Theor Appl Genet (2000) 101:473–477

R.E. Vaillancourt · H.D. Jackson

A chloroplast DNA hypervariable region in eucalypts

Received: 20 March 1999 / Accepted: 16 December 1999

Abstract Eucalypt chloroplast DNA (cpDNA) provides useful markers for phylogenetic and population research including gene flow and maternity studies. All cpDNA studies in *Eucalyptus* to date have been based on the RFLP technique, which requires relatively large amounts of clean DNA. The objective of this study was to develop PCR-based cpDNA markers for *Eucalyptus*. The chloroplast genome of *Eucalyptus*, like that of most angiosperms, possesses inverted repeats (IR). The two junctions between the IRs and the large single copy (LSC) regions are defined as J_{LA} and J_{LB}. The region surrounding the J_{LA} junction was sequenced from 26 *Eucalyptus* DNA samples (21 of *E. globulus*, plus 5 other species), and the J_{LB} region was sequenced using 5 of these samples. The samples were chosen to represent all major haplotypes identified in previous cpDNA RFLP studies. The J_{LA} products were 150–210 bp in size, while those from J_{LB} were approximately 500 bp in size. Eighteen mutations were scored in total. Extensive variation was found in the IR within the intergenic spacer between *rpl*2 and the IR/LSC junctions. Many of these characters were complex indels. One sample of *E. globulus* possessed a relatively large (38 bp) insertion which caused displacement of the IR/LSC junctions. This is the first report of intraspecific variation in the position of IR/LSC junctions; interspecific variation was also found. The LSC region near J_{LB} had a low rate of mutation and none were informative. The LSC region near J_{LA} possessed 2 informative mutations. The variation revealed from these sequences reflects the differentiation previously uncovered in an extensive RLFP analysis on the same samples. These results suggest that the J_{LA} region will provide very useful cpDNA polymorphisms for future studies in *Eucalyptus*.

Communicated by A.L. Kahler

R.E. Vaillancourt (✉) · H.D. Jackson
Cooperative Research Centre for Sustainable Production Forestry and School of Plant Science, University of Tasmania, GPO Box 252–55, Hobart, Tasmania 7001, Australia
e-mail: R.Vaillancourt@utas.edu.au
Fax: +61 3 6226 2698

Key words Chloroplast DNA · J_{LA} · PCR · *Eucalyptus* · Inverted repeat

Introduction

Eucalypt species are fast becoming the hardwood species of choice in plantations throughout the world, and *E. globulus* is the most important of these species in temperate regions of the world (Eldridge et al. 1993). Chloroplast DNA (cpDNA) has been proven to be useful in phylogenetic (Steane et al. 1991; Sale et al. 1996), population (Byrne and Moran 1994; Jackson et al. 1999) and evolutionary studies in *Eucalyptus* (Steane et al. 1998; Jackson et al. 1999; McKinnon et al. 1999). All cpDNA studies in *Eucalyptus* to date have been based on the restriction fragment length polymorphism (RFLP) technique. This requires relatively large amounts of clean DNA, which is time-consuming to obtain in eucalypts. The cpDNA microsatellites developed for gymnosperms (Powell et al. 1995; Vendramin et al. 1996) do not transfer well to *Eucalyptus* (D.A. Steane personal communication). Consequently, development of cpDNA markers in *Eucalyptus* that would enable the RFLP technique to be avoided would prove useful in studies that require a large sample size.

In three independent RFLP studies (Steane et al. 1998; Jackson et al. 1999; McKinnon et al. 1999) the cpDNA of *Eucalyptus* was found to be very variable with respect to the size of restriction fragments when probed with *Petunia* probe nos. 20 and 19 (Sytsma and Gottlieb 1986). This variation, which appeared as a wobble in band size, could not be used in these studies because it could not be ascribed to simple restriction site mutations or indels. These *Petunia* probes hybridize to a cpDNA region that overlaps the junctions between the large single-copy region (LSC) and the inverted repeats (IR). Byrne et al. (1993) found that the chloroplast genome of *Eucalyptus* possesses the IR structure characteristic of most angiosperms, such as found in *Petunia* and *Nicotiana*. The IR is a duplicated section of cpDNA, and

