentiating *E. urophylla*, *E. saligna* and *E. grandis*. For this purpose we analysed the 5S repeat from *E. urophylla*, *E. saligna* and *E. grandis*, grown from seed that was collected from native trees known to represent these species in Australia and Indonesia. Three commercially grown *E. grandis* individuals were also included. Sequence analysis showed that the 5S spacer is sufficiently variable for distinguishing *E. grandis*, *E. urophylla* and *E. saligna* from each other. It also revealed that two of the three commercially grown *E. grandis* in South Africa are hybrids. Despite the utility of this approach the variable sites in the *Eucalyptus* 5S repeat appear to be unsuitable for constructing species-specific PCR-primers.

**Key words:** *Eucalyptus* hybrids, *E. grandis*, *E. urophylla*, 5S rRNA, diagnostic PCR

### INTRODUCTION

Species of *Eucalyptus* L'Hérot. are extensively grown in commercial forest plantations, particularly in the tropics and southern hemisphere (ELDRIDGE *et al.* 1994). The genus *Eucalyptus* includes more than 500 different species that differ markedly in their value for forestry (PRYOR & JOHNSON 1971; CHIPPENDALE 1988; ELDRIDGE *et al.* 1994). The majority of the more commonly utilised plantation species belong to the sub-genus *Symphyomyrtus* (PRYOR & JOHNSON 1971; CHIPPENDALE 1988; ELDRIDGE *et al.* 1994). Nearly all economically important species are included in the *Symphyomyrtus* sections, *Transversaria*, *Exsertaria* and *Maidenaria* (PRYOR & JOHNSON 1971, ELDRIDGE 1994). *E. grandis* W. Hill: Maiden in the section *Transversaria* is the most widely used species in plantations (BURGESS 1988, ANONYMOUS 1990, ELDRIDGE *et al.* 1994). In addition to *E. grandis*, the section *Transversaria* also includes other commonly utilized species, e.g. *E. urophylla* S.T. Blake and *E. saligna* Smith (VAN WYK 1985; GRIFFIN *et al.* 1988; ELDRIDGE 1994).

*Eucalyptus* hybrids, especially those with *E. gran-*

*dis*, *E. saligna* and *E. urophylla* in their genetic backgrounds, are widely used in South African plantations (ELDRIDGE *et al.* 1994). All three species confer a variety of valuable traits (e.g. accelerated growth rates, cold tolerance, drought tolerance, etc.) to their hybrid progeny. Many of these hybrids were naturally generated during the early developmental stages of plantation forestry, when various *Eucalyptus* species were planted in combination. During this period, several generations of open pollination and natural selection produced *Eucalyptus* populations (land races) that were better suited to the South African environments, in which they have been planted. Within these populations, seed was then harvested from trees displaying favourable characters. As a result, many of the commercially used South African *Eucalyptus* land races are believed to be natural hybrids (ELDRIDGE *et al.* 1994).

Most of the South African *E. grandis* plantations are thought to represent land races that are natural *E. saligna* × *E. grandis* hybrids, although this has not been proven experimentally. Currently, a wide variety of such natural hybrids, as well as pure *Eucalyptus* species, are used in breeding programmes involving other hybrids. From the management and developmental

perspective, it would be useful and important to know the genetic background of the trees used in these programmes. Knowledge of whether breeding and planting stock are true to type, is equally important.

For routine identifications of *Eucalyptus* species and hybrids, taxonomic keys such as the one proposed by CHIPPENDALE (1988) are generally used. However, the morphology-based identification systems generally provide unsatisfactory diagnoses for *E. grandis*, and closely related species such as *E. saligna* and *E. urophylla*. This is because these species have many overlapping and variable morphological features (CARR & CARR 1959; PRYOR & JOHNSON 1971; CHIPPENDALE 1988; GUNN & McDONALD 1991; ELDRIDGE *et al.* 1994). Application of the morphology-based identification systems on hybrids of these species is even more problematic and results in highly ambiguous diagnoses (JOHNSON 1976; GRIFFIN *et al.* 1988). It is, therefore, virtually impossible to accurately identify eucalypts like *E. grandis*, *E. saligna* and *E. urophylla* or any of their hybrids using morphology alone.

An alternative approach for obtaining more robust and unambiguous identifications of the genetic background of breeding and planting stock is to use DNA-based methods. Three regions have thus far been targeted for studying the classification and phylogeny of eucalypts. These include the 5S ribosomal RNA (rRNA) gene and spacer (UDOVICIC *et al.* 1995); the internal transcribed spacers (ITS) associated with the 26S, 5.8S and 18S rRNA genes (STEANE *et al.* 1998); and regions located on the chloroplast genome (MCKINNON *et al.* 2001). Analyses of eucalypt chloroplast genes have revealed that they would not be appropriate for identifying hybrids. This is because the chloroplasts are inherited maternally (BYRNE *et al.* 1993) and also because chloroplast and species evolution are possibly discordant (STEANE *et al.* 1998; MCKINNON *et al.* 1999). Analyses of both the ITS and 5S regions in *Eucalyptus* has shown that they are potentially suitable for resolving some of the relationships among groups of *Eucalyptus* (UDOVICIC *et al.* 1995; STEANE *et al.* 1998), although neither study addressed the phylogeny of the section *Transversaria*.

The 5S rRNA, together with several ribosomal proteins and the 26S and 5.8S rRNA molecules, forms part of the large ribosomal subunit (GARRETT *et al.* 1981; NOLLER 1984; BRIMACOMBE & STIEGE 1985; GÖRINGER & WAGNER 1988). In plants, the 5S genes are located in arrays of tandem repeats that are separate from the 26S–5.8S–18S-repeating units (LONG & DAWID 1980; SASTRI *et al.* 1992). Each 5S repeat consists of the 5S gene [~ 120 base pairs (bp) in length] (LONG & DAWID 1980), which contains few phylogenetically informative sites, and a non-transcribed

intergenic spacer region, which is phylogenetically more useful (SCOLES *et al.* 1988; STEELE *et al.* 1991; SASTRI *et al.* 1992; UDOVICIC *et al.* 1995; CRONN *et al.* 1996; BAKER *et al.* 2000). Furthermore, each repeating unit within an individual is not necessarily identical. This is because the degree of similarity between the various repeats depends on a process referred to as 'gene conversion', whereby nonreciprocal recombinational events homogenise multiple copies of a gene at a specific locus (GRAUR & LI 2000).

Our goal was to determine whether the 5S region is sufficiently variable to allow development of a DNA-based diagnostic procedure, to be used in commercial *Eucalyptus* breeding programmes. For this purpose we aimed to determine the sequences for the 5S repeat in three commonly used commercial *E. grandis* individuals in South Africa, as well as for *E. urophylla*, *E. saligna* and *E. grandis*, that were collected in natural stands and certified as true to type. Secondly, we wished to determine whether the 5S repeat is sufficiently variable to allow differentiation of the various eucalypts. Our final objective was assess whether the 5S repeat is suitable for constructing diagnostic PCR primer sets for differentiating *E. urophylla*, *E. saligna* and *E. grandis*.

