High synteny and colinearity among *Eucalyptus* genomes revealed by high-density comparative genetic mapping

Corey J Hudson, Anand RK Kullan, Jules S Freeman, Danielle A Faria, Dario Grattapaglia, Andrzej Kilian, Alexander A Myburg, Brad M Potts and René E Vaillancourt.

Corey J Hudson (corresponding author), Jules S Freeman, Brad M Potts and René E Vaillancourt

School of Plant Science and CRC for Forestry, University of Tasmania, Private Bag 55 Hobart, Tasmania 7001, Australia.

Anand RK Kullan and Alexander A Myburg

Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0002, South Africa.

Danielle A Faria

EMBRAPA Genetic Resources and Biotechnology - Parque Estação Biológica - PqEB - Av. W5 Norte (final), Brasília, DF - Brazil - 70770-917.

Dario Grattapaglia

EMBRAPA Genetic Resources and Biotechnology - Parque Estação Biológica - PqEB - Av. W5 Norte (final), Brasília, DF - Brazil - 70770-917. Universidade Catolica de Brasília- SGAN, 916 modulo B, 70790-160 DF Brasilia, Brazil.

Andrzej Kilian

Diversity Arrays Technology Pty Ltd, PO Box 7141, Yarralumla, ACT 2600, Australia.

Corresponding author: Corey J Hudson

Email: cjhudson@utas.edu.au

Phone: +61 3 62262603

Fax: +61 3 62262698

Abstract

Understanding genome differentiation is important to compare and transfer genomic information between taxa, such as from model to non-model organisms. Comparative genetic mapping can be used to assess genome differentiation by identifying similarities and differences in chromosome organisation. Following release of the assembled *Eucalyptus grandis* genome sequence (January 2011; http://www.phytozome.net/), a better understanding of genome differentiation between *E. grandis* and other commercially important species belonging to the subgenus *Symphyomyrtus* is required. In this study, comparative genetic mapping analyses were conducted between *E. grandis*, *E. urophylla* and *E. globulus* using high density linkage maps constructed from Diversity Array Technology and microsatellite molecular markers. There were 236 – 393 common markers between maps, providing the highest resolution yet achieved for comparative mapping in *Eucalyptus*. In two intra-section comparisons (section *Maidenaria* - *E. globulus* and section *Latoangulatae* - *E. grandis vs. E. urophylla*), ~1% of common markers were non-syntenic and within chromosomes 4.7-6.8% of markers were non-colinear. Consistent with increasing taxonomic distance, lower synteny (6.6% non-syntenic markers) was observed in an inter-section comparison between *E. globulus* and *E. grandis* x *E. urophylla* consensus linkage maps. Two small chromosomal translocations, or duplications, were identified in this comparison representing possible genomic differences between *E. globulus* and section *Latoangulatae* species. Despite these differences, the overall high level of synteny and colinearity observed between section *Maidenaria* – *Latoangulatae* suggests that the genomes of these species are highly conserved indicating that sequence information from the *E. grandis* genome will be highly transferable to related *Symphyomyrtus* species.

Keywords: Eucalyptus, tree genomics, comparative mapping, chromosome rearrangement

Introduction

Genome sequences of model species are often used as references for related species. This enables the transfer of genetic information between species and accelerates research in non-model species. For example, the mouse genome sequence has long been used as a surrogate genetic resource in human disease research to map health-related quantitative trait loci (QTLs) and identify disease related genes that are difficult to study in human-based studies (Carver and Stubbs 1997). Similarly in plants, the genome sequences of rice (International Rice Genome Sequencing Project, 2005) and poplar (Tuskan et al. 2006) have been used as reference sequences for closely related grass (Krishnan et al. 2009) and tree species (Hamanishi and Campbell 2011), respectively. The extrapolation of genetic information from one organism to another is the essence of comparative genomics and a key component of this is comparative genetic mapping (Krutovsky et al. 2004). In such studies, the positions of homologous molecular markers mapped on genetic linkage maps of multiple pedigrees are compared through assessment of (1) synteny, the location of loci on homologous linkage groups, and (2) colinearity, the congruent ordering of loci on homologous linkage groups. By highlighting structural differences and similarities between species an understanding of chromosome and genome evolution can be gained (Laurie and Devos 2002; Paterson et al. 2000). The understanding of chromosome synteny and colinearity is also important for transferring genetic information (e.g. molecular marker, QTL and candidate gene positions; Celton et al. 2009) confidently between species.

Comparative mapping studies conducted in several plant families have consistently shown that closely related species exhibit high synteny and colinearity (Paterson et al. 2000; Paterson et al. 2009; Tang et al. 2008). However, chromosome segment duplications, mobility of DNA sequences, gene deletion and localised chromosomal rearrangements may create deviations from colinearity (Paterson et al. 2000). Despite the possibility of these occurrences, the relatively few comparative mapping studies conducted in forest trees to date have revealed high synteny and colinearity among closely related species. For example, comparison of amplified fragment length polymorphism (AFLP), expressed sequence tag polymorphism (ESTP) and microsatellite molecular marker loci in

pine tree linkage maps identified high synteny and colinearity between the species *Pinus sylvestris* and *P. taeda* which are estimated to have diverged some 70 MYA (Komulainen et al. 2003). Comparative mapping also suggested high genome conservation of spruce (*Picea*) species, with comparison of AFLP, ESTP and microsatellite linkage maps of *P. glauca*, *P. abies* and the species complex *P. mariana* x *P. rubens* revealing a remarkable conservation of gene content and order among conifer species (Pelgas et al. 2006). In *Populus*, Paolucci et al. (2010) compared the position of 86 microsatellite markers mapped in *P. alba* to their position in the *P. trichocarpa* genome sequence. A high level of synteny and colinearity was detected between these species, with 86% of markers being colinear (Paolucci et al. 2010). Comparative mapping has also been performed between *Fagaceae* species (e.g. oak, chestnut and beech). Although a lack of transferable markers has limited comparative mapping efforts between these species, results have indicated that strong macrosynteny exists between the closely related oak and chestnut genera (reviewed in Kremer et al. 2007).

The first mapped assembly (V1.0) of the *Eucalyptus grandis* genome was released in January 2011 (http://www.phytozome.net/). In order to exploit the full potential of this valuable genetic resource, a key question to address is, how confidently and to what extent can the sequence information of the *E. grandis* reference genome be transferred to other eucalypts? *Eucalyptus* represents an old Rosid lineage estimated to extend back to the Late Cretaceous (70 MYA, Crisp et al. 2004; Grattapaglia et al. submitted; Ladiges et al. 2003) and comprises some 700 extant species which are primarily endemic to the continent of Australia (Williams and Brooker 1997). Eucalypts are commonly the dominant or codominant flora in a wide range of environments, including tall forests, open woodlands and mallee shrublands (Byrne 2008). Due to their fast growth rates, adaptability, and excellent wood and pulp qualities, eucalypt species and hybrids now constitute the most widely planted hardwood crop world-wide for commercial forestry production (Eldridge et al. 1993; Grattapaglia and Kirst 2008). A recent taxonomic revision of the genus by Brooker (2000) recognised 13 *Eucalyptus* subgenera with most species belonging to the largest subgenus *Symphyomyrtus* (Brooker 2000). Molecular dating estimates suggest that considerable radiation of *Symphyomyrtus* species occurred 10–36 MYA (Crisp et

al. 2004) and today, *Symphyomyrtus* species and hybrids from sections *Latoangulatae* (e.g. *E. grandis* and *E. urophylla*), section *Maidenaria* (e.g. *E. globulus and E. nitens*) and section *Exsertaria* (e.g. *E. camaldulensis*) account for most of the 20 million hectares of eucalypt plantations established world-wide (Doughty 2000; Eldridge et al. 1993; Iglesias-Trabado and Wilstermann 2008; Myburg et al. 2007). Considering their importance to the commercial forestry sector, a better understanding of synteny and colinearity of eucalypt genomes is required.

The genome synteny and colinearity of three commercially important *Symphyomyrtus* species; *E. grandis* (W. Hill ex Maiden) and *E. urophylla* (S.T. Blake), both from section *Latoangulatae*, and *E. globulus* (Labill.) from section *Maidenaria* was examined in this study. *E. grandis* is one of the most widely planted subtropical eucalypts (Eldridge et al. 1993) with substantial plantations established in South America and Africa. *E. urophylla* is one of only two eucalypts that occurs exclusively outside of Australia being native to eastern Indonesia (Payn et al. 2007). It is often crossed with *E. grandis* to produce hybrid progeny which have higher disease resistance and are better adapted to tropical conditions in comparison to pure *E. grandis* (Pepe et al. 2004). *E. globulus* from section *Maidenaria* is regarded as the premier eucalypt for plantation forestry in temperate regions of the world due to its favourable wood qualities and broad adaptability (Eldridge et al. 1993; Grattapaglia and Kirst 2008).

Comparative mapping studies conducted in *Eucalyptus* to date have suggested high synteny between species (Brondani et al. 2006; Marques et al. 2002; Myburg et al. 2003). However, these past studies have been limited by their use of low density linkage maps which have contained few shared markers between species. In crop species, non-colinearity has been detected in regions previously thought to be colinear upon more thorough investigation (Lai et al. 2004; Peng et al. 2004). Hence, it is possible that small genome rearrangements may exist between eucalypt species which have not been detected in previous mapping studies. Comparative mapping with several hundred common homologous markers offers an opportunity to investigate colinearity at a much finer scale and substantially improve our understanding of genome evolution in *Symphyomyrtus*.

Common, transferable markers that segregate in multiple species, or pedigrees of interest, are essential for comparative mapping studies (Myburg et al. 2007; Neale and Krutovsky 2005). A variety of molecular marker types, including AFLP, random fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), microsatellite and gene-based markers have been used in eucalypt linkage mapping studies. Each of these DNA-based marker techniques vary in their DNA quality and quantity requirements, developmental costs, technical expertise, repeatability, degree of genetic information per locus and ease of establishing homology across studies and taxa (Myburg et al. 2007). A major limitation of the types of markers used in linkage maps published in the genus to date is that none have allowed both high-throughput genotyping and easy transferability among pedigrees, therefore limiting the resolution of comparative mapping efforts. Recently, a DArT marker genotyping system was developed for *Eucalyptus* (Sansaloni et al. 2010; Steane et al. 2011). DArT is a microarray hybridisation-based method that simultaneously assays hundreds to thousands of restriction fragment-based markers across a genome. The eucalypt DArT array contains 7680 selected, polymorphic markers which were generated from the DNA of 64 eucalypt species (Sansaloni et al. 2010). Therefore, the eucalypt DArT array can be considered to be a genus-generic genotyping tool as it can be used for genotyping most eucalypt species represented on the array. It is expected that DArT genotyping will provide 1000-2000 polymorphic DArT markers per eucalypt pedigree for linkage mapping studies (Sansaloni et al. 2010) and therefore, this technology offers a high-throughput, genome-wide genotyping platform with the potential to produce several hundreds of markers homologous between studies. Furthermore, the DNA fragments corresponding to the DArT markers on the eucalypt array have been sequenced, which has allowed their positions in the *E. grandis* genome to be determined (Kullan et al. accepted).