in each duplicate (25 kb in size) genes are in the opposite order. The IRs are separated by two single-copy regions, one large (LSC, 80 kb in size) and the other small (SSC, 21 kb in size; Byrne et al. 1993). There are two junctions between the LSC and IR, and these are defined as J_{LA} and J_{LB} in Shinozaki et al. (1986). Goulding et al. (1996) developed primers, anchored in conserved genes on either side of the junctions (for J_{LA} *rpl*2 and *trn*H; for J_{LB} *rpl*2 and *rpl*22), for amplifying both regions. The gene encoding the ribosomal protein *rps*19 is often proximal to J_{LB} on the LSC side of the junction and therefore can be found with these primers (Goulding et al. 1996). Goulding et al. (1996) found that the DNA region surrounding J_{LA} evolved rapidly by expansion and contraction in *Nicotiana*, while the J_{LB} region was much more conserved. In the study presented here, we investigated the potential of the J_{LA} and J_{LB} regions as a source of highly polymorphic polymerase chain reaction (PCR)-based cpDNA markers for *Eucalyptus*.

Materials and methods

The DNA used in this study were 22 samples from Jackson et al. (1999), representing all haplotypes therein, and 4 samples from Steane et al. (1998). The taxa belong to subgenus *Symphyomyrtus* (20 samples of *Eucalyptus globulus*; and 1 each of *E. dalrympleana*, *E. archeri*, *E. morrisbyi* and *E. vernicosa*) except for *E. obliqua* from subgenus *Monocalyptus*, which served as an outgroup. Primers for PCR amplification of the J_{LA} and J_{LB} regions were those developed by Goulding et al. (1996): for the J_{LA} region, *rpl*2, 5′-GATAATTTGATTCTTCGTCGCC-3′ (which is in both inverted repeats) and *trn*H, 5′-CGGATGTAGCCAAGTGGATC-3′ (in the LSC side of J_{LA}); for the J_{LB} region, *rpl*2 and *rpl*22, 5′-ACTCTTCGTGCTTTGTAGC-3′ (in the LSC side of J_{LB}). PCR reaction mixtures (50 µl final volume) contained 5 µl of *Taq* polymerase 10× buffer (Promega), 2 U *Taq* polymerase, 200 µ*M* of each dNTP, 2 m*M* $MgCl_2$, 100 µg/ml of BSA (bovine serum albumin), 7.5 pmol of each primer, 5 µl 50% glycerol and 20 ng of genomic DNA. Amplification was carried out in a PTC-200 (MJ Research) thermocycler programmed as follows: an initial denaturation for 5 min at 95°C followed by 35 cycles of 1 min at 94°C, 1 min at 46°C, 1 min at 72°C; and a final extension of 5 min at 72°C. For sequencing, PCR products were cleaned using the QIAquick column (QIAGEN) and then quantified using a fluorometer (Hoefer). Sequencing was performed using the BigDye terminator chemistry of Perkin Elmer (PE Applied Biosystems) on an ABI automated sequencer. Each PCR product was sequenced in both directions. Sequence alignments were done manually. To judge whether sequence characters were diagnostic of haplotypes, we mapped the characters onto the phylogenetic tree obtained from cpDNA RFLP analysis on the same samples by Jackson et al. (1999; Fig. 2 therein).

Results

The region of cpDNA spanning J_{LA} was sequenced in 26 samples of *Eucalyptus*. Five of these were also sequenced for the region of cpDNA spanning $J_{LB,}$ which enabled the exact position of the IR/LSC junctions to be determined (Fig. 1). Eighteen variable characters were scored in total (Table 1), 11 of which were found in the IR. All IR mutations occurred in the intergenic spacer between *rpl*2 and the end of the IR. A few of these mutations were simple base pair (bp) substitutions (4 out of

Table 1 Description of all variable characters scored from sequencing the J_{LA} and J_{LB} regions of the chloroplast DNA of *Eucalyptus*. Characters were scored by comparison with the outgroup