## MATERIALS AND METHODS

### Plant material and DNA extraction

Three commercially grown *E. grandis* trees (designated as A, B and C) were randomly selected for this study. All three trees were thought to represent pure *E. grandis*. Typical of *E. grandis* grown in South Africa, trees A, B and C were superior individuals selected from open pollinated seedling stands. A representative of each of *E. urophylla*, *E. saligna* and *E. grandis* were also included. Each of these representatives was grown from seed, which was certified as true to type and has been collected from natural stands in Australia or Indonesia. For outgroup purposes, commonly grown South African *Eucalyptus* representatives from the section *Maidenaria* were used. These included *E. nitens*, *E. dunnii* and *E. smithii* for which seeds had originated in native stands in Australia, and are believed to be true to type. For testing the primer sets generated in this study, 60 progeny of a cross between individuals A and C were included. All the *Eucalyptus* trees used in this study have been planted in South African plantations and representatives are maintained by the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria. DNA was isolated from young leaves using the method described by

MURRAY and THOMPSON (1980).

### Amplification, cloning and sequencing of 5S rDNA repeats

The 5S gene and spacer regions from each individual were amplified using primers 5SR and 5SFUL (UDOVICIC *et al.* 1995). Each PCR reaction contained 1 mM dNTPs (0.25 mM of each), 2.5 mM MgCl<sub>2</sub>, 0.2 μM primer, 0.25 ng·μl<sup>-1</sup> DNA, 0.05 U·μl<sup>-1</sup> Super-Therm DNA polymerase and 1× reaction buffer [Southern Cross Biotechnology (Pty) Ltd, Cape Town, South Africa]. Reactions were performed on a GeneAmp® PCR System 9700 (Perkin Elmer, Warrington, U.K.). The cycling conditions were as follows: an initial denaturation step of 5 min at 92 °C, followed by 30 cycles of 10 sec at 92 °C, 10 sec at 60 °C, 10 sec at 72 °C and a final extension at 72 °C for 12 min.

PCR products were visualized by agarose gel electrophoresis, and the ~400 bp fragment was excised from the gel and purified using a gel extraction kit (QIAquick Gel Extraction Kit, Qiagen GmbH, Germany). The purified products were cloned using the pGEM-T Easy Vector System (Promega Corporation, Madison, Wisconsin, USA) and plasmids harvested with alkaline lysis (SAMBROOK *et al.* 1989). Due to the fact that the 5S repeat occurs as multiple copies, at least 4 randomly selected clones from each plant were sequenced using plasmid-specific primers T7 (5'-TAA TAC GAC TCA CTA TAG GG-3') and SP6 (5'-TAT TTA GGT GAC ACT ATA G-3'). Sequencing reactions were performed on an ABI PRISM™ 377 automated DNA sequencer with an ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer).

### Sequence analyses

Nucleotide sequences were analysed with Sequence Navigator version 1.0.1 (Applied Biosystems) and manually aligned by inserting gaps. The degree of variation in the 5S gene and spacer regions within each individual was quantified using nucleotide diversity. Nucleotide diversity, which is a measure of the average number of differences per nucleotide between all pairs of sequences obtained from a single individual, was calculated according to NEI (1987). The degree of variation between species was estimated from pair-wise distances that were calculated from the mean character differences with PAUP (Phylogenetic Analysis Using Parsimony, version 4.0b; Sinauer Associates, Sunderland, Mass., SWOFFORD 1998).

Aligned sequences were subjected to phylogenetic analyses using the parsimony and distance methods included in PAUP. Trees were generated from the complete data set (gene and spacer sequences), as well as from separate subsets, including either the gene or spacer sequences. In the parsimony analyses, gaps were treated as fifth characters in heuristic searches with tree-bisection-reconnection branch swapping. The distance analyses, using the neighbour joining algorithm, were based on 'uncorrected distances' (uncorrected "p"), where gaps were treated as missing characters. In both cases bootstrap analyses based on 1000 replications, were performed to estimate the confidence of branches. The possible secondary structure of the rRNA molecule encoded by the sequenced 5S gene was inferred with the internet-based programme *mfold* (<http://bioinfo.math.rpi.edu/~mfold/rna/form1.cgi>), which is based on the Zuker algorithm (ZUKER *et al.* 1999).

### Specific primers

The sequences obtained in this study were evaluated for the presence of unique differences that would allow construction of diagnostic primers. Only one of the three eucalypts from the section *Transversaria*, *E. urophylla*, contained such unique nucleotide positions. The *E. urophylla*-specific 5S repeat primers were EU-f (5'-AAG CCG GAA GTA GAA CTC TG-3') and EU-r (5'-GAT CGA AAT CTC GAC CCG GG-3'). Interestingly, one of the outgroup species, *E. nitens*, also harboured a number of unique 5S positions that were potentially useful for diagnostic PCR. An *E. nitens*-specific primer [EN-r (5'-ATG TCC GGA GAG TTC TAG TC-3')] was, therefore, designed. This primer is used in combination with 5SR, which was designed by UDOVICIC *et al.* (1995). The PCR reaction and cycling conditions with both primer sets were similar to those for amplifying the 5S gene and spacer regions described above, except that an annealing temperature of 92 °C was used. All PCR reactions were repeated twice and the amplified products were electrophoresed on 2 % agarose gels (3 V·cm<sup>-1</sup>; room temperature) and visualized with a UV-transilluminator.

Both the *E. nitens* and *E. urophylla* primer sets were tested on all the *Eucalyptus* individuals included in this study. The *E. urophylla* primer set was further tested on 60 progeny of the cross between trees A and C, thought to represent pure *E. grandis*. Segregation of the *E. urophylla*-fragment in the progeny of the cross between A and C was evaluated using the chi-square test for goodness-of-fit to a 1:1 ratio.

RESULTS

Amplification, sequencing and alignment of 5S repeat

PCR with the primers 5SR and 5SFUL amplified an array of products, with the smallest fragment approximately ~400 bp in length. Electrophoretic separation of these products generated a 'ladder' of fragments that were multimers of the ~400 bp product. Since these ~400 bp fragments represent the 5S gene and spacer repeating units (UDOVICIC *et al.* 1995), they were excised from the gels, cloned and sequenced. The sequences for each individual studied, ranged from 406 to 411 nucleotides in length. There was no correlation

between the plants used and the length of the sequence.

For this study, 5S repeat sequences were reported only if they were recovered at least twice from a single individual. Because we could not discount the fact that sequences recovered only once may be the result of PCR or sequencing errors, these sequences were not used even though they may possibly have formed part of the 'pool' of 5S repeats within an individual. Alignment of the representative sequences from each of the nine different plants analysed, was relatively simple, requiring the insertion of alignment gaps at only nine positions (Fig. 1). The first 95 bp and the last 23 bp of these sequences represented the 5S gene and no gaps were required for the alignment. All nine of the alignment gaps were situated in the spacer region.