In the present study we focus on the subgenus *Symphyomyrtus* and examine genome synteny and colinearity among *E. grandis*, *E. urophylla* and *E. globulus* through comparative mapping with high-density DArT / microsatellite linkage maps. These are the highest marker-density maps yet used for comparative

mapping in eucalypts, allowing comparative mapping at a resolution which is unprecedented in the genus and rivalled by few studies in forest trees.

Materials and Methods

Synteny and colinearity of *E. grandis*, *E. urophylla* and *E. globulus* genomes were examined through comparison of five genetic linkage maps which were generated using three mapping pedigrees (Table 1). Three maps were generated in an *E. grandis* x *E. urophylla* (GU) double pseudo-backcross (BC) mapping pedigree (Kullan et al. accepted). Two of these were pure species maps for each of *E. grandis* and *E. urophylla*. The third map from this pedigree was an *E. grandis* x *E. urophylla* consensus linkage map (GU) which was generated by integrating linkage data from both backcross families. An outcrossed F_2 pedigree was used to generate the *E. globulus* F_2 Lighthouse linkage map while the *E. globulus* F_1 FAM4 linkage map was constructed using an inter-provenance F_1 family. Two intra-section analyses and one inter-section comparative mapping analysis were conducted using these five linkage maps. These were (1) *E. globulus* F_1 FAM4 *vs.* *E. globulus* F_2 Lighthouse (section *Maidenaria* intra-specific comparison), (2) *E. urophylla vs. E. grandis* (section *Latoangulatae* inter-specific comparison), and (3) *E. globulus* F_2 Lighthouse *vs.* GU consensus (section *Maidenaria vs*. section *Latoangulatae* comparison). Information from an 'additional' three *E. globulus* linkage maps (*i.e.* not used in comparative mapping analyses; Table 1 and Table S4; ESM_1.pdf) were used to investigate the linkage group position of non-syntenic markers detected.

For genotyping individuals, total genomic DNA was extracted using either a CTAB extraction protocol (Doyle and Doyle 1990) after grinding fresh leaf tissue in liquid nitrogen or through 'FastPrep' sample preparation (BIO 101/Savant FastPrep FP120; MP Biomedicals, Solon, OH) and subsequent DNA extraction with a Qiagen DNeasy 96 Plant kit (QIAGEN, Valencia, CA). DNA quality and concentration were estimated either by spectrophotometry (Nanodrop 8000, Thermo Scientific, USA) and/or by electrophoresis using 1% agarose gels stained with ethidium bromide and visualised under UV light; with comparison to a molecular weight marker (lambda *Hin*dIII, Promega). Fifty-five microsatellite markers were genotyped in the *E. globulus* F_2 Lighthouse family following the

screening of 245 microsatellite markers; 213 from Embra (Brondani et al. 1998; Brondani et al. 2002; Brondani et al. 2006; Faria et al. 2011; D. Faria unpublished), 36 from CSIRO (Byrne et al. 1996; Freeman et al. 2006; Glaubitz et al. 2001) and 3 from CRC (Steane et al. 2001). Markers were preferentially selected based on their degree of polymorphism (fully informative markers preferred) and estimated genome position based on previous linkage mapping studies (e.g. Brondani et al. 2006; Freeman et al. 2006). A small number of un-linked markers were also genotyped in *E. globulus* F_1 FAM1 (4 microsatellite markers) and F_1 FAM5 (5 microsatellite markers) families. Microsatellite markers were amplified using QIAGEN® Multiplex PCR Kits. Multiplex reactions typically contained 4-7 markers with conditions following the manufacturers microsatellite cycling protocol, except that the total final reaction volume was decreased (reagent ratios stayed the same) to either 12.5 or 5 µL. A 58˚C annealing temperature was used with PCR amplification conducted in either PTC-225 (MJ Research, Watertown, MA, USA) or Gene Amp PCR System 9700 programmable thermocyclers (Applied Biosystems, Foster City, CA, USA). Amplified PCR products from each multiplex were then simultaneously separated by capillary electrophoresis, using either a CEQ™ 8000 Genetic Analysis System (Beckman Coulter, Brea, CA, USA) with allele sizes estimated using the CEQ™ fragment analysis software by comparison to a CEQ™ DNA Size Standard-400, or, Applied Biosystems ABI PRISM 3100 or 3700 sequencers. ABI sequencing data was collected using Genescan® and analysed with Genotyper® software with comparison to the internal ROX™ 35-500 bp size standard (Applied Biosystems). Thirty-six microsatellite markers were genotyped in the *E. globulus* F_2 KI x Taranna mapping population, details of which have been reported in Freeman et al. (2006). Microsatellite genotyping of the GU pedigree has been described in Kullan et al. (accepted). For all mapping families prior to DArT genotyping, microsatellite marker genotype scores were used to check the pedigree of individuals by comparison to parental genotypes.

DArT genotyping was performed by Diversity Arrays Technology (DArT) Pty. Ltd. (Yarralumla, ACT, Australia) for all mapping pedigrees. For each individual, 10-15 µL of genomic DNA at 50-70 ng/µL was supplied. Samples were digested with *Pst*I/*Taq*I restriction enzymes and genotyped with a 7680-marker *Eucalyptus*

DArT array (Sansaloni et al. 2010). DArT genotyping scoring parameters were used to group markers into quality classes for linkage mapping, these included; call rate, reproducibility, *P* and or *Q*; an ANOVA based estimate of marker quality which reflects how well the two phase clusters (present = 1 *vs.* absent = 0) are separated (for further details see Jaccoud et al. 2001; Sansaloni et al. 2010).

Linkage analysis was performed in JoinMap 4.0 (Van Ooijen 2006) using similar mapping strategies and stringency levels for all mapping families. For *E. globulus* pedigrees, parental maps were constructed separately before constructing consensus maps for each linkage group. All loci were tested for goodness-of-fit to expected Mendelian segregation ratios using Chi-square goodness-of-fit tests in JoinMap 4.0, with the appropriate classification test set to allow for dominance in the testing of 3:1 segregating DArT markers. Linkage groups were defined at a minimum logarithm of odds (LOD) of three. Following the selection and formation of linkage groups, the Strongest Cross Link (SCL) information was inspected. Any assignment of ungrouped loci to linkage groups based on their SCL value were made in an iterative, '*one marker at a time*' fashion. Using this strategy, SCL values were recalculated and re-inspected after each change and the effect of marker reassignment assessed; which can help detect and avoid erroneous linkage group assignment (Van Ooijen 2006). Linkage group numbering and the orientation of linkage groups followed Brondani et al. (2006).

The regression algorithm (Stam 1993) was used for marker ordering within linkage groups. Default JoinMap 4.0 settings (recombination frequency < 0.40, goodness-of-fit jump threshold 5.0 and ripple 1) and Kosambi's mapping function were used with the default minimum LOD value increased to 3. *E. globulus* parental linkage maps were constructed in multiple stages, commencing with all microsatellite markers and highest quality DArT markers (classified as either 100% reproducibility and ≥ 80% call rate or $P \geq 90$) with lesser quality markers added in subsequent rounds (minimum DArT marker quality; reproducibility 85% and call rate 75% or $P \geq 70$). Following each mapping round, marker order was compared to the previous map, changes in marker order were reviewed and problematic markers were removed where necessary based on marker quality class and the following mapping parameters; maximum Chi-square goodness-of-

fit threshold, nearest neighbour fit, genotype probability function (Van Ooijen 2006) and the level of marker segregation distortion compared to surrounding markers. A maximum Chi-square goodness-of-fit threshold of 2.0 was applied for acceptance of marker fit in the *E. globulus* F_1 FAM4 linkage map and in the *E. globulus* F_2 Lighthouse framework (Keats et al. 1991) linkage map. An additional mapping round using relaxed marker fit parameters was performed in the *E. globulus* F_2 Lighthouse pedigree in order to map as many markers as possible. In this so-called comprehensive map a relaxed, maximum Chi-square goodness-of-fit threshold of 3.0 was applied. Following the construction of male and female maps in *E. globulus* pedigrees, the colinearity between markers segregating from both parents was inspected. Any non-colinearity was investigated and problematic markers were removed from linkage groups and marker-orders re-calculated. Having established colinearity between parental maps, consensus maps were then constructed using the '*Combine groups for map integration'* function of JoinMap 4.0. Linkage group marker-orders of consensus maps were compared to individual parental maps, any marker-order disagreements were investigated and poorly fitting markers were removed using the same criteria applied in parental map construction.

Individual component maps were also constructed in the *E. grandis* x *E. urophylla* pseudo-backcross pedigree before constructing an integrated GU consensus map. GU component maps were constructed in a single stage, using only markers with > 90% reproducibility, > 75% call rate and a Q-value > 60%. For these maps, a maximum Chi-square goodness-of-fit threshold of 3.0 was used for the acceptance of marker order fit with all other mapping parameters being as described above. The '*Combine groups for map integration'* function of JoinMap 4.0 was also utilised for construction of the GU consensus linkage map (Kullan et al. accepted).