Character number and location	Character type	Character description and position (pos)[a]
Inverted repeats		
1	Transversion	G↔T, pos 38
2	Substitution	Outgroup state=C, 2.1=A, 2.2=T, pos 51
3	Transversion	T↔A, pos 58
4	Indel	5-bp insertion at pos 59–63
5	Multistate	Various insertions at pos 64–70: 5.0=gap; 5.1=AATTATT, 5.2=AATTATA; 5.3=AATTAAA, 5.4=AATTAT_
6	Multistate	Various number of As at pos 71–85: 6.2=A_2, 6.8=AATAAAAA, 6.9=A_9, 6.10=$A_{10,}$ 6.15=A_{15}
7	Indel	48-bp insertion at pos 80–127
8	Indel	13-bp deletion at pos 132–144
9	Indel	12 bp insertion at pos 134–145
10	Transversion	T↔G, pos 132
11	Indel	38 bp insertion at pos 94–131, originating from both sides of the J_{LB} junction
LSC of J_{LA}		
12	Indel	5-bp insertion, pos 146–150.
13	Transversion	G↔C, pos 146
LSC of J_{LB}		
14	Transition	T↔C, in *rps*19
15	Transversion	C↔A, in *rps*19
16	Transition	T↔C, in *rps*19
17	Transversion	G↔T, in *rps*19
18	Transition	T↔C, in *rps*19