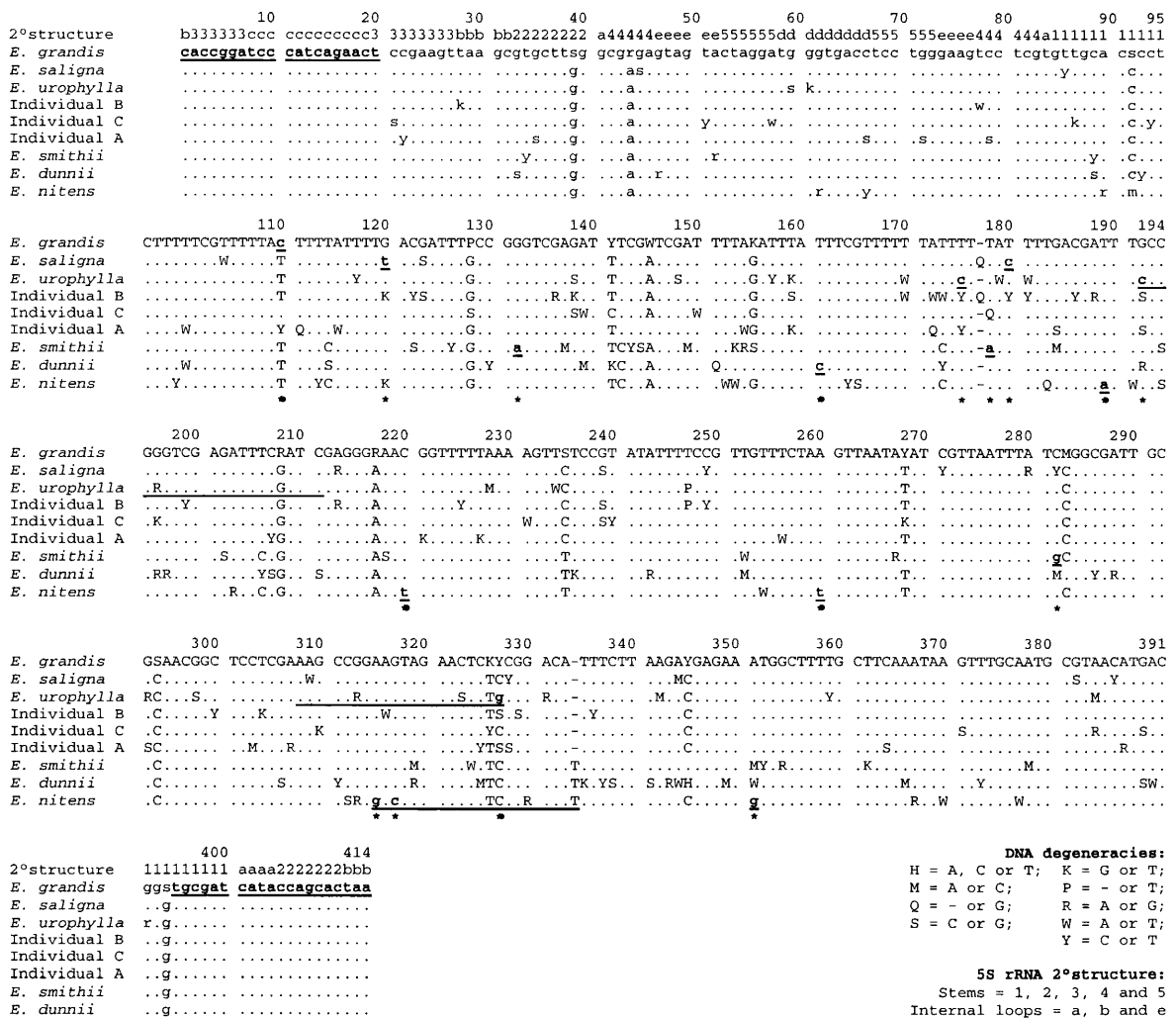


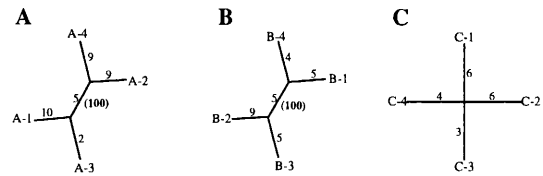
Figure 1. An alignment of the consensus sequences for the 5S gene and spacer regions of the nine *Eucalyptus* individuals studied. Spacer regions are indicated in capital letters, genes are indicated in lower case letters, and alignment gaps are indicated with -. Individual-specific fixed polymorphisms (bold and underlined lower case letters) are indicated with \*. All primer sites are underlined and those for 5SR and 5SFUL (UDOVICIC *et al.* 1995) are indicated in bold. The loop and stem regions of the probable secondary structure [inferred with *mfold* (<http://bioinfo.math.rpi.edu/~mfold/rna/form1.cgi>)] of the 5S gene are indicated according to BARCISZEWSKA *et al.* (1996), where our stems 1 to 5 corresponds to their stems A to E.

Of the 414 aligned nucleotide positions, 42.2 % were variable. Of the 296 aligned nucleotide positions in the spacer region 50.8 % (150) were variable, while 24.6 % (29 of 118) of the nucleotide positions in the genes were variable. Analysis of the inferred secondary structure, further indicated that most of the variable sites in the genes (22 nucleotides) were located in the stems and the remaining seven polymorphic nucleotides, were located in loops E and D (Fig. 1).

The three distinct domains of the 5S spacers are the 5'-upstream region, 3'-downstream region and the mid-spacer region (APPELS *et al.* 1992; SCOLES *et al.* 1988; SASTRI *et al.* 1992). The *Eucalyptus* 3'-downstream region that signals termination of transcription for RNA polymerase III is characterised by the stretch of Ts (nucleotides 97 to 119) immediately after the gene (Fig. 1). Additional AT-rich regions, which are thought to assist in termination of transcription in cases where transcripts read through the first terminator, are located at nucleotide positions 157–163, 166–183, 240–247 and 262–270 (Fig. 1). The 5'-upstream domain consists of the 60–90 nucleotides upstream from the 5'-end of the gene. The *Eucalyptus* 5S spacer mid-spacer region corresponds to positions 120–334.

**Within- and between-individual variation**

To determine the extent to which the 5S repeats of an individual evolve in concert, *i.e.* gene conversion, we analysed and compared within-individual and between-individual variation. Sequence analyses revealed that each of the reported 5S repeats analysed per plant, were



**Figure 2.** Dendograms generated from the sequences for the four 5S repeats analysed for individual A (A), individual B (B) and individual C (C). In all three cases confidence of internal branches was estimated with 1,000 bootstrap replications. Branch lengths are represented in single digits and bootstrap values are indicated in parentheses.

unique. The within-individual sequence variations were preferentially located in the spacer regions, which harboured more polymorphic nucleotides than the 5S genes (Table 1, Fig. 1). However, quantification of the variation per nucleotide between the gene and spacer regions within each individual, using nucleotide diversity as a parameter, revealed that the diversity in the spacers, was generally not significantly higher than that for the genes (Table 1). The only two exceptions were *E. dunnii* and *E. grandis* individual B, where the nucleotide diversity in the genes was significantly less than the diversity in the spacer (Table 1).