In each comparative mapping analysis, common markers (mapping to both maps) were identified and syntenic markers were defined as those which had mapped to homologous linkage groups. Homologous linkage groups were plotted in MapChart 2.2 (Voorrips 2002) and the colinearity of syntenic markers was assessed. In linkage mapping, establishing the correct map position of tightly

linked markers in high density linkage maps is notoriously problematic (see Collard et al. 2009; Ferreira et al. 2006; Hackett and Broadfoot 2003). For example, in the simulation study of Collard et al. (2009) marker-orders within linkage group regions of high marker density were found to differ between independent analyses of the same marker data when using identical analysis parameters. It was found that identical marker-orders were obtained between independent analyses when these same datasets were pruned (e.g. tightly linked markers removed) to have a 1 centiMorgan (cM) marker resolution (Collard et al. 2009). Therefore, to avoid interpreting the possible error associated with ordering markers in high-density regions as regions of non-colinearity, a 1 cM threshold was applied for declaring non-colinearity. Markers were thus scored as non-colinear when a shift in rank-order marker position occurred and the shift in marker position exceeded 1 cM in both maps. The number of non-syntenic, non-colinear markers and their relative marker shift distance (cM) were recorded for each comparison. In order to graphically examine genome conservation, the position of common markers in each comparative mapping comparison were plotted in matrix plots (following Kaló et al. 2004) where the relative positions of common markers from each map were plotted on different axes. In these plots, syntenic and colinear markers occur on the diagonal plane across the plot within homologous linkage group 'cells', non-colinear syntenic markers occur off the diagonal plane within homologous linkage group cells, whereas non-syntenic markers occur in non-homologous linkage group cells, respectively.

To investigate potential causes of non-synteny, JoinMap 4.0 marker grouping (SCL-Value) and marker ordering fit statistics (mean chi-square goodness-of-fit and nearest neighbour fit values) were firstly re-examined in each map to ensure the correct linkage group assignment of markers. The linkage group assignments of non-syntenic markers were also compared to the three additional *E. globulus* linkage maps mentioned above. To investigate whether duplications of DArT marker loci within species genomes could explain non-synteny, BLAST searches of non-syntenic DArT markers (GenBank accession numbers HR865291-HR872186) were performed against the *E. grandis* genome assembly (V1.0 http://www.phytozome.net/). Non-syntenic DArT marker TBLASTX searches were also performed against the *Arabidopsis* Information Resource (TAIR)

database (http://www.arabidopsis.org/Blast/) in order to determine whether non-syntenic DArT markers resided within genes and/or had some biological function. Where necessary, the nucleotide sequence similarity of DArT markers were examined using the sequence alignment BLAST tool (bl2seq) available at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Results

Three independent mapping pedigrees were used to generate the five DArT and microsatellite genetic linkage maps used for comparative mapping analyses in this study (Table 2 and electronic supplementary material Tables S1-S3 available in the online resource file ESM_1.pdf). In each of the five *E. globulus* families, the total number of DArT and microsatellite markers used for linkage map construction ranged from 1000 (*E. globulus* KI x T F_2) to 1523 (*E. globulus* Lighthouse F_2). Markers which exceeded marker-ordering fit parameters were removed during map construction. This resulted in the removal of 19 (*E. globulus* KI x T F_2 consensus map) to 48 % (*E. globulus* F1 FAM4 consensus map) of markers throughout map construction (*i.e.* from parental map construction through to building the integrated consensus map). On average, 35 % of the starting dataset markers were removed during map construction in *E. globulus* families. Fifteen percent of markers were removed during construction of the GU consensus map (see Kullan et al. accepted). The GU consensus map was the highest density map (2290 markers, Table 2 and Table S1; ESM_1.pdf), having double and four times the number of markers than the parental consensus maps constructed in *E. globulus* F_2 Lighthouse (1060 markers; Table 2 and Table S2; ESM_1.pdf) and *E. globulus* F_1 FAM4 families (569 markers; Table 2 and Table S3; ESM_1.pdf), respectively. The mean marker interval length ranged from 0.48 cM for the GU consensus linkage map to 2.04 cM for the *E. globulus* F_1 FAM4 linkage map. All maps used for comparative mapping contained a moderate level of markers with distorted segregation ($\alpha \leq 0.05$ level). This level was highest in the parental *E. grandis* and *E. urophylla* backcross parent maps which contained 27.5% and 36.3% of mapped markers with segregation distortion, respectively, and less in the three consensus linkage maps (13.4 - 26.6%; Tables S3-5; ESM_1.pdf). Total map lengths of the consensus and *E. urophylla* linkage maps were very similar (1107-1151 cM; Table 2), while the *E. grandis* parental linkage

map had a total map length of 925 cM (Table 2). All maps used in this study comprised 11 linkage groups in accordance with the haploid chromosome number of *Eucalyptus* (Bachir and Abdellah 2006).

The two independently constructed *E. globulus* linkage maps (F_1 FAM4 and F_2 Lighthouse) had similar total map lengths; 1138 and 1151 cM, respectively. These map lengths agreed with, and fell roughly in the intermediate range, of previously published estimates for this species (701-1405 cM; reviewed in Myburg et al. 2007). On average, the two maps shared 21 DArT/microsatellite markers per linkage group (range 6-33). Only three DArT markers out of a total of two hundred and thirty-six (1.3%) common markers were found to be non-syntenic between these maps and 11 syntenic DArT markers were non-colinear (4.7%; Table 3). Nine of the eleven non-colinear markers were single-marker rearrangements and only one rearrangement involved two markers; this rearrangement occurred over a distance of less than 2 cM. Six of the non-colinear single-marker rearrangements occurred over distances less than 5 cM and only one rearrangement, which had good marker fit statistics in both maps, measured greater than 10 cM. The similar map lengths and high synteny and colinearity of *E. globulus* linkage maps suggested that the marker order of the *E. globulus* F_2 Lighthouse linkage map could be considered representative of the species and be used confidently in the inter-sectional comparison.

The inter-specific comparison within section *Latoangulatae* (*E. grandis vs. E. urophylla*) demonstrated a similar degree of genome similarity to the intra-specific comparison within section *Maidenaria* (*E. globulus vs. E. globulus*). *Eucalyptus grandis* and *E. urophylla* linkage maps shared on average 22 common DArT/microsatellite markers per linkage group (range 6-37), and only two (0.8%) DArT markers were found to be non-syntenic, while 17 (6.8%) syntenic markers were scored as non-colinear (Table 3). This included the only detected event of non-colinearity between microsatellite markers in any of the three map comparisons; a small (1.6 cM) marker-order inversion between Embra 226 and Embra 98 on linkage group 7.

The *E. globulus* F_2 Lighthouse and GU consensus linkage maps contained the highest number of mapped markers and as expected a greater number of common markers (393; Table 2 and Fig.1) were observed between these maps in comparison to both intra-section comparisons which used less marker-dense maps. A similar percentage of non-colinearity was detected between the *E. globulus* F_2 Lighthouse and GU consensus linkage maps (23 DArT markers or 6.3%; Table 3) to that found in both intra-section comparisons. Similarly, the majority of non-colinear loci in this comparison involved only single markers (19 out of 23) over small distances; 18 out of 23 non-colinear markers had a relative marker shift of < 5 cM between maps. However, a higher degree of non-synteny was observed in the *E. globulus* F_2 Lighthouse *vs.* GU comparison (26 DArT markers or 6.6%) in comparison to both intra-section comparisons (0.8-1.3%; Table 3). The overall degree of synteny and colinearity between *E. globulus* F_2 Lighthouse and GU consensus linkage maps can be seen in the matrix plot figure below (Fig. 1).

Most non-syntenic markers in the *E. globulus* F_2 Lighthouse *vs.* GU consensus occurred singly; with most occurring on linkage groups 7 and 5, respectively. There were four occurrences of non-synteny in the *E. globulus* F_2 Lighthouse *vs.* GU consensus comparison in which a pair of tightly linked DArT markers (termed marker-pairs) mapped to different linkage groups in the two maps (Table S5 ESM_1.pdf, numbered 1-4 in Fig. 1). For each marker pair, sufficient sequence similarity was detected between the two markers to classify them as identical (putatively redundant) DArT markers based on DArT marker sequence similarity (blseq2, NCBI; Table S5 ESM_1.pdf). Therefore, the 26 non-syntenic markers in this map comparison represent 22 unique loci. Another non-syntenic region was identified in the *E. globulus* F_2 Lighthouse *vs.* GU consensus comparison which contained three tightly linked DArT markers (ePt-504105, ePt-568705, ePt-637503, Fig. 1, #5). These markers mapped to *E. globulus* F_2 Lighthouse LG 7 (44.6-45.0 cM) and GU consensus LG 4 (28.0-31.3 cM). Results of DArT marker pair-wise BLAST comparisons indicated that these three markers were unique markers. Therefore, this non-syntenic region containing three unique (non-redundant) loci may represent a small chromosomal translocation between *E. globulus* and section *Latoangulatae* species. The species origin of these markers in the GU pedigree was investigated and all three were found to have originated

from the *E. urophylla* grandparent used to generate the F_1 hybrid of *E. grandis* and *E. urophylla*.

Examination of JoinMap 4.0 marker-linkage group SCL values and linkage group marker fit statistics of the 31 non-syntenic markers found across all comparisons did not suggest incorrect linkage group assignment (*i.e.* mapping error) for any of them. The non-syntenic DArT markers were variable in their sequence length (298 to 1013 bp), but most (19 out of 30) returned high BLAST matches (> 80% similarity over > 90% of DArT sequence) to the *E. grandis* genome enabling their position to be investigated; DArT marker ePt-599923 had a poor quality sequence and was excluded. Three non-syntenic DArT markers (ePt-574289, ePt-566325 and ePt-566325; 'both' under 'Genome support' in Table 4) returned high similarity BLAST matches to two different chromosome assemblies in the draft *E. grandis* genome assembly (V1.0, www.phytozome.net). In each case, the *E. grandis* chromosome scaffolds corresponded to the two different linkage groups to which the non-syntenic marker had been mapped. Thus, it appears that duplicated loci occur on different chromosomes within the *E. grandis* genome for these markers. A further 13 non-syntenic markers (including two marker pairs) returned high BLAST matches to a single *E. grandis* genome scaffold corresponding to one of the linkage groups to which the non-syntenic marker had been mapped. For each of these, the *E. grandis* genome scaffold base-pair position of the non-syntenic markers coincided approximately with its' mapped position (*i.e.* linkage group cM distance; data not shown). Of these 13 non-syntenic markers (or marker pairs), *E. grandis* genome BLAST results supported the placement of eight and five DArT markers in the *E. globulus* F_2 Lighthouse and GU consensus linkage maps, respectively (Table 4).