[a] See Fig. 1

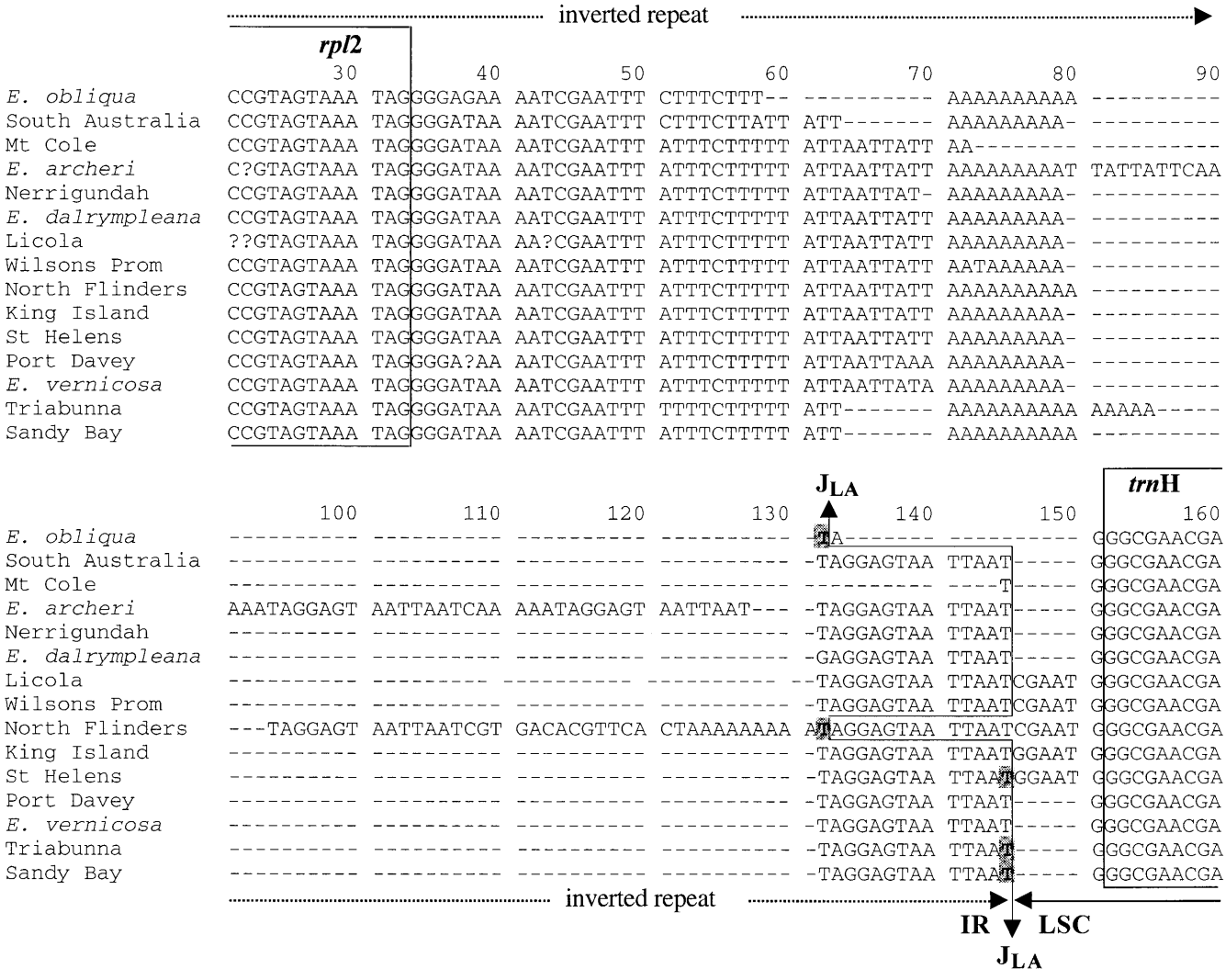

Fig. 1 Alignment of the sequence from the J_{LA} region in 15 *Eucalyptus* samples. Only samples with different haplotypes are shown. *Eucalyptus globulus* samples are labelled by locality name only. *Hyphens* represent gaps inserted for alignment. Samples where the end of the inverted repeat (*IR*) is *shaded* represent those that have been sequenced for J_{LB}. The position of the J_{LA} junction (*J_{LA}*), the IR and the large single-copy (*LSC*) region was then extrapolated for all samples as shown. The first 20 (within *rpl*2) and the last 65 (within *trn*H) base pairs of the J_{LA} products are not shown because they were highly conserved with tobacco

11), and many were complex sometimes involving large indels. The 5 mutations (1.2%) in the large single-copy region contiguous to J_{LB} (J_{LB}-LSC) were simple bp substitutions which mainly distinguished the outgroup from the ingroup. Two mutations were found in the large single-copy region contiguous to J_{LA} (J_{LA}-LSC), one of which was an indel.

The sample of *E. globulus* from North Flinders Island possessed a large insertion of 38 bp in the J_{LA} region (Table 1, character 11; Fig. 1). The sequence of this insertion is identical to the sequence flanking J_{LB}. The first 14 bp of this insert are the usual last 14 bp of the IR (as in the sample from St Helens) followed by a C, while the next 23 bp of the insert corresponds to the first 23 bp of *rps*19 (J_{LB}-LSC). A complete copy of *rps*19 was found outside of the inverted repeat, in the LSC portion of the J_{LB} region. Therefore, the 23-bp insertion of the *rps*19 sequence into J_{LA} results in a shift in the IR/LSC junctions in the North Flinders sample relative to other *E. globulus* samples. Four other indels are also possible duplications of nearby fragments (characters 4, 5, 7 and 12; Table 1), and all these were judged to be insertions by outgroup comparisons. The region between positions 64 and 80 has a high degree of variability, which makes character interpretation difficult. For example, character 5 may be an insertion, however there appear to have been a number of mutations in the fragment since the original insertion event. This would make it difficult to determine the original sequence that was inserted. Therefore, the variation in this region was ascribed to only 3 characters (5, 6 and 7), 2 of which have many independent states (Table 1).

The 18 sequence characters were mapped onto a phylogenetic tree obtained from cpDNA RFLP analysis by Jackson et al. (1999; Fig. 2). Most of the characters mapped to only one position on the tree (i.e. non-

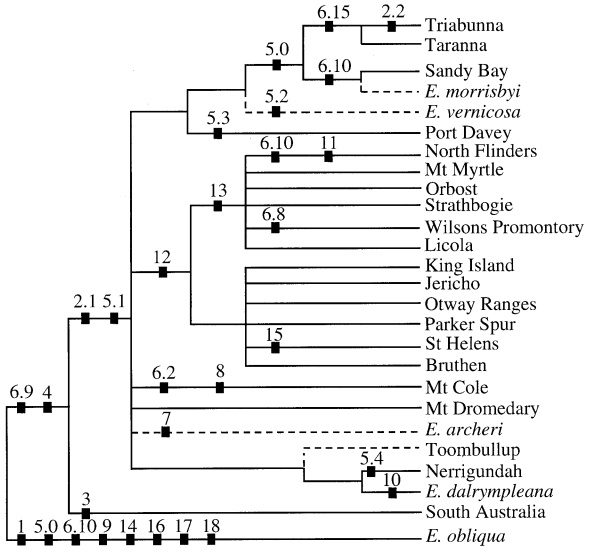

Fig. 2 Sequence evolution of the J_{LA} and J_{LB} regions of the chloroplast DNA of 21 *E. globulus* samples and five other eucalypt species as inferred by mapping the variable characters onto a phylogenetic tree obtained by Jackson et al. (1999) from the same samples but using RFLP markers. The clade defined by character 12 was redrawn to accommodate the better resolution afforded by character 13. *Branches* with *dotted lines* indicate samples that were analysed by Steane et al. (1998) and incorporated into the analysis of Jackson et al. (1999) on the basis of RFLP characters