Cluster analysis of the four sequences obtained from each individual, indicated that the 5S repeats in each plant could not be separated into more than one group. The only exceptions were *E. grandis* individuals A and B. These plants appeared to harbour two types of 5S

**Table 1.** The number of polymorphic nucleotides and nucleotide diversities in the gene and spacer regions of the eucalypts analysed.

| <i>Eucalyptus</i> individual | Number of polymorphic nucleotides |        | Nucleotide diversity <sup>2</sup> |         |        |                      |
|------------------------------|-----------------------------------|--------|-----------------------------------|---------|--------|----------------------|
|                              | Gene <sup>1</sup>                 | Spacer | Gene                              |         | Spacer |                      |
| <i>E. grandis</i>            | 4                                 | 13     | 0.017                             | ± 0.008 | 0.022  | ± 0.004              |
| <i>E. saligna</i>            | 2                                 | 10     | 0.009                             | ± 0.009 | 0.026  | ± 0.012              |
| <i>E. urophylla</i>          | 2                                 | 17     | 0.015                             | ± 0.010 | 0.039  | ± 0.014              |
| <i>E. nitens</i>             | 4                                 | 14     | 0.018                             | ± 0.008 | 0.032  | ± 0.012              |
| <i>E. dunnii</i>             | 4                                 | 37     | 0.018                             | ± 0.008 | 0.066  | ± 0.010 <sup>3</sup> |
| <i>E. smithii</i>            | 3                                 | 23     | 0.013                             | ± 0.009 | 0.042  | ± 0.011              |
| A                            | 5                                 | 18     | 0.021                             | ± 0.009 | 0.042  | ± 0.016              |
| B                            | 2                                 | 23     | 0.008                             | ± 0.005 | 0.047  | ± 0.013 <sup>3</sup> |
| C                            | 5                                 | 13     | 0.021                             | ± 0.014 | 0.027  | ± 0.005              |

<sup>1</sup>) The number and position of possible polymorphic nucleotides in the 40 positions representing the 5SR and 5SFUL primer sites (Fig. 1; positions 1–20 and 394–414) is not known.

<sup>2</sup>) Nucleotide diversity and variance were determined according to NEI (1987). Nucleotide diversity of genes and spacers are significantly different at *P* < 0.05.

repeat (Fig. 2 A and B). The third *E. grandis* test tree, individual C, harboured a single 5S repeat type (Fig. 2C). Sequence information from one of the clones in each of the clusters was selected for subsequent parsimony and distance analyses. These were clones 2 and 3 for individual A, clones 1 and 3 for individual B and clone 1 for individual C.

Between-individual variation was estimated from the pair-wise distances determined with PAUP (Table 2). Since the genes harboured almost no informative characters, between-individual distances were based on the mean character differences in the spacer sequences. The latter values are numerically equivalent to the nucleotide diversities that were used to quantify the within-individual variation in the spacer sequences (Table 1). The values for the between-individual differences were generally lower than, but sometimes equal to, the values estimating within-individual diversity (Table 2). The average differences between *E. grandis* individual A and pure *E. grandis* and *E. urophylla* were zero, indicating that some of the 5S repeats are more similar to those for *E. grandis*, while others are more similar to those in *E. urophylla*. The average differences between *E. grandis* individual B and *E. saligna* and *E. urophylla* were also zero, indicating that some of the 5S repeats in B were more similar to those of *E. saligna*, while others were more similar to those of *E. urophylla*. The average differences between *E. grandis* individual C and pure *E. grandis* were zero, indicating that all of the 5S repeats in C are similar to those in *E. grandis*.

### Phylogenetic analysis

In order to determine the relationships between the

selected sequences for *E. grandis* individuals A, B and C and the pure *E. grandis*, *E. saligna* and *E. urophylla*, distance- and parsimony-based analyses were used. In these analyses, *E. nitens*, *E. dunnii* and *E. smithii* were used as outgroup taxa. Since similar results were generated with the distance- and parsimony-based methods, only those obtained for the parsimony-based analyses are reported. Because the 5S gene harboured only two parsimony-informative characters, whereas the spacer contained 53 informative characters, both regions were used to generate a single most parsimonious tree with PAUP (Fig. 3). This tree was 209 steps in length, with consistency (CI) and homoplasy (HI) indexes of 0.9187 and 0.0813, respectively. The tree was also congruent with that inferred from the spacer sequences alone (results not shown).

All of the sequences from an individual plant grouped together and separate from those of other individuals. This was true for all but two of the plants studied. These exceptions were *E. grandis* individuals A and B, which apparently obtained their 5S repeats from parents that were not conspecific. One of the sequences from both individuals A and B grouped with those from *E. urophylla*. The second sequence from individual A clustered with *E. grandis*, whereas the second sequence from individual B clustered with *E. saligna*. Based on the 5S sequence data, *E. grandis* individual C grouped together with pure *E. grandis*.