Fourteen of the thirty-one non-syntenic markers had been mapped in at least one 'additional' linkage map (*i.e.* not used in that particular map comparison, Table 4; LGA column). When combining this linkage map information with *E. grandis* BLAST results, some conclusions can be drawn for at least some non-syntenic markers. Firstly, detection of duplicated DArT marker loci on different chromosome scaffolds within the *E. grandis* genome for three non-syntenic DArT markers (ePt-574289, ePt-566325 and ePt-637292; Table 4) indicates that these

markers were correctly positioned in each linkage map. Of the remaining 28 non-syntenic markers, only three markers (all in the *E. globulus* F_2 Lighthouse *vs.* GU consensus comparison) had linkage group support for the placement of the marker from both the *E. grandis* genome and additional linkage maps. For example, the placement of DArT markers ePt-566850, ePt-504766 and ePt-572057 in the *E. globulus* F_2 Lighthouse map was supported by both the *E. grandis* genome and the fact that these markers were mapped to similar linkage group positions in additional linkage maps. This suggests the possible erroneous mapping of these markers in the GU consensus linkage map. Only one case was found where additional mapping information conflicted with *E. grandis* genome placement. In this case, the marker pair ePt-599965/ePt-643259 was mapped to LG 3 in both *E. globulus* F_2 Lighthouse and *E. globulus* F_1 FAM4 linkage maps. However, the *E. grandis* genome supported the placement of this marker pair on LG 5; to which it had been mapped in the GU consensus linkage map (Table 4). This discrepancy could indicate a putative duplication or translocation of this locus to chromosome 3 in *E. globulus*.

Of the remaining 23 non-syntenic markers, seven had linkage support only and a further eight had genomic placement support based on *E. grandis* genome BLAST results only. While these data provide support for the correct placement of these non-syntenic markers to one of the linkage groups that it had been mapped to (*i.e.* either in the *E. globulus* F_2 Lighthouse or GU consensus map), it does not provide sufficient evidence to differentiate between mapping errors, or alternatively some genomic mechanism that may be responsible for the mapping of these markers to non-homologous linkage groups. Eight non-syntenic markers did not have any additional supportive evidence for their placement on either linkage group to which they were mapped.

Only five non-syntenic markers (or marker pairs; Table S5 ESM_1.pdf) returned moderate matches to TAIR accessions. DArT marker ePt-566850 showed moderate homology with an *Arabidopsis* transposable element gene (AT2G10840.1; 31/68 amino acid identities, e-value 0.029) and appeared to be duplicated within chromosome 11 of the *E. grandis* genome; this marker returned one full-length, high similarity BLAST match and four partial, high similarity

matches within a 27 kbp area (3,726,917 - 3,754,232 scaffold 11; http://www.phytozome.net/).

Discussion

The degree of genome synteny and colinearity between three commercially important *Eucalyptus* species were assessed through comparative mapping analyses using high marker-density genetic linkage maps. In each of these maps, markers were ordered within linkage groups in an iterative process. This involved calculating marker-orders, inspecting marker-order fit statistics and the removal of any poorly fitting markers and subsequent recalculation of marker-orders. This approach was also applied to consensus map construction in addition to closely inspecting marker colinearity between consensus and parental maps to ensure that consensus maps accurately represented parental map marker-orders. This mapping methodology in combination with the stringent mapping parameters applied and the marker-ordering power provided by the large progeny sizes of the mapping families used in this study resulted in the construction of robust linkage maps having high marker-order accuracy.

A very high degree of synteny and colinearity was detected in each of three comparative analyses. Based on the level of colinearity observed, it appears that no major inversions or translocations within chromosomes have occurred between species. In both intra-section comparisons approximately 1% of common markers were found to be non-syntenic. Less synteny (26 non-syntenic markers, 6.6%; Table 3) was observed in an inter-sectional comparison between *E. globulus* F_2 Lighthouse and *E. grandis* x *E. urophylla* (GU) consensus linkage maps and two possible small inter-chromosomal translocations, or duplications, were identified. An inspection of marker grouping and marker-order fit statistics of the markers involved in these rearrangements did-not suggest any erroneous marker-linkage group assignment of the non-syntenic markers. Therefore, it is suspected that these differences are likely to be real and not the result of mapping errors.

The first possible translocation, or marker duplication, involved the non-syntenic DArT marker pair ePt-599965/ ePt-643259 (marker pair 3; Fig. 1); which was the only case where evidence supported the mapping of non-syntenic markers to

different (non-homologous) linkage groups. Specifically, these markers mapped to linkage group 3 in the *E. globulus* F_2 Lighthouse linkage map (mapped to this same linkage group in the *E. globulus* F_1 FAM5 consensus map) and to linkage group 5 of the GU consensus linkage map (with placement on this linkage group supported by the *E. grandis* genome sequence; Table 4). The second possible translocation or duplication detected involved a small non-syntenic region containing three tightly linked DArT markers. These markers mapped to linkage group 7 of the *E. globulus* F_2 Lighthouse linkage map and to linkage group 4 of the GU consensus linkage map. Although no *E. grandis* genome or additional linkage map information supported the mapping of these markers to either of the non-homologous linkage groups to which they were mapped, it seems unlikely that the mapping of three tightly linked markers to different linkage groups would occur by chance.

Apart from the detection of two possible small genomic differences and a low level of non-synteny in the section *Latoangulatae* - *Maidenaria* map comparison, *E. grandis*, *E. urophylla* and *E. globulus* genomes were overall, highly syntenic and colinear as expected. Comparative mapping studies have shown that large regions of synteny and colinearity can be observed between highly divergent species, even if they happen to differ in chromosome number and/or have substantial differences in DNA content (e.g. Hougaard et al. 2008; Kaló et al. 2004; Wu et al. 2009). Although the genome size of section *Latoangulatae* species have been reported to be ~20% larger than *E. globulus* (Grattapaglia and Bradshaw 1994; Praça et al. 2009), all eucalypts have the same chromosome number (haploid n=11; Bachir and Abdellah 2006) and previous smaller-scale comparative mapping studies have detected high synteny and colinearity among eucalypt species. For example, in a previous comparative study of *E. grandis* and *E. globulus*, Myburg et al. (2003) found perfect synteny and colinearity between 82 AFLP markers mapped to parental maps in a F_2 backcross mapping family. Their results supported the hypothesis of many dispersed regions of genome expansion (as opposed to a small number of gross chromosomal changes) as the likely cause for the substantial genome size differentiation between *E. grandis* and *E. globulus* (Myburg et al. 2003). Based on this finding we did not expect to find large regions of non-synteny between species in this study. Brondani et al. (2006)

also detected high synteny between parental *E. grandis* and *E. urophylla* linkage maps constructed with microsatellite markers. Although a much greater degree of non-colinearity (18% of 122 fully informative microsatellite markers) was reported compared to this study (4.7-6.9%; Table 3), much of the observed non-colinearity was reported to be likely attributed to analytical causes such as scoring errors and allele drop-outs generating apparent recombination events (Brondani et al. 2006), which may account for the different levels of non-colinearity between studies. Brondani et al. (2006) also reported, as found in this study, that generally only one or two markers were involved in each non-colinear rearrangement and there was no indication of major chromosomal rearrangements between *E. grandis* and *E. urophylla*.

Although all linkage maps were carefully constructed using the same marker-ordering algorithm and similar mapping parameter stringency, it is possible that analytical causes may be responsible for some of the non-colinearity observed. The GU consensus linkage map was built by integrating the recombination information from all four individual component maps generated in this double pseudo-backcross *E. grandis* x *E. urophylla* mapping pedigree. Although the two component maps of this pedigree (individual *E. grandis* and *E. urophylla* linkage maps) analysed in this study were highly colinear, a small number of marker rearrangements were detected. Such marker-order heterogeneity between component maps can adversely influence consensus map marker-orders (Gustafson et al. 2009) due to the '*combine groups for map integration*' function of JoinMap 4.0 using mean recombination frequencies and combined LOD scores when calculating integrated map recombination values (Van Ooijen 2006). Additionally, any bias associated with incorporating data from the different sized backcross mapping populations (Studer et al. 2010) could have also influenced the accuracy of marker-order in this consensus map. Furthermore, more general sources of error, including missing data, particularly in dense map regions (Hackett and Broadfoot 2003), and genotyping errors (Cheema and Dicks 2009; Slate 2008) could have lead to incorrect marker-orders (at a local scale) in any of the maps used for comparative mapping. Another potential source of error may arise from the mapping of markers with segregation distortion; which are a normal phenomenon in wide crosses (Semagn et al. 2006). However, at least two lines of

evidence suggest that the mapping of markers with distorted segregation ratios had little or no effect on marker non-colinearity. Firstly, the *E. globulus* Lighthouse F_2 linkage map contained 182 markers with distorted segregation and the majority (59%; data not shown) of these were added in 'comprehensive' mapping rounds (see Material and Methods). Despite adding such a high proportion of segregation distorted markers to the higher-quality framework map, the trusted marker order of the framework map was not affected. Secondly, the 11 markers found to be non-colinear between *E. globulus* linkage maps did not occur within linkage group regions of high segregation distortion in either map. Therefore, it seems more likely that a combination of factors, including the more complex mapping pedigrees of this study compared to previous eucalypt comparative mapping studies (e.g. Brondani et al. 2006; Myburg et al. 2003), the construction of maps by different operators, and other potential sources of error as described above could have contributed to the non-colinearity observed between maps in this study.

Mapping errors could also possibly explain the non-synteny observed between maps. However, when considering the agreement between the proportion of non-syntenic markers with taxonomic distance, the independent support for the mapping of markers to non-homologous linkage groups for at least one non-syntenic marker pair, and the fact that no obvious mapping errors were detected when re-examining the marker fit statistics of the non-syntenic markers, it appears that at least some of the non-synteny detected is likely to be real. It is possible that genomic mechanisms (e.g. transposable element activity and duplication or deletion events) may be responsible for this observed non-synteny. The non-syntenic marker ePt-566850 appeared to be duplicated within a small genomic region on chromosome 11 of the *E. grandis* genome and was found to have moderate homology with an *Arabidopsis* transposable element. Although there was no evidence for the duplication of this marker across non-homologous chromosomes, transposable elements have been shown to play an important role in the movement of genes between chromosomes and have been implicated as a cause of non-synteny between many closely related plant species (Bennetzen 2000; Bennetzen 2007; Bennetzen et al. 2005; Morgante et al. 2007). For example, transposable element activity has been suggested to explain cases of non-synteny

between *Picea* species (Pelgas et al. 2006), and Wu et al. (2009) suggested that the mapping of orthologous markers to non-homologous linkage groups in *Capsicum* and *Solanum* may have been a consequence of transposon-mediated marker transposition. Therefore, it is possible that transposable element activity may have similarly resulted in the movement of DNA between chromosomes in the eucalypt species examined and thus contributed to the low-level of non-synteny observed in this study.

