High-density genetic linkage maps with over 2400 sequence-anchored DArT markers for genetic dissection in an F2 pseudo-backcross of *Eucalyptus grandis* x *E. urophylla*

Anand Raj Kumar Kullan[1], Maria M van Dyk[1], Nicoletta Jones[2], Arnulf Kanzler[2], Arlene Bayley[3], Alexander A. Myburg[1*]

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

2 - Sappi Forests Research, Shaw Research Centre, PO Box 473, Howick, 3290, South Africa

3 - Sappi Technology Centre, PO Box 6, The Innovation Hub, Lynnwood Pretoria 0087, South Africa

* Corresponding Author:

Department of Genetics
University of Pretoria,
Pretoria,
0002,
South Africa

Email: zander.myburg@fabi.up.ac.za
Tel: +27-12-420-4945

Key words: Molecular marker, Consensus genetic linkage map, Comparative mapping

Abstract

Traits that differentiate cross-fertile plant species can be dissected by genetic linkage analysis in interspecific hybrids. Such studies have been greatly facilitated in *Eucalyptus* tree species by the recent development of Diversity Arrays Technology (DArT) markers. DArT is an affordable, high-throughput marker technology for the construction of high-density genetic linkage maps. *E. grandis* and *E. urophylla* are commonly used to produce fast-growing, disease tolerant hybrids for clonal eucalypt plantations in tropical and subtropical regions. We analyzed 7680 DArT markers in a F2 pseudo-backcross mapping pedigree based on an F1 hybrid clone of *E. grandis* and *E. urophylla.* A total of 2440 markers (31.7%) were polymorphic and could be placed in linkage maps of the F1 hybrid and two pure-species backcross parents. An integrated genetic linkage map was constructed for the pedigree resulting in 11 linkage groups (n = 11) with 2290 high-confidence (LOD ≥ 3.0) markers and a total map length of 1107.6 cM. DNA sequence analysis of the mapped DArT marker fragments revealed that 43% were located in protein coding regions and 90% could be placed in the recently completed draft genome assembly of *E. grandis*. Together with the anchored genomic sequence information, this linkage map will allow detailed genetic dissection of quantitative traits and hybrid fitness characters segregating in the F2 progeny and will facilitate the development of markers for molecular breeding in *Eucalyptus*.

Introduction

Eucalyptus tree species and their hybrids form the basis of the largest hardwood plantation crop in the world, occupying approximately 19.6 million hectares (www.git-forestry.com). Interspecific hybridization is important for the improvement of eucalypt plantations (Griffin *et al*. 1988, Eldridge *et al*. 1993, Khurana and Khosla 1998, Potts and Dungey 2004) yielding highly productive genotypes that are deployed in clonal eucalypt plantations in tropical and subtropical regions (Wright 1997, Campinhos and Ikemori 1989, Bison *et al*. 2006). *E. grandis*, a subtropical eucalypt in the section *Latoangulatae,* has been extensively used for the production of pulp due to its rapid growth, good form and easy vegetative

propagation. The species, however, has a low survival rate in humid and tropical areas, due to susceptibility to fungal diseases (Wingfield *et al*. 1989). *E. urophylla*, a tropical eucalypt native to islands of Indonesia and also a member of the section *Latoangulatae*, is more tolerant to fungal diseases than *E. grandis.* Interspecific hybrids of *E. grandis* and *E. urophylla* combine the fast growth and better rooting ability of *E. grandis* with the disease tolerance, adaptability and greater coppicing capability of *E. urophylla* (Vigneron and Bouvet 2000, Campinhos and Ikemori *et al*. 1989). Hybrids of *E. grandis* and *E. urophylla* are mainly grown in Brazil (Camphinos and Ikemori, 1977, Bison *et al*. 2006), the Congo (Vigneron and Bouvet 2000) and South Africa (Darrow 1995, Wright 1997). *E. grandis* x *E. urophylla* hybrids often exhibit superior growth and quality compared to the pure species, but the genetic architecture of hybrid superiority (Verhaegen *et al*. 1997, Grattapaglia *et al*. 1996) remains to be fully characterized in this hybrid combination.

Genetic linkage maps are useful for studying genome-wide patterns of inheritance of qualitative and quantitative traits, developing markers for molecular breeding, map-based cloning and comparative genomic studies. In the past two decades, important advances have been made in the construction of genetic maps for *Eucalyptus* species. The first generation of *Eucalyptus* genetic maps were constructed with Restriction Fragment Length Polymorphism (RFLP) markers (