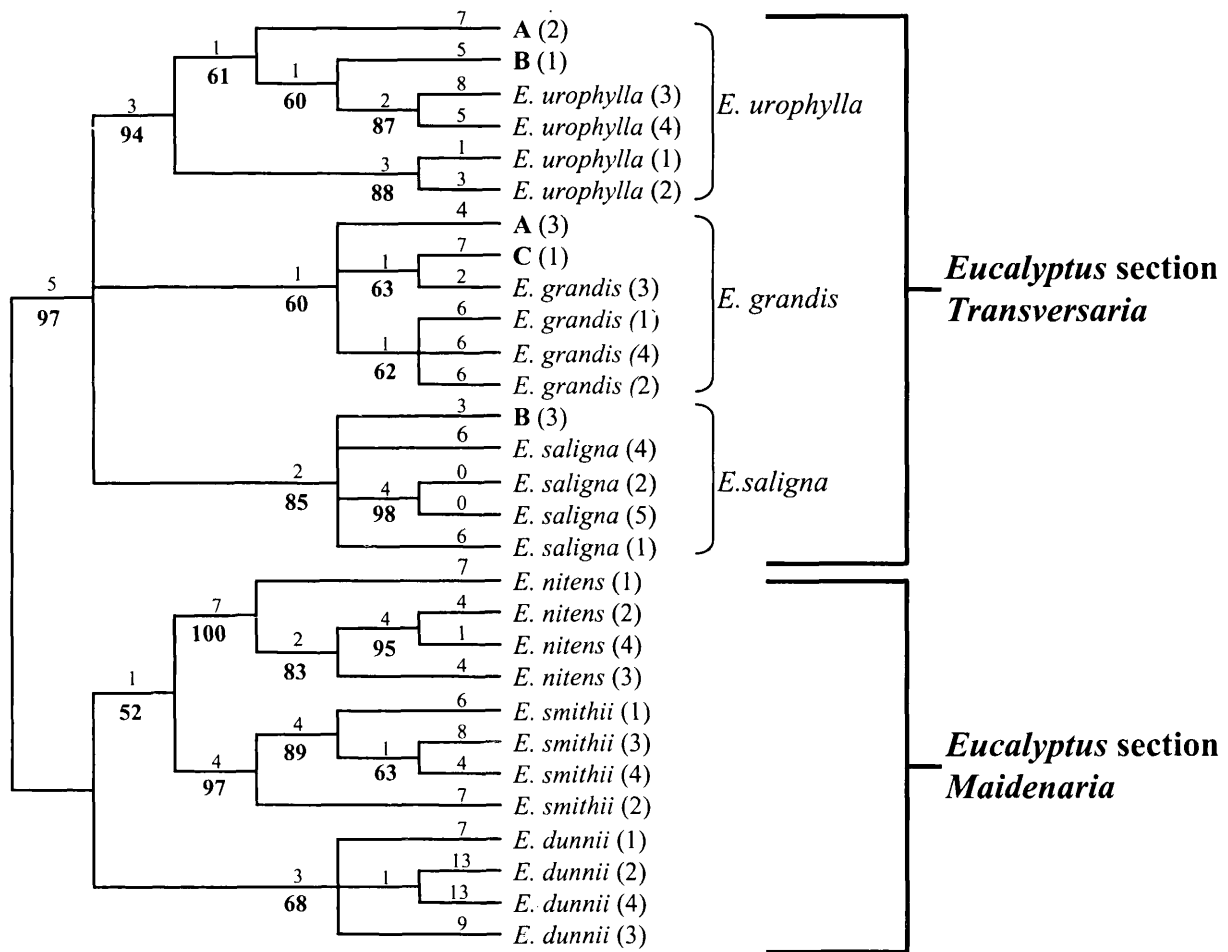
### Specific PCR

From the sequences obtained in this study, it was possible to construct specific primers for the *E. urophylla* and *E. nitens* individuals. The sequence variation

**Table 2.** Average differences among the 5S-spacer sequences for the nine different *Eucalyptus* species and individuals studied.

| <i>Eucalyptus</i><br>individual | Average difference <sup>1</sup> |                   |                     |                  |                  |                   |       |       |   |
|---------------------------------|---------------------------------|-------------------|---------------------|------------------|------------------|-------------------|-------|-------|---|
|                                 | <i>E. grandis</i>               | <i>E. saligna</i> | <i>E. urophylla</i> | <i>E. nitens</i> | <i>E. dunnii</i> | <i>E. smithii</i> | A     | B     | C |
| <i>E. grandis</i>               | –                               |                   |                     |                  |                  |                   |       |       |   |
| <i>E. saligna</i>               | 0.010                           | –                 |                     |                  |                  |                   |       |       |   |
| <i>E. urophylla</i>             | 0.014                           | 0.017             | –                   |                  |                  |                   |       |       |   |
| <i>E. nitens</i>                | 0.044                           | 0.044             | 0.051               | –                |                  |                   |       |       |   |
| <i>E. dunnii</i>                | 0.027                           | 0.030             | 0.034               | 0.030            | –                |                   |       |       |   |
| <i>E. smithii</i>               | 0.037                           | 0.041             | 0.044               | 0.034            | 0.020            | –                 |       |       |   |
| A                               | 0.000                           | 0.007             | 0.000               | 0.041            | 0.024            | 0.030             | –     |       |   |
| B                               | 0.003                           | 0.000             | 0.000               | 0.040            | 0.024            | 0.033             | 0.000 | –     |   |
| C                               | 0.000                           | 0.014             | 0.017               | 0.047            | 0.030            | 0.041             | 0.003 | 0.007 | – |

<sup>1</sup>Average differences were determined with PAUP (SWOFFORD 1998), using the consensus sequences for each of the individuals studied. Consensus sequences were generated with MacClade (MADDISON & MADDISON 1992).



**Figure 3.** A single most parsimonious tree generated from the combined 5S sequences for the gene and the spacer region studied in each of the *Eucalyptus* species included, as well as the three test trees. The number of the specific clone sequenced for each individual is indicated in parenthesis alongside the taxon name. The test trees A, B and C are indicated in bold. This tree is rooted to the species of the section *Maidenaria*. Branch lengths are indicated above the branches, while bootstrap values based on 1,000 replications are indicated in bold digits below the internodes.

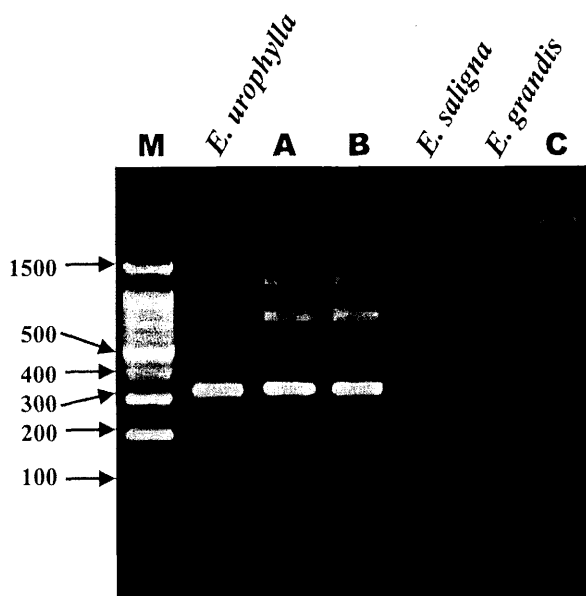
in other species did not allow the design of such primers. Use of the *E. nitens* primer and 5SR amplified a ~330 bp product, only from the *E. nitens* individual. The *E. urophylla*-specific primer set amplified a ~350 bp fragment from presumed *E. grandis* individuals A and C, as well as the *E. urophylla* representative (Fig. 4). Application of these primers to the 60 progeny of a controlled cross between individuals A and C resulted in amplification of this fragment from 33 individuals. This fragment was absent from the remaining 27 progeny, which approximates a 1:1 segregation ( $P > 0.05$ ).

**DISCUSSION**

The primary objective of this study was to assess the value of the 5S rRNA gene and spacer sequence for

molecular diagnostics in common, commercially-grown *Eucalyptus* species. For this purpose, we specifically analysed the 5S repeat from *Eucalyptus* species that have been used in Southern African plantations. Our results revealed that the 5S spacer region appears to contain sufficient informative sites to allow phylogenetic separation of a small set of closely related *Eucalyptus* species. Although these sequences appear to be marginally useful for designing specific PCR-primers, this region is generally not adequately variable for developing PCR-based diagnostic methods. Our study further showed that the current morphology-based classification used in commercial *Eucalyptus* forestry is not adequately robust for diagnosis of either ‘pure’ species or hybrids.