Three non-syntenic DArT markers in this study were found to have duplicated loci in the *E. grandis* genome. The genome position of these loci corresponded to the alternative linkage groups in which the DArT markers had been mapped and these loci likely represent regions of ancient duplication which have been conserved following the divergence of *Latoangulatae* and *Maidenaria* sections. It is also possible that more recent species specific duplications have occurred, or conversely, that duplicated loci have been lost either at an intra-specific or species level following divergence. For example, just as transposable element activity can create structural changes within genomes (Bennetzen 2007), other genetic mechanisms including unequal homologous recombination and illegitimate recombination can generate small deletions in plant genomes (Bennetzen et al. 2005; Lysak et al. 2009). These processes are believed to be responsible for the gene deletions which have been documented in maize. For example, it was found that approximately 50% of duplicated genes have been lost in homologous chromosomal regions following the hybridisation of two closely related progenitors which gave rise to maize (Lai et al. 2004). Many small deletions detected in rice and *Arabidopsis* genomes are also believed to have resulted from these mechanisms (Bennetzen et al. 2005). Signatures of the mechanisms responsible for plant genome evolution have been detected in many characterised plant genomes (Bennetzen 2007) and it is now believed that plant genomes are remarkably more dynamic than once thought (Feschotte et al. 2002). Therefore, if these processes of duplication, transposition and deletion have been acting in *Eucalyptus*, greater opportunity would exist for these mechanisms to occur during the longer temporal separation of section *Latoangulatae* and *Maidenaria* species than between species within sections. This could account for the greater degree of non-synteny observed in the inter-sectional comparison.

In conclusion, the very high overall level of synteny and colinearity observed within and between species, and across two sections of subgenus *Symphyomyrtus* (sections *Latoangulatae* and *Maidenaria*), suggests that the genomes of these species are highly similar and that information from the *E. grandis* genome sequence will be highly transferable. Such high transferability can also be predicted for closely related *Symphyomyrtus* sections (e.g. *Exsertaria*; Steane et al. 2011) which contain other commercially important eucalypt species (e.g. *E. camaldulensis*). However, it is important to consider that a few small genome differences between species are expected to exist, e.g. at least two likely small genomic differences were detected between section *Latoangulatae* and *Maidenaria* species in this study, and that the number of such rearrangements is expected to increase between species having undergone more ancient species divergence.

Acknowledgements

We thank Norske Skog, Forestry Tasmania, Gunns Ltd, Australian Bluegum Plantations Pty Ltd, Western Australian Plantation Resources (WAPRES), David Pilbeam of the Southern Tree Breeding Association (STBA) and STBA for access to, and maintenance of *E. globulus* mapping family trials. Funding for this project was provided by the Australian Research Council (DP0770506 & DP110101621) as well as the Cooperative Research Centre for Forestry (Australia). We also thank Sappi Forest Research (South Africa) who generated and maintained the *E. grandis* x *E. urophylla* backcross mapping plant materials and the following organisations who provided financial support that contributed to the *E. grandis* x *E. urophylla* linkage mapping work; Sappi, Mondi, the Technology and Human Resources for Industry Program (THRIP), the National Research Foundation (NRF) and the Department of Science and Technology (DST) of South Africa.

References

Bachir O, Abdellah B (2006) Chromosome numbers of the 59 species of *Eucalyptus* L'Herit (*Myrtaceae*). *Caryologia* **59**, 207-212.

Bennetzen JL (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *The Plant Cell* **12**, 1021-1029.

Bennetzen JL (2007) Patterns in grass genome evolution. *Current Opinion in Plant Biology* **10**, 176-181.

Bennetzen JL, Ma J, Devos K (2005) Mechanisms of recent genome size variation in flowering plants. *Annals of Botany* **95**, 127-132.

Brondani RPV, Brondani C, Grattapaglia D (2002) Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. *Molecular Genetics and Genomics* **267**, 338-347.

Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theoretical and Applied Genetics* **97**, 816-827.

Brondani RPV, Williams ER, Brondani C, Grattapaglia D (2006) A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. *BMC Plant Biology* **6:20**.

Brooker MIH (2000) A new classification of the genus *Eucalyptus* L'Her. (Myrtaceae). *Australian Systematic Botany* **13**, 79-148.

Byrne M, Marquez-garcia M, Uren T, Smith D, Moran G (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. *Australian Journal of Botany* **44**, 331-341.

Byrne M (2008) Phylogeny, diversity and evolution of eucalypts. In 'Plant Genome - Biodiversity and Evolution'. (Eds AK Sharma and A Sahrma) pp. 303-346. (Science Publishers: Enfield (NH))

Carver E, Stubbs L (1997) Zooming in on the human-mouse comparative map: Genome conservation re-examined on a high-resolution scale. *Genome Research* **7**, 1123-1137.

Celton J-M, Chagne D, Tustin S, Terakami S, Nishitani C, Yamamoto T, Gardiner S (2009) Update on comparative genome mapping between *Malus* and *Pyrus*. *BMC Research Notes* **2**, 182-188.

Cheema J, Dicks J (2009) Computional approaches and software tools for genetic linkage map estimation in plants. *Briefings in Bioinformatics* **10**, 595-608.

Collard B, Mace E, McPhail M, Wenzl P, Cakir M, Fox G, Poulsen D, Jordan D (2009) How accurate are the marker orders in crop linkage maps generated from large marker datasets? *Crop and Pasture Science* **60**, 362-372.

Crisp M, Cook L, Steane D (2004) Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present day communities? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**, 1551-1571.

Doughty R (2000) 'The *Eucalyptus*: a natural and commercial history of the gum tree.' (John Hopkins University Press: Baltimore, MD)

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* **12**, 13-15.

Eldridge KG, Davidson J, Harwood C, van Wyk G (1993) 'Eucalypt breeding and domestication.' (Clarendon Press: Oxford).

Faria D, Mamani E, Pappas G, Grattapaglia D (2011) Genotyping systems for *Eucalyptus* based on tetra-, penta-, and hexanucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests. *Tree Genetics & Genomes* **7**, 63-77.

Ferreira A, Flores da Silva M, da Costa e Silva L, Cruz CD (2006) Estimating the effects of population size and type on the accuracy of genetic maps. *Genetics and Molecular Biology* **29**, 187-192.

Feschotte C, Jiang N, Wessler S (2002) Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics* **3**, 329-341.

Freeman JS, Potts BM, Shepherd M, Vaillancourt RE (2006) Parental and consensus linkage maps of *Eucalyptus globulus* using AFLP and microsatellite markers. *Silvae Genetica* **55**, 202-217.

Glaubitz JC, Emebiri LC, Moran GF (2001) Dinucleotide microsatellites from *Eucalyptus sieberi*: Inheritance, diversity, and improved scoring of single-based differences. *Genome* **44**, 1041-1045.

Grattapaglia D, Bradshaw HD (1994) Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. *Canadian Journal of Forest Research* **24**, 1074-1078.

Grattapaglia D, Kirst M (2008) Tansley review: *Eucalyptus* applied genomics: from gene sequences to breeding tools. *New Phytologist* **179**, 911-929.

Grattapaglia D, Vaillancourt RE, Shepherd M, Thumma BR, Foley W, Kulheim C, Potts BM, Myburg AA (submitted) Progress in *Myrtaceae* genomics: *Eucalyptus* as the pivotal genus.

Gustafson J, Ma X-F, Korzun V, Snape J (2009) A consensus map of rye integrating mapping data from five mapping populations. *TAG Theoretical and Applied Genetics* **118**, 793-800.

Hackett CA, Broadfoot LB (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* **90**, 33-38.

Hamanishi ET, Campbell MM (2011) Genome-wide responses to drought in forest trees. *Forestry* **84**, 273-283.

Hougaard BK, Madsen LH, et al. (2008) Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and Arachis. *Genetics* **179**, 2299-2312.

Iglesias-Trabado G, Wilstermann D (2008) *Eucalyptus universalis*; global cultivated eucalypt forests map 2008 (Version 1.0.1). GIT Forestry Consulting - EUCALYPTOLOGICS. http://git-forestry.com/download_git_eucalyptus_map.htm. Accessed 1/5/2011.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* **436**, 793-800.

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research* **29**, (4):e25.

Kaló P, Seres A, Taylor S, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis T, Kiss G (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. *Molecular Genetics and Genomics* **272**, 235-246.

Keats BJB, Sherman SL, et al. (1991) Guidelines for human linkage maps: An international system for human linkage maps (ISLM, 1990). *Genomics* **9**, 557-560.

Komulainen P, Brown GR, Mikkonen M, Karhu A, García-Gil MR, O'Malley D, Lee B, Neale DB, Savolainen O (2003) Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. *TAG Theoretical and Applied Genetics* **107**, 667-678.

Kremer A, Casasoli M, *et al.* (2007) Fagaceae trees. In 'Genome mapping and molecular breeding in plants'. (Ed. C Kole) pp. 161-188. (Springer-Verlag: Berlin)

Krishnan A, Guiderdoni E, et al. (2009) Mutant resources in rice for functional genomics of the grasses. *Plant Physiology* **149**, 165-170.

Krutovsky KV, Troggio M, Brown GR, Jermstad KD, Neale DB (2004) Comparative mapping in the *Pinaceae*. *Genetics* **168**, 447-461.

Kullan ARK, van Dyk MM, Jones N, Kanzler A, Bayley A, Myburg AA (accepted) High-density genetic linkage maps with over 2400 sequence-anchored DArT markers for genetic dissection in an F_2 pseudo-backcross of *Eucalyptus grandis* x *E. urophylla*. *Tree Genetics & Genomes.*

Ladiges PY, Udovicic F, Nelson G (2003) Australian biogeographical connections and the phylogeny of large genera in the plant family *Myrtaceae*. *Journal of Biogeography* **30**, 989-998.

Lai J, Ma J, et al. (2004) Gene loss and movement in the Maize genome. *Genome Research* **14**, 1924-1931.

Laurie D, Devos K (2002) Trends in comparative genetics and their potential impacts on wheat and barley research. *Plant Molecular Biology* **48**, 729-740.

Lysak M, Koch M, Beaulieu J, Meister A, Leitch I (2009) The dynamic ups and downs of genome size evolution in *Brassicaceae*. *Mol Biol Evol* **26**, 85-98.