homoplasious), which indicates agreement between the two data sets. For example, character 2 from the J_{LA} sequence places the South Australian sample as basal to the ingroup (Table 1, Fig. 2), as in the original RFLP analysis. Another important character was number 12, which supported a clade (Fig. 2) identical to an RFLP clade and similarly had a clear geographical pattern to its distribution amongst *E. globulus* samples. However, the structure within this RFLP clade was previously defined only by homoplasious RFLP characters. Therefore, this part of the tree has been redrawn, taking into account the differentiation obtained with the non-homoplasious sequence character 13. The mapping of multistate characters 5 and 6 onto the RFLP tree was more difficult, with one possible interpretation being shown in Fig. 2. The sequence data added further resolution to some of the RFLP clades. For example, the Triabunna, Taranna and Sandy Bay *E. globulus* samples (identical according to cpDNA RFLPs) are now all differentiated by the sequence data (Fig. 2).

Discussion

Sequencing the J_{LA} region has provided useful molecular markers. In eucalypts there is considerable sequence variation in the IR region close to the LSC. More informative characters were found in the J_{LA} than the J_{LB} region. The sequence variation uncovered using the J_{LA} region largely reflects the differentiation previously uncovered using extensive RLFP analysis on the same samples (Jackson et al. 1999). Jackson et al. (1999) identified two major cpDNA clades within Tasmanian samples of *E. globulus*. Our results show that these can be unambiguously distinguished by the presence or absence of a 5-bp indel (character 12). By mapping the sequence characters onto the RFLP phylogenetic tree of Jackson et al. (1999), we were able to show that the two types of data harbour the same phylogenetic signal. Some characters in the J_{LA} region will be useful in phylogenetic studies, although some difficulties may be encountered with the more complex characters. With characters that have many different states it is difficult to judge how many different mutations have actually occurred and whether these have followed any particular series. Despite these difficulties, these mutations are still useful in fingerprinting haplotypes.

Many of the mutations in the J_{LA} region are indels, and often these appeared to be duplications of nearby fragments. The reason why indels are so common in the junction regions is not well understood, but probably relates to the recombination and gene conversion between the two IRs that is presumed to occur continually (Goulding et al. 1996). Sequence mutations were also found to occur within the indel characters, indicating that this intergenic DNA is evolving at a high rate within *Eucalyptus*. The initial 20 bp of the intergenic spacer (bp 34–63), upstream of *rpl*2, is highly conserved between *Eucalyptus* and *Nicotiana*, suggesting that the mechanism mutating the DNA is mostly active close to the IR/LSC junction. In eucalypts, the gene arrangement across the IR/LSC junctions is similar to that found in *Nicotiana* (Goulding et al. 1996). A complete copy of *rps*19 was found within the LSC of J_{LB}, as in most angiosperms, however the truncated copy of *rps*19 (*rps*19′) commonly found in the IR side of J_{LA} (Goulding et al. 1996) was only found in 1 *Eucalyptus* sample (North Flinders). Six different *rps*19′ structures were observed among 13 *Nicotiana* species studied by Goulding et al. (1996), and these correlated to interspecific variation in the position of the IR/LSC junctions. In *Eucalyptus*, the position of the junctions was variable between and within species, with this being the first case of intraspecific variation reported.

The J_{LA} region has provided useful PCR markers for cpDNA that should be widely applicable in *Eucalyptus*. Some of the mutations result in fragment size variation large enough to be easily resolved on high-resolution agarose. These size mutations would be particularly well suited to maternity analysis (in mixed seed lots) and seed dispersal studies, since *Eucalyptus* chloroplasts are believed to be maternally transmitted (Byrne et al. 1993). In fact, the maternal inheritance of chloroplast needs to be verified in *Eucalyptus* on a large scale, since in the earlier analysis of Byrne et al. (1993) only a relatively small numbers of intraspecific F_1s were used and therefore the possibility of rare ‘leakage’ of chloroplasts through the pollen tube (Rajora and Mahon 1995) was not assessed. Interspecific hybrids (F_1) are commonly used in tropical *Eucalyptus* plantations, and these could