Sequence analysis of the *Eucalyptus* 5S repeat revealed that two of the three commercially grown eucalypts used in this study, were not *E. grandis*, as



**Figure 4.** PCR products generated with the primers EU-r and EU-f specific to the *E. urophylla* individual after electrophoretic separation on an agarose gel. The *E. urophylla*-specific PCR product was present in *E. urophylla*, individual A (lane A) and individual B (lane B). This fragment could not be amplified from *E. saligna*, *E. grandis* and individual C. Lane M contains the molecular weight marker with sizes indicated in base pairs.

initially supposed. Only one (individual C) of the three randomly selected South African *E. grandis* trees appears to represent this species and not a hybrid (see for example Figs. 3 and 4). *E. grandis* individual A apparently represents an interspecific hybrid between *E. urophylla* and *E. grandis*, because it harbours 5S repeats typical of both species. Surprisingly, individual B appeared to lack *E. grandis*-specific 5S repeats entirely. This plant apparently arose from an interspecific hybridisation event between *E. urophylla* and *E. saligna*. Our results thus confirm the suggested but untested hypothesis that commonly grown *E. grandis* trees in South Africa, represent local land races, which have emerged through natural hybridisation between commonly grown species.

For studying *Eucalyptus* hybrids, previous researchers have used morphological, biochemical and isozyme characters (e.g. JOHNSON 1976; WHIFFIN 1981; COOKE & LADIGES 1991), as well as RAPDs (random amplified polymorphic DNAs) (SALE *et al.* 1996) and RFLPs (restriction fragment length polymorphisms) (e.g. TYSON *et al.* 1998; JACKSON *et al.* 1999). To the best of our knowledge, this study is the first to demonstrate the hybrid nature of *Eucalyptus* trees using nuclear DNA sequence data. Our results further illustrate the potential of molecular tools such as DNA

sequencing, for identification and characterisation of the genetic background of *Eucalyptus* trees used in breeding programmes. Because hybridisation has become a major strategy for the improvement of *Eucalyptus*, such relatively rapid techniques will make it possible to identify the genetic background of outstanding individuals resulting from open pollinated forests.

From a long-term research standpoint we are interested in developing molecular diagnostic procedures for characterising and identifying commercially grown *Eucalyptus* genotypes in South Africa. Prior to beginning full-scale population studies of these trees, which would be a prerequisite for developing such diagnostic methods, we wished to consider whether the 5S repeat would represent an adequate tool. For this purpose we used a relatively small set of commercially developed eucalypts and examined various aspects of their 5S repeats, which included phylogenetic relatedness, within- and between-individual variation, the incidence of individual- and species-specific polymorphisms, the occurrence of polymorphic sites suitable for designing specific PCR-primers, etc. Our findings, however, suggest that the 5S repeat is not suitable for *Eucalyptus* diagnostics, mainly because it lacks sites fit for primer design and because it is inordinately variable within individuals. In the following sections these results are discussed more fully.

Phylogenetic analyses suggest that the repeats within the various eucalypts generally appeared to be more closely related to one another than to those of other individuals, with hybrid individuals A and B as the only exceptions (Fig. 3). Unfortunately, none of the South African trees A, B and C turned out to represent 'pure' *E. grandis*, and it is thus impossible to speculate on the degree of intra-specific variation. Our results do, however, suggest that the repeats within an individual have arisen from an ancestor, after the speciation event and that the observed within-individual variation is the result of incomplete gene conversion. Even though repeats within an individual are evolutionarily more closely related to each other than to those of other species (Fig. 3), repeats within an individual are more different to each other than to those of other species. This implies that within-individual diversity exceeds that of between-individual diversity (Tables 1 and 2). For example, the nucleotide diversity in the *E. grandis* 5S spacer (0.022) is much higher than that expected from comparisons with *E. saligna* (0.01) and *E. urophylla* (0.014).

The *Eucalyptus* 5S polymorphic nucleotides appeared to be distributed equally across the gene and the spacer. This was evident from the fact that the nucleotide diversities in these regions were not significantly different (Table 1). There were, however, definite



differences between the gene and the spacer, regarding the extent to which polymorphisms were fixed within an individual. The 5S genes of the *Eucalyptus* species studied, were devoid of individual-specific fixations, whereas the spacers harboured 16 fixed within-individual polymorphisms (Fig. 1). This contradicts the neutral expectation that fixed polymorphisms should occur randomly across both the *Eucalyptus* 5S gene and the spacer. The non-random distribution of fixed polymorphisms is also reflected by the fact that almost all of the parsimony-informative characters were situated in the spacer regions.

Similar trends of unequal distribution of fixed polymorphisms, as well as parsimony informative characters in the 5S gene and spacer have been observed in other plants (MCINTYRE *et al.* 1992; PLAYFORD *et al.* 1992; CRONN *et al.* 1996). In these plants, polymorphism and divergence of the 5S spacers also appeared to be correlated, whereas polymorphism and divergence were decoupled in the genes (KELLOGG & APPELS 1995). The non-random distribution of fixed individual-specific polymorphisms, therefore, suggests that different forces are involved in the evolution of the 5S genes and spacers of *Eucalyptus* and other plants. These dissimilar evolutionary forces and the fact that within-individual diversity generally outweighs between-individual diversity, could potentially also reduce its phylogenetic value. This is suggested in the topology of the single most parsimonious tree that was generated from the *Eucalyptus* 5S sequences (Fig. 3). Higher within-individual diversities gave rise to longer terminal branches, especially in the case of *E. saligna*, *E. grandis* and *E. urophylla*, whereas lower between-individual diversities gave rise to relatively short internal branches (Fig. 3). Therefore, as has been shown for other plants (see for example HALANYCH 1991; STEELE *et al.* 1991), phylogenies of large sets of closely related *Eucalyptus* species using this region, might not be reliable. Including a greater number of samples will potentially increase the number of homoplasies and reduce the resolution even further.

The inferred secondary structure for the 5S rRNA molecule encoded by the *Eucalyptus* 5S gene was similar to those characterised and inferred for other organisms (ERDMANN & WOLTERS 1986; BARCISZEWSKA *et al.* 1996). The *Eucalyptus* 5S rRNA structure consisted of the characteristic five single stranded loops (A to D) and the five helical double stranded stems (stems 1–5). Inspection of the variable nucleotides in the *Eucalyptus* 5S revealed that the majority of the polymorphic positions were located in the stems rather than in the loops. These stem substitutions can potentially disrupt or alter secondary structure, thereby reducing the number of functional rRNA molecules

produced (ROMBY *et al.* 1990; BARCISZEWSKA *et al.* 1996). However, several workers have shown that the extent to which these stem substitutions influence secondary structure, appears to be limited by the occurrence of compensating substitutions (STEELE *et al.* 1991; SCHNEEBERGER & CULLIS 1992; KELLOGG & APPELS 1995; TRONTIN *et al.* 1999). This is consistent with the commonly held view that conservation of 5S rRNA secondary structure is superior to that of the gene sequence (ERDMANN & WOLTERS 1986; ROMBY *et al.* 1990). Thus, a multiplicity of substitutions in the gene sequence will be allowed, as long as the 5S rRNA structure is retained.

One of the aims of this study was to construct PCR primers that would selectively recognise particular species of *Eucalyptus*. For this purpose we examined the individual-specific fixed polymorphisms within the 5S spacer region. *E. grandis* harboured one, *E. saligna* two and *E. urophylla* harboured three, such polymorphisms (Fig. 1). Of these, only two of the *E. urophylla*-specific positions (192 and 327), were appropriate for use as diagnostic bases in a primer set. With the *E. urophylla* primer set, we were thus able to differentiate this individual and its hybrids from the other *Eucalyptus* species included in this study. This was because it amplified the expected fragment only in the *E. urophylla* individual and the hybrid individuals A and B (Fig. 4). Whether this primer set will be of value as a routinely used diagnostic tool for identifying *E. urophylla* in the genetic background of commercial forestry *Eucalyptus*, remains to be tested.