Marques C, Brondani R, Grattapaglia D, Sederoff R (2002) Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. *Theoretical and Applied Genetics* **105**, 474-478.

Morgante M, De Paoli E, Radovic S (2007) Transposable elements and the plant pan-genomes. *Current Opinion in Plant Biology* **10**, 149-155.

Myburg AA, Griffin AR, Sederoff RR, Whetten RW (2003) Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F_1 hybrid based on a double pseudo-backcross mapping approach. *Theoretical and Applied Genetics* **107**, 1028-1042.

Myburg AA, Potts BM, Marques CM, Kirst M, Gion J, Grattapaglia D, Grima-Pettenatti J (2007) Eucalypts. In 'Genome mapping and molecular breeding in plants '. (Ed. C Kole) pp. 115-160. (Springer-Verlag: Berlin)

Neale DB, Krutovsky KV (2005) Comparative genetic mapping in trees: the group of conifers. In 'Biotechnology in agriculture and forestry, Vol. 55. Molecular marker systems in plant breeding and crop improvement'. (Eds H Lorz and G Wenzel) pp. 267-277. (Springer-Verlag Berlin)

Paolucci I, Gaudet M, Jorge V, Beritognolo I, Terzoli S, Kuzminsky E, Muleo R, Scarascia Mugnozza G, Sabatti M (2010) Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. *Tree Genetics & Genomes* **6**, 863-875.

Paterson AH, Bowers JE, et al. (2000) Comparative genomics of plant chromosomes. *Plant Cell* **12**, 1523-1540.

Paterson AH, Bowers JE, Feltus FA, Tang H, Lin L, Wang X (2009) Comparative genomics of grasses promises a bountiful harvest. *Plant Physiol.* **149**, 125-131.

Payn KG, Dvorak WS, Myburg AA (2007) Chloroplast DNA phylogeography reveals the island colonisation route of *Eucalyptus urophylla* (Myrtaceae). *Australian Journal of Botany* **55**, 673-683.

Pelgas B, Beauseigle S, Acheré V, Jeandroz S, Bousquet J, Isabel N (2006) Comparative genome mapping among *Picea glauca*, *P. mariana* × *P. rubens* and

P. abies, and correspondence with other *Pinaceae*. *TAG Theoretical and Applied Genetics* **113**, 1371-1393.

Peng JH, Zadeh H, et al. (2004) Chromosome bin map of expressed sequence tags in homoeologous Group 1 of hexaploid wheat and homoeology with rice and *Arabidopsis*. *Genetics* **168**, 609-623.

Pepe B, Surata K, Suhartono F, Sipayung M, Purwanto A, Dvorak WS (2004) Conservation status of natural populations of *Eucalyptus urophylla* in Indonesia and international efforts to protect dwindling gene pools. *Forest Genetic Resources* **31**, 62-64.

Praça MM, Carvalho CR, Novaes CRDB (2009) Nuclear DNA content of three *Eucalyptus* species estimated by flow and image cytometry. *Australian Journal of Botany* **57**, 524-531.

Sansaloni C, Petroli C, Carling J, Hudson C, Steane D, Myburg A, Grattapaglia D, Vaillancourt R, Kilian A (2010) A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. *Plant Methods* **6**, 16.

Semagn K, Bjornstad A, Ndjiondjop MN (2006) Principles, requirements and prospects of genetic mapping in plants. *African Journal of Biotechnology* **5**, 2569-2587.

Slate J (2008) Robustness of linkage maps in natural populations: a simulation study. *Proceedings of the Royal Society B* **275**, 695-702.

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant Journal* **3**, 739-744.

Steane DA, Vaillancourt RE, Russell J, Powell W, Marshall D, Potts BM (2001) Development and characterisation of microsatellite loci in *Eucalyptus globulus* (*Myrtaceae*). *Silvae Genetica* **50**, 89-91.

Steane DA, Myburg AA, Sansaloni CP, Petroli CD, Grattapaglia D, Kilian A, Vaillancourt RE (2011) DArT arrays for genetic mapping and diversity analysis of *Eucalyptus*. *Molecular Phylogenetics and Evolution* **59**, 206-224.

Studer B, Kolliker R, et al. (2010) EST-derived SSR markers used as anchor loci for the construction of a consensus linkage map in ryegrass (*Lolium* spp.). *BMC Plant Biology* **10**, 177.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and collinearity in plant genomes. *Science* **320**, 486-488.

Tuskan GA, DiFazio S, et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596-1604.

Van Ooijen J (2006) JoinMap 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V.: Wageningen, Netherlands.

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. *Heredity* **93**, 77-78.

Williams JE, Brooker MIH (1997) Eucalypts: an introduction. In 'Eucalypt ecology: individuals to ecosytems'. (Eds J Williams and J Woinarski) pp. 1-15. (Cambridge University Press: Cambridge U.K.)

Wu F, Eannetta N, Xu Y, Durrett R, Mazourek M, Jahn M, Tanksley S (2009) A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. *TAG Theoretical and Applied Genetics* **118**, 1279-1293.

Table 1 Pedigree and cross details of linkage maps used in each of three comparative mapping analyses. Details are also given for 'additional' linkage maps that were not used for comparative mapping but were used to inspect the linkage group position of non-syntenic markers

Pedigree	*N*	Cross details	Linkage map generated from pedigree used for comparative mapping
Used for comparative mapping			
E. globulus Lighthouse F_2	503	An outcrossed F_2 family generated from two inter-provenance F_1 individuals; BA0010 (Wilsons Promontory 614LH x King Island (KI) KI440) and BA0012 (Wilsons Promontory 615LH x Taranna (T) TA423)	Integrated parental consensus
E. globulus F_1 FAM4	183	Inter-provenance cross between Southern Tasmania tree #5797 x Flinders Island tree #5617	Integrated parental consensus
E. grandis x *E. urophylla* double pseudo-backcross F_2	547	An *E. grandis* x *E .urophylla* F_1 hybrid (GUSAP1) was backcrossed to (non-parental) individuals of both parental species (*E. grandis* tree P1381 and *E. urophylla* tree E142) to obtain two F_2 backcross (BC) families (Kullan et al. accepted)	1) *E. grandis* x *E. urophylla* consensus linkage map (GU); this integrated mapping data from both BC families
			2) Pure species *E. grandis* linkage map from *E. grandis* BC family (n=180)
			3) Pure species *E. urophylla* linkage map from *E. urophylla* BC family (n=367)
'Additional' pedigrees used to inspect non-syntenic markers only			
E. globulus F_1 FAM1	183	Inter-provenance cross between Eastern Otways tree #7479 x South-eastern Tasmania tree #5507	Integrated parental consensus[a]
E. globulus F_1 FAM5	184	Inter-provenance cross between Western Otways tree #4845 x Strzelecki Ranges tree #5474	Integrated parental consensus[a]

E. globulus KI x T F_2	172[b]	An outcrossed F_2 family generated from two inter-provenance F_1 individuals; G1060 (T7 x KI157) and G1026 (KI5 x T144) (Freeman et al. 2006)	Integrated parental consensus[a]

[a]Map used to inspect linkage group assignment of non-syntenic markers only. [b]Fifty-one additional individuals were added to the *E. globulus* KI x T F_2 family of Freeman et al. (2006).

Table 2 Linkage information for five genetic maps used for comparative mapping. Number of markers mapped (#M), linkage group length (cM; Kosambi) and mean marker interval length (MMI, cM) per linkage group are provided for each map. *glob* = *E. globulus*, LH = Lighthouse, GU = *E. grandis* x *E. urophylla* pseudo-backcross pedigree

	Linkage group											
Map	1	2	3	4	5	6	7	8	9	10	11	Total (Average)
glob F_1 FAM4												
#M	36	63	78	33	76	30	43	80	38	48	44	569 (52)
cM	91	111	126	98	117	87	91	98	110	113	97	1138 (103)
MMI	2.6	1.7	1.6	3.0	1.5	2.9	2.1	1.2	2.9	2.4	2.2	(1.99)
glob F_2 LH												
#M	62	124	114	59	120	115	113	119	94	81	59	1060 (96)
cM	95	123	103	79	102	141	92	130	85	98	103	1151 (105)
MMI	1.5	1.0	0.9	1.3	0.8	1.2	0.8	1.1	0.9	1.2	1.7	(1.12)
GU Consensus												
#M	180	235	257	163	220	236	169	274	210	156	190	2290 (208)
cM	89	102	106	80	110	137	84	119	89	98	95	1108 (101)
MMI	0.5	0.4	0.4	0.4	0.5	0.5	0.5	0.4	0.4	0.6	0.5	(0.48)
E. grandis backcross parent												
#M	75	109	107	77	109	105	89	121	83	42	74	991 (90)
cM	90	92	72	69	82	90	82	99	80	83	86	925 (84)
MMI	1.2	0.8	0.7	0.9	0.8	0.8	0.9	0.8	1.0	2.0	1.2	(0.99)
E. urophylla backcross parent												
#M	67	107	104	58	90	101	83	132	63	62	91	958 (87)
cM	85	107	113	87	99	125	87	123	79	94	108	1107 (101)
MMI	1.3	1.0	1.1	1.5	1.1	1.2	1.0	0.9	1.3	1.5	1.3	(1.19)

Table 3 Number of common markers, syntenic and colinear markers in each of three comparative mapping analyses. *glob* = *E. globulus*, LH = Lighthouse, GU = *E. grandis* x *E. urophylla* pseudo-backcross pedigree

Comparison	Common markers			Non-syntenic markers (% for total)			Non-colinear markers ([a]% for total)		
	DArT	SSR	Total	DArT	SSR	Total	DArT	SSR	Total
glob F_1 FAM4 *vs*. *glob* F_2 LH	230	6	236	3	0	3 (1.3)	11	0	11 (4.7)
E. grandis vs. *E. urophylla*	226	24	251	2	0	2 (0.8)	16	1	17 (6.8)
glob F_2 LH *vs.* GU consensus	377	16	393	26	0	26 (6.6)	23	0	23 (6.3)

[a]calculated as the percentage of syntenic markers

Table 4 Linkage group (LG) and position (cM) for non-syntenic (N.S.) markers in all three comparative mapping analyses. *E. grandis* genome BLAST results "Genome support" and information from additional linkage maps (in LGA column) provides support for the putatively correct marker linkage group assignment (LGA) or each non-syntenic marker