have abnormal chloroplast inheritance. This verification will be especially important if the cpDNA is to be used for transformation purposes (Daniell et al. 1998). One of the possible advantages of transforming the chloroplast is that maternal inheritance would lead to an easier containment of foreign genes in eucalypts, which have limited seed dispersal but more widespread pollen dispersal (Skabo et al. 1998). Chloroplast DNA polymorphism may prove useful in finding the original Australian provenance of eucalypt landraces dispersed throughout the world. For example, while naturally distributed in southeastern Australia and Tasmania, *E. globulus* was introduced to Portugal, Spain, India, USA, and many South American countries during the last century (Eldridge et al. 1993). The origins of the landraces are unknown in most cases because of poor records and possibly multiple introductions. In these countries, breeders are now attempting to fuse new collections of native Australian *E. globulus* with landrace material, and knowledge of the landrace origin could help minimize inbreeding effects. This cpDNA region may also provide useful markers for uncovering pathways of plant migration in eucalypts following events such as glaciations (Dumolin-Lapègue et al. 1997; King and Ferris 1998; Jackson et al. 1999). The fact that the J_{LA} region proved to be polymorphic in *Nicotiana* (Goulding et al. 1996) and *Eucalyptus*, genera from two distant families, suggests that this region may be applicable to other plant groups.

Acknowledgments We thank Gay McKinnon for technical assistance, Dot Steane and Brad Potts for their advice. This work was supported by the Cooperative Research Centre for Sustainable Production Forestry and a grant from the Australian Research Council. The experiments carried out in this project comply with Australian laws.

References

Byrne M, Moran GF (1994) Population divergence in the chloroplast genome of *Eucalyptus nitens*. Heredity 73:18–28

Byrne M, Moran GF, Tibbits WN (1993) Restriction map and maternal inheritance of chloroplast DNA in *Eucalyptus nitens*. J Hered 84:218–220

Daniell H, Datta R, Varma S, Gray S, Lee SB (1998) Containment of herbicide resistance through genetic engineering of the chloroplast genome. Nat Biotechnol 16:345–348

Dumolin-Lapègue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. Genetics 146:1475–1487

Eldridge K, Davidson J, Harwood C, Van Wyk G (1993) Eucalypt domestication and breeding. Oxford University Press, New York

Goulding SE, Olmstead RG, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. Mol Gen Genet 252:195–206

Jackson HD, Steane DA, Potts BM, Vaillancourt RE (1999) Chloroplast DNA evidence for reticulate evolution in *Eucalyptus* (Myrtaceae). Mol Ecol 8:739–751

King RA, Ferris C (1998) Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. Mol Ecol 7:1151–1161

McKinnon GE, Steane DA, Potts BM, Vaillancourt RE (1999) Incongruence between chloroplast and species phylogenies in *Eucalyptus* subgenus *Monocalyptus* (Myrtaceae). Am J Bot 86:1038–1046

Powell W, Morgante M, Andre C, McNicol JW, Machray GC, Doyle JJ, Tingey SV, Rafaski JA (1995) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. Proc Natl Acad Sci USA 92:7759–7763

Rajora OP, Mahon JD (1995) Paternal plastid DNA can be inherited in lentil. Theor Appl Genet 90:607–610

Sale MM, Potts BM, West AK, Reid JB (1996) Relationships within *Eucalyptus* (Myrtaceae) using PCR-amplification and southern hybridisation of chloroplast DNA. Aust J Syst Bot 9:273–282

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torozawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO 5:2043–2049

Skabo S, Vaillancourt RE, Potts BM (1998) Fine-scale genetic structure of *Eucalyptus globulus* ssp. *globulus* forest revealed by RAPDs. Aust J Bot 46:583–594

Steane DA, West AK, Potts BM, Ovenden JM, Reid JB (1991) Restriction fragment length polymorphisms in chloroplast DNA from six species of *Eucalyptus*. Aust J Bot 39:399–414

Steane DA, Byrne M, Vaillancourt RE, Potts BM (1998) Chloroplast DNA polymorphism signals complex interspecific interactions in *Eucalyptus* (Myrtaceae). Aust J Bot 11:25–40

Sytsma KJ, Gottlieb LD (1986) Phylogenetics of the *Lisianthius skinneri* (Gentianaceae) species complex in Panama utilizing DNA restriction fragment analysis. Evolution 39:594–608

Vendramin GG, Lelli L, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. Mol Ecol 5:595–598