The diagnostic PCR-primers were also used to assess whether our results did not arise from a 'sampling artefact'. As mentioned earlier the 5S gene and spacer region occurs as multiple repeats in the genome. Since we have sequenced only a limited number of repeats in each individual, it might have been possible that we sampled by chance the various 5S 'types' specific to each of the plants. This would be especially true if alternative 5S repeat 'types', occur in very low copy numbers. However, these low copy repeats, should they exist, would readily be detected with PCR. Using the *E. urophylla* and *E. nitens*-specific primers, we showed that the *E. nitens*-repeat was present only in the *E. nitens* individual studied (Fig. 4). The *E. urophylla*-repeat was present only in *E. urophylla* and the *E. grandis* trees shown to be hybrids. We are, therefore, relatively confident that the findings reported in this study are not based on 'under sampling' of the 5S repeat-pool in the *Eucalyptus* genome.

The 5S repeats of *E. urophylla* and possibly the other *Eucalyptus* species, appear to be located at a single locus. This is because the *E. urophylla*-specific PCR product segregated in a Mendelian fashion, in the

progeny of a cross between *E. grandis* (individual C) and the *E. grandis* now known to be an *E. grandis* × *E. urophylla* hybrid (individual A). The expected ~330 bp fragment was present in 55 % of the 60 offspring, while it was absent from the remaining 45 %, which is equivalent to a 1:1 segregation of alleles at a single locus. If the 5S rRNAs of all *Eucalyptus* species turn out to be located on a single locus, they will share this feature with plants such as Douglas fir, tomato and sugar beet (LAPITAN *et al.* 1991; SCHONDELMAIER *et al.* 1997; AMARASINGHE & CARLSON 1998). In contrast, *Pinus*, *Triticum* and *Acacia* are examples of plants with multiple 5S loci (DVORÁK *et al.* 1989; MORAN *et al.* 1992; PLAYFORD *et al.* 1992).

The results of this study suggest that the 5S repeat will have limited value as diagnostic and phylogenetic marker for South African *Eucalyptus* plantation forestry. While 5S sequences can be used to identify and characterize important planting and breeding stock, analyses of large sets of samples may be extremely challenging. The sequences determined in this study have resulted in the construction of two putative diagnostic PCR primer sets. These could be of some use in routine practices such as determining seed sources, verification of species identity and certifying forestry products such as seed and wood for pulping. However, alternative nuclear regions, which may prove to be more useful for *Eucalyptus* diagnostics should be examined in future.

## REFERENCES

- AMARASINGHE, V. & CARLSON, J.E. 1998: Physical mapping and characterization of 5S rRNA genes in Douglas-fir. *J. Hered.* **89**: 495–500.
- ANONYMOUS 1990: Commercial timber resources and roundwood processing in South Africa 1988–1989. Forestry Branch, Department of Environmental Affairs, Pretoria.
- APPELS, R., BAUM, B.R. & CLARKE, B.C. 1992: The 5S DNA unit of bread wheat (*Triticum aestivum*). *Plant Syst. Evol.* **183**: 183–194.
- BAKER, W.J., HEDDERSON, T.A. & DRANSFIELD, J. 2000: Molecular phylogenetics of *Calamus* (Palmea) and related rattan genera based on 5S nrDNA spacer sequence data. *Mol. Phylog. Evol.* **14**: 218–231.
- BARCISZEWSKA, M.Z., ERDMAN, V.A. & BARCISZEWSKI, J. 1996: Ribosomal 5S RNA: Tertiary structure and interactions with proteins. *Biol. Rev.* **71**: 1–25.
- BRIMACOMBE, R. & STIEGE, W. 1985: Structure and function of ribosomal RNA. *Biochem. J.* **229**: 1–17.
- BURGESS, I.P. 1988: Provenance trials of *Eucalyptus grandis* and *E. saligna* in Australia. *Silvae Genet.* **37**: 221–227.
- BYRNE, M., MORAN, G.F. & TIBBITS, W.N. 1993: Restriction map and maternal inheritance of chloroplast DNA in *Eucalyptus nitens*. *J. Hered.* **84**: 218–220.
- CARR, D.J. & CARR, S.G.M. 1959: Floral morphology and the taxonomy of *Eucalyptus*. *Nature* **184**: 1549–1552.
- CHIPPENDALE, G.M. 1988: *Eucalyptus*, *Angophora* (Myrtaceae). Flora of Australia 19. Australian Government Publishing Service, Canberra, 542 pp.
- COOKE, I.O. & LADIGES, P.Y. 1991: Morphological variation within *Eucalyptus nitens* s. lat. and re-cognition of a new species, *E. denticulata*. *Austr. Syst. Bot.* **4**: 375–390.
- CRONN, R.C., ZHAO, X., PATERSON, A.H. & WENDEL, J.F. 1996: Polymorphisms and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J. Mol. Evol.* **42**: 685–705.
- DVORÁK, J., ZHANG, H.-B., KOTA, R.S. & LASSNER, M. 1989: Organization and evolution of the 5S ribosomal RNA gene family in wheat and related species. *Genome* **32**: 1003–1016.
- ELDRIDGE, K., DAVIDSON, J., HARWOOD, C. & VAN WYK, G. 1994: *Eucalypt Domestication and Breeding*. Clarendon Press, Oxford, 288 pp.
- ERDMANN, V.A. & WOLTERS, J. 1986: Collection of published 5S, 5.8S and 4.5S ribosomal RNA sequences. *Nucleic Acids Res.* **14**(suppl): r1–r59.
- GARRETT, A., DOUTHWAITE, S. & NOLLER, H.F. 1981: Structure and role of the 5S RNA-protein complexes in biosynthesis. *Trends Biochem. Sci.* **6**: 137–139.
- GÖRINGER, H.U. & WAGNER, R. 1988: 5S RNA structure and function. *Meth. Enzymol.* **164**: 721–747.
- GRAUR, D. & LI, W.-H. 2000: *Fundamentals of Molecular Evolution*. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, 481 pp.
- GRIFFIN, A.R., BURGESS, I.P. & WOLF, L. 1988: Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'Hérot. – a review. *Austr. J. Bot.* **36**: 41–66.
- GUNN, B.V. & MCDONALD, M.W. 1991: *Eucalyptus urophylla* seed collections. *Forest Genet. Res. Inf., FAO Rome* **19**: 34–37.
- HALANYCH, K.M. 1991: 5S ribosomal RNA sequences inappropriate for phylogenetic reconstruction. *Mol. Biol. Evol.* **8**: 249–253.
- JACKSON, H.D., STEANE, D.A., POTTS, B.M. & VAILLANCOURT, R.E. 1999: Chloroplast DNA evidence for reticulate evolution in *Eucalyptus* (Myrtaceae). *Mol. Ecol.* **8**: 739–751.
- JOHNSON, L.A.S. 1976: Problems of species and genera in *Eucalyptus* (Myrtaceae). *Plant Syst. Evol.* **125**: 155–167.
- KELLOGG, E.A. & APPELS, R. 1995: Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. *Genetics* **140**: 325–343.
- LAPITAN, N.L.V., GANAL, M.W. & TANKSLEY, S.D. 1991: Organization of the 5S ribosomal RNA genes in the genome of tomato. *Genome* **34**: 509–514.
- LONG, E.O. & DAWID, I.B. 1980: Repeated genes in Eukaryotes. *Ann. Rev. Biochem.* **49**: 727–764.
- MADDISON, W.P. & MADDISON, D.R. 1992: *MacClade: Analysis of Phylogeny and Character Evolution*. Version 3.0. Sinauer Associates, Sunderland, Massachusetts.
- MCINTYRE, C.L., WINBERG, B., HOUCHEINS, K., APPELS, R. & BAUM, B.R. 1992: Relationships between *Oryza* species (*Poaceae*) based on 5S DNA sequences. *Plant Syst. Evol.* **183**: 249–264.
- MCKINNON, G.E., STEANE, D.A., POTTS, B.M. & VAILLAN-