Comparison/Marker	N.S. marker position				[a]LGA support, linkage group (map)	[b]Genome support
	Map, LG and cM		Map, LG and cM			
glob F_2 LH *vs. glob* F_1 FAM1						
	glob F_2 LH		*glob* F_1 FAM1			
ePt-568818	7	83.0	5	6.0	2 (GU Con) 7 (FAM5)	-
ePt-574289	7	71.6	5	15.3	5 (GU Con)	both
ePt-643897	8	87.6	6	69.7	-	*glob* F_2 LH
E. grandis vs. E. urophylla						
	E. grandis		*E. urophylla*			
ePt-636534	5	41.6	1	50.5	5 (FAM5)	-
ePt-637292	2	33.8	8	71.9	-	both
glob F_2 LH *vs. E. grandis* x *E. urophylla* Consensus (GU)						
	glob F_2 LH		GU			
ePt-503782	8	90.3	11	87.1	-	GU
ePt-504105	7	45	4	31	-	-
ePt-504766	10	27.1	5	69.3	10 (FAM1)	*glob* F_2 LH
ePt-564413	6	40.6	8	98.1	-	*glob* F_2 LH
ePt-565169	3	99.8	4	61.3	-	*glob* F_2 LH
ePt-566325	7	71.9	5	95.9	7 (FAM4)	both
ePt-566850	11	1.7	2	34.7	11 (FAM4)	*glob* F_2 LH
ePt-567610	9	19.8	8	0.9	-	GU
ePt-568036	4	11.7	2	72.2	4 (FAM5)	-
ePt-568705	7	44.6	4	31.3	-	-
ePt-568818	7	83	2	55.1	7 (FAM5) 5 (FAM4)	-
ePt-568865/ePt-572057[c]	6	139	5	35.2	6 (572057; FAM1)	*glob* F_2 LH
ePt-571831	3	86.8	5	46.7	3 (FAM4)	-
ePt-574289	7	71.6	5	73.5	5 (FAM4)	both
ePt-574367/ePt-643036[c]	3	2.5	5	58.9	-	-
ePt-599923	7	81.8	3	36.8	7 (FAM1)	no sequence
ePt-599965/ePt-643259[c]	3	53.7	5	12.4	3 (both FAM5)	GU
ePt-600068	7	44.9	2	26.2	-	-
ePt-636589/ePt-638853[c]	7	47.6	3	82.8	-	-
ePt-637503	7	44.8	4	28	-	-
ePt-640753	9	22.6	3	23.6	-	GU
ePt-641639	4	48	8	37	-	*glob* F_2 LH

[a]Linkage group assignment support (LGA); linkage group and map (LG number (map)) to which the non-syntenic marker was mapped to in another linkage map not used in that comparative mapping comparison. FAM1=*E. globulus* F_1 FAM1, FAM4=*E. globulus* F_1 FAM4, FAM5=*E. globulus* F_1 FAM5, F_2 LH=*E. globulus* F_2 Lighthouse, F_2 KIxT=*E. globulus* F_2 KI x Taranna and GU con=*E. grandis* x *E. urophylla* consensus.
[b]Marker - *E. grandis* genome (V1.0 release; http://www.phytozome.net/) BLAST results supports the placement of the non-syntenic marker in this map. "Both" indicates support for both maps. [c]Marker pair, refer to Table S5 (ESM_1.pdf)

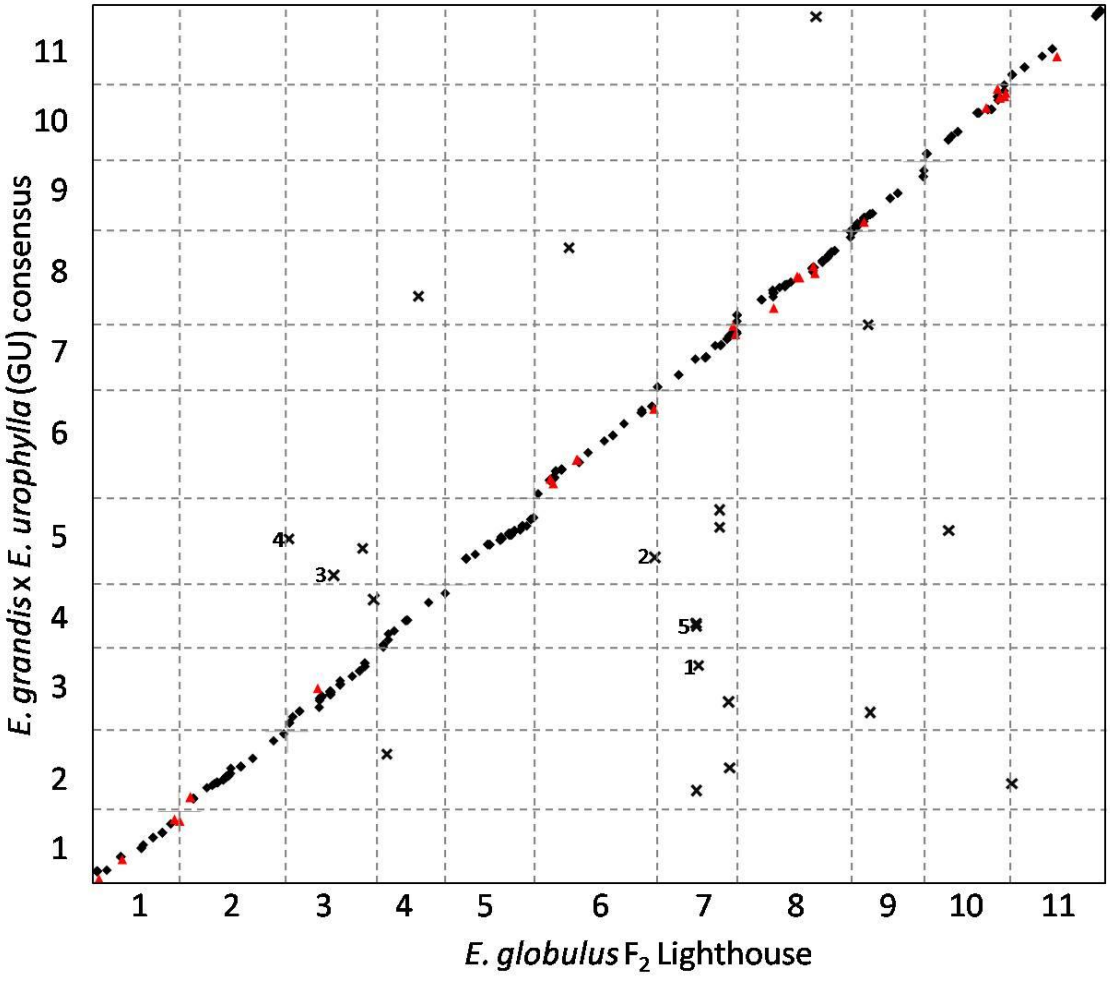

Fig. 1 Matrix plot showing the relative mapped position of 393 common DArT and microsatellite markers in the *E. globulus* F_2 Lighthouse (section *Maidenaria*, x-axis) and *E. grandis* x *E. urophylla* (GU) consensus (section *Latoangulatae*, y-axis) linkage maps. Broken lines indicate the proportional linkage group length (cM) of the 11 linkage groups (axes numbered). Diamond symbols represent syntenic colinear markers, triangles are non-colinear syntenic markers and cross symbols are non-syntenic markers. Numbers within the matrix indicate non-syntenic marker pairs (#1-4, see Table S5; ESM_1.pdf) and the position a triplet of unique non-syntenic markers (#5; discussed in text) which may represent a small translocation or marker duplication between *E. globulus* (section *Maidenaria*) and section *Latoangulatae* species

Table S1 Summary of *E. grandis* and *E. urophylla* parental linkage maps and the *E. grandis* x *E. urophylla* consensus (GU) linkage map

Map and LG	cM	Marker interval cM		Markers mapped			
		Average	Max	DArT	SSR	Total	[a]Seg. dist. (%)
***E. grandis* map**							
1	89.9	1.19	17.7	70	5	75	16 (21.3)
2	92.3	0.84	9.3	103	6	109	25 (22.9)
3	71.6	0.68	6.7	103	4	107	35 (32.7)
4	69.4	0.90	6.8	75	2	77	23 (29.8)
5	82.0	0.75	10.6	107	2	109	43 (39.4)
6	90.5	0.78	7.3	102	3	105	26 (24.7)
7	82.0	0.93	11.6	86	3	89	23 (25.8)
8	98.7	0.81	8.6	118	3	121	13 (15.7)
9	79.6	0.95	9.1	80	3	83	30 (36.1)
10	82.5	1.96	9.8	41	1	42	17 (40.4)
11	86.2	1.16	8.1	72	2	74	19 (25.6)
Total	924.7			957	34	991	270
Average		0.99		87	3.1	90.0	27.5%
***E. urophylla* map**							
1	85.1	1.27	8.7	61	6	67	13 (19.4)
2	107.1	1.00	11.4	101	6	107	27 (25.2)
3	113.4	1.09	6.7	99	5	104	40 (38.4)
4	87.4	1.50	7.9	54	4	58	12 (20.6)
5	98.5	1.09	13.1	86	4	90	60 (66.6)
6	124.8	1.23	8.9	98	3	101	25 (24.7)
7	87.1	1.04	7.8	78	5	83	56 (67.4)
8	123.1	0.93	9.8	126	6	132	69 (52.2)
9	79.0	1.25	11.2	59	4	63	12 (19.0)
10	93.7	1.51	11.2	61	1	62	14 (22.5)
11	108.2	1.28	8.4	89	2	91	20 (21.9)
Total	1107.3			912	46	956	348
Average		1.19		82.9	4.2	86.9	36.3%
***E. grandis* x *E. urophylla* consensus (GU)**							
1	88.8	0.49	4.4	173	7	180	39 (21.6)
2	102.1	0.43	5.0	228	7	235	46 (19.5)
3	105.5	0.41	3.1	251	6	257	85 (33.0)
4	79.8	0.48	5.5	157	6	163	25 (15.3)
5	110.4	0.50	9.6	218	2	220	104 (47.2)
6	136.9	0.58	8.2	232	4	236	60 (25.4)
7	83.5	0.49	7.0	163	6	169	68 (40.2)
8	119.1	0.45	6.6	263	11	274	61 (22.2)
9	88.5	0.42	6.5	203	7	210	31 (14.7)
10	97.7	0.62	5.0	155	1	156	44 (28.2)
11	95.3	0.50	4.8	186	4	190	47 (24.7)
Total	1107.6			2229	61	2290	610
Average		0.48		202.6	5.5	208.1	26.6%

[a]Seg. distortion; number and percentage of markers with segregation distortion $\alpha \leq 0.05$

Table S2 Summary of *E. globulus* F_2 Lighthouse comprehensive parental and integrated consensus linkage maps

Map and LG	cM	Marker interval cM		Markers mapped			
		Average	Max	DArT	SSR	Total	[a]Seg. dist. (%)
Female map							
1	96.8	2.42	14.15	36	5	41	9 (22)
2	129.7	1.44	16.07	86	5	91	3 (3.3)
3	103.6	1.02	13.46	100	3	103	17 (16.5)
4	81.1	1.69	10.35	46	3	49	9 (18.4)
5	98.4	1.30	15.50	72	5	77	18 (23.4)
6	140.6	1.67	16.82	77	8	85	10 (11.8)
7	96.2	1.08	22.09	87	3	90	24 (26.7)
8	125.0	1.37	14.78	88	4	92	8 (8.7)
9	84.3	1.26	13.07	65	3	68	24 (35.3)
10	97.1	1.67	18.31	55	4	59	3 (5.1)
11	94.6	2.96	15.66	30	3	33	3 (9.1)
Total	1147.5			742	46	788	128
Average		1.48		67.5	4.2	71.6	16.2%
Male map							
1	84.5	1.80	13.04	42	6	48	11 (22.9)
2	113.2	1.40	13.90	77	5	82	2 (2.4)
3	106.3	1.18	9.00	88	3	91	19 (20.9)
4	82.8	1.53	11.64	52	3	55	11 (20)
5	101.5	1.49	18.97	66	3	69	12 (17.4)
6	139.6	1.89	16.18	70	5	75	9 (12)
7	78.3	0.97	14.10	77	5	82	29 (35.4)
8	130.1	1.57	18.68	79	5	84	8 (9.5)
9	84.1	1.24	12.99	66	3	69	32 (46.4)
10	100.7	1.98	27.00	49	3	52	3 (5.8)
11	107.4	2.90	14.08	33	5	38	5 (13.2)
Total	1128.7			699	46	745	141
Average		1.54		63.5	4.2	67.7	18.9%
Integrated consensus map							
1	94.5	1.55	13.51	56	6	62	12 (19)
2	123.3	1.00	8.86	118	6	124	5 (4)
3	102.7	0.91	9.08	112	2	114	24 (21)
4	79.0	1.36	11.09	57	2	59	7 (12)
5	101.5	0.85	12.16	115	5	120	28 (23)
6	141.4	1.24	13.01	107	8	115	17 (15)
7	91.7	0.82	11.54	109	4	113	35 (31)
8	130.2	1.10	13.82	114	5	119	10 (8)
9	85.3	0.92	11.72	91	3	94	33 (35)
10	98.1	1.23	18.71	77	4	81	3 (4)
11	103.4	1.78	12.07	54	5	59	8 (14)
Total	1151.3			1010	50	1060	182
Average		1.10		91.8	4.5	96.4	17.2%

[a]Seg. distortion; number and percentage of markers with segregation distortion $\alpha \leq 0.05$

Table S3 Summary of *E. globulus* F_1 FAM4 parental and integrated consensus linkage maps

Map and LG	cM	Marker interval cM		Markers mapped			
		Average	Max	DArT	SSR	Total	[a]Seg. dist. (%)
Female map							
1	40.3	2.52	19.94	16	1	17	17 (0)
2	69.0	2.38	17.47	30		30	30 (13.3)
2B	35.9	3.26	15.68	12		12	12 (0)
3	128.0	2.78	32.15	47		47	47 (10.6)
4	80.7	3.23	24.03	25	1	26	26 (19.2)
5	111.1	3.37	23.35	34		34	34 (0)
6	33.0	2.36	7.30	13	1	15	15 (6.7)
6B	13.6	3.40	8.97	5		5	5 (0)
7	69.6	4.09	20.60	18		18	18 (0)
8	97.2	1.52	23.22	64	1	65	65 (0)
9	91.6	3.39	45.20	28		28	28 (17.9)
10	107.8	8.29	27.58	14		14	14 (0)
11	98.7	3.08	18.08	32	1	33	33 (27.3)
Total	976.4			338	5	344	29
Average		2.95		26		26	8.40%
Male map							
1	87.7	3.02	18.94	28	2	30	0 (0)
2	108.1	3.00	20.69	37		37	6 (16.2)
3	106.2	2.79	26.58	38		39	0 (0)
4	29.7	2.47	10.68	13		13	6 (46.2)
5	91.0	1.86	18.71	50		50	25 (50)
6	67.2	4.48	21.36	14	1	16	0 (0)
7	90.6	3.12	21.25	29		30	0 (0)
8	90.5	2.66	22.24	34	1	35	0 (0)
9	76.4	2.64	23.40	30		30	21 (70)
10	80.6	2.24	13.21	36		37	0 (0)
11	98.7	2.74	21.37	35	2	37	1 (2.7)
Total	926.5			344	6	354	59
Average		2.70		31		32	16.70%
Integrated consensus map							
1	91.2	2.61	19.97	34	2	36	0 (0)
2	111.2	1.79	15.97	63		63	9 (14.3)
3	125.6	1.63	10.02	77		78	5 (6.4)
4	97.6	3.05	24.10	32	1	33	6 (18.2)
5	117.1	1.56	19.09	76		76	25 (32.9)
6	86.6	2.99	21.28	28	1	30	1 (3.3)
7	90.8	2.16	13.14	42		43	0 (0)
8	97.8	1.24	11.63	79	1	80	0 (0)
9	110.0	2.97	23.40	38		38	21 (55.3)
10	112.7	2.40	21.25	47		48	0 (0)
11	97.5	2.27	22.83	42	2	44	9 (20.5)
Total	1137.9			558	7	569	76
Average		2.04		50.8		51.8	13.40%

[a]Seg. distortion; number and percentage of markers with segregation distortion $\alpha \leq 0.05$

Table S4 Summary of 'additional' *E. globulus* integrated consensus linkage maps (F_1 FAM1, F_1 FAM5 and KI x T F_2) which were used to investigate the linkage group position of non-syntenic markers

Map and LG	cM	Marker interval cM		Markers mapped					
		Average	Max	DArT	SSR	Gene	AFLP	Total	[a]Seg. dist. (%)
***E. globulus* F_1 FAM1**									
1	109.9	2.56	16.9	52	1			53	0 (0)
2	114.1	2.72	22.3	52				52	0 (0)
3	114.1	1.78	11.6	78		1		79	14 (17.7)
4	80.9	2.70	23.6	40				40	0 (0)
5	101.8	1.27	11.9	101				101	3 (3)
6	84.3	2.41	25.5	43	1			44	9 (20.5)
7	82.3	1.58	11.1	55		1		56	2 (3.6)
8	97.7	1.53	13.8	81	1			82	18 (22)
9	76.8	1.87	11.8	55				55	0 (0)
10	66.1	2.13	15.6	36				36	5 (13.9)
11	105.6	2.51	16.7	57	2			59	2 (3.4)
Total	1033.5			650	5	2	0	657	53
Average		1.97		59.1				59.7	8.1%
***E. globulus* F_1 FAM5**									
1	111.0	3.3	23.8	45	1			46	6 (13.0)
2	69.3	1.5	20.7	52				52	2 (3.8)
3	110.7	1.5	12.7	88		1		89	25 (28.1)
4	83.3	3.8	17.8	29				29	0 (0)
5	104.0	1.9	22.7	66				66	1 (1.5)
6	106.0	3.7	14.0	40	1			41	2 (4.9)
7	85.4	1.8	19.8	69				69	0 (0)
8	107.7	1.6	16.1	79	1			80	36 (45.0)
9	78.9	2.7	10.9	38				38	1 (2.6)
10	86.7	1.6	14.2	69		1		70	5 (7.1)
11	112.8	2.4	19.1	54	1			55	3 (5.5)
Total	1055.9			629	4	2	0	635	81
Average		2.1		57.2				57.7	12.8%
***E. globulus* KI x T F_2**									
1	96.2	2.0	14.3	42	6		9	57	0 (0)
2	124.0	3.0	23.6	46	4		9	59	27 (45.8)
3	121.3	1.7	18.4	81	2	1	10	94	10 (10.6)
4	83.6	2.5	16.1	46	4	2	9	61	3 (4.9)
5	109.9	1.8	8.7	71	3		17	91	25 (27.5)
6	151.5	3.1	26.4	50	4		12	66	28 (42.4)
7	112.4	1.7	9.5	74	4	1	14	93	4 (4.3)
8	147.2	2.1	12.1	85	3		12	100	10 (10)
9	90.6	2.5	14.9	59	1		5	65	6 (9.2)
10	102.5	2.3	17.0	61	2	1	4	68	16 (23.5)
11	123.2	2.9	15.2	45	3		8	56	5 (8.9)
Total	1262.5			660	36	5	109	810	134
Average		2.2		60.0	3.3		9.9	73.6	16.5%

[a]Seg. distortion; number and percentage of markers with segregation distortion $\alpha \leq 0.05$

Table S5 NCBI BLAST (blseq2) sequence identity and e-values for non-syntenic marker-pairs mapped to the same positions within each consensus map in the *E. grandis* x *E. urophylla* (GU) consensus *vs. E. globulus* F_2 Lighthouse (*glob* F_2 LH) comparative mapping analysis

		Linkage map position					Blast sequence similarity	
		GU consensus		glob F_2 LH		# base		
Pair	Marker	LG	cM	LG	cM	pairs[a]	Identity	e-value
1	ePt-636589[b]	3	82.9	7	47.5	971	309/317	7e-109
	ePt-638853	3	82.8	7	47.5	297		
2	ePt-568865	5	35.2	6	138.9	483	483/485	0
	ePt-572057	5	35.3	6	138.9	483		
3	ePt-599965	5	12.4	3	53.7	820	781/850 (gaps 33/850)	0
	ePt-643259	5	12.4	3	53.7	845		
4	ePt-574367	5	58.9	3	2.5	434	350/378 (gaps 2/378)	1e-158
	ePt-643036	5	59.0	3	2.5	379		

[a]Marker sequence length with adapter sequence/cloning vector fragments removed. [b]Marker 636589 was a poor sequence, the raw sequence length for this marker is shown