- COURT, R.E. 1999: Incongruence between chloroplast and species phylogenies in *Eucalyptus* subgenus *Monocalyptus* (Myrtaceae). *Am. J. Bot.* **86**: 1038–1046.
- McKINNON, G.E., VAILLANCOURT, R.E., JACKSON, H.D. & POTTS, B.M. 2001: Chloroplast sharing in the Tasmanian eucalypts. *Evolution* **55**: 703–711.
- MORAN, G.F., SMITH, D., BELL, J.C. & APPELS, R. 1992: The 5S rRNA genes in *Pinus radiata* and the spacer region as a probe for relationships between *Pinus* species. *Plant Syst. Evol.* **183**: 209–221.
- MURRAY, M.G. & THOMPSON, W.F. 1980: Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **8**: 4321–4325.
- NEI, M. 1987: *Molecular Evolutionary Genetics*. Columbia University Press, New York, 512 pp.
- NOLLER, H.F. 1984: Structure of ribosomal RNA. *Ann. Rev. Biochem.* **53**: 119–162.
- PLAYFORD, J., APPELS, R. & BAUM, B.R. 1992: The 5S DNA units of *Acacia* species (*Mimosaceae*). *Plant Syst. Evol.* **183**: 235–247.
- PRYOR, L.D. & JOHNSON, L.A.S. 1971: *Classification of the Eucalypts*. Australian National University Press, Canberra, 102 pp.
- ROMBY, P., BAUDIN, F., BRUNEL, C., LEAL DE STEVENSON, I., WESTHOF, E., ROMANIUK, P.J., EHRESMANN, C. & EHRESMANN, B. 1990: Ribosomal 5S rRNA from *Xenopus laevis* oocytes: conformation and interaction with transcription factor IIIA. *Biochimie* **72**: 437–452.
- SALE, M.M., POTTS, B.M., WEST, A.K. & REID, J.B. 1996: Molecular differentiation within and between *Eucalyptus risdonnii*, *E. amigdalina* and their hybrids using RAPD markers. *Austr. J. Bot.* **44**: 559–569.
- SAMBROOK, J., FRITSCH, E.F. & MANIATIS, T. 1989: *Molecular Cloning. A Laboratory Manual*. Second Edition. Cold Spring Harbour Laboratory Press, pp 1.25–1.28.
- SASTRI, D.C., HILU, K., APPELS, R., LAGUDAH, E.S., PLAYFORD, J. & BAUM, B.R. 1992: An overview of evolution in plant 5S DNA. *Plant Syst. Evol.* **183**: 169–181.
- SCHONDELMAIER, J., SCHMIDT, T., HESLOP-HARRISON, J.S. & JUNG, C. 1997: Genetic and chromosomal localization of the 5S rDNA locus in sugar beet (*Beta vulgaris* L.) *Genome* **40**: 171–175.
- SCHNEEBERGER, R.G. & CULLIS, C.A. 1992: Intraspecific 5S rRNA gene variation in flax, *Linum usitatissimum* (*Lina- ceae*). *Plant Syst. Evol.* **183**: 265–280.
- SCOLES, G.J., GILL, B.S., XIN, Z.-Y., CLARKE, B.C., MCINTYRE, C.L., CHAPMAN, C. & APPELS, R. 1988: Frequent duplication and deletion events in the 5S rRNA genes and the associated spacer regions of the *Triticeae*. *Plant Syst. Evol.* **160**: 105–122.
- STEANE, D.A., BYRNE, M., VAILLANCOURT, R.E. & POTTS, B.M. 1998: Chloroplast DNA polymorphism signals complex interspecific interaction in *Eucalyptus* (Myrtaceae). *Austr. J. Bot.* **11**: 25–40.
- STEELE, K.P., HOLSINGER, K.E., JANSEN, R.K. & TAYLOR, D.W. 1991: Assessing the reliability of 5S rRNA sequence data for phylogenetic analysis in green plants. *Mol. Biol. Evol.* **8**: 240–248.
- SWOFFORD, D.L. 1998: PAUP\*. Phylogenetic Analysis Using Parsimony (\* and other methods). Version 4. Sinauer Associates, Sunderland Massachusetts.
- TRONTIN, J.-F., GRANDEMANGE, C. & FARVE, J.-M. 1999: Two highly divergent 5S rDNA unit size classes occur in composite tandem array in European larch (*Larix decidua* Mill.) and Japanese larch (*Larix kaempferi* (Lamb.) Carr.). *Genome* **42**: 837–848.
- TYSON, M., VAILLANCOURT, R.E. & REID, J.B. 1998: Determination of clone size and age in a Mallee eucalypt using RAPDs. *Austr. J. Bot.* **46**: 161–172.
- UDOVICIC, F., MCFADDEN, G.I. & LADIGES, P.Y. 1995: Phylogeny of *Eucalyptus* and *Angophora* based on 5S rDNA spacer sequence data. *Mol. Phylogen. Evol.* **4**: 247–256.
- VAN WYK, G. 1985: Genetic variation in wood preservation of fast growing *Eucalyptus grandis* families. In: *Forest products research international – achievements and the future*, Proceedings of the IUFRO Symposium Pretoria April 1985, Vol 1, pp. 1–16, CSIR, Pretoria.
- WHIFFIN, T. 1981: Analysis of hybridisation between *Eucalyptus pauciflora* Sieber ex Spreng. and *E. radiata* Sieber ex DC (Myrtaceae). *Bot. J. Linn. Soc.* **83**: 237–250.
- ZUKER, M., MATHEWS, D.H. & TURNER, D.H. 1999: Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. In: *RNA Biochemistry and Biotechnology*. (eds. J. Barciszewski & B.F.C. Clarke) NATO ASI Series, Kluwer Academic Publishers, pp. 11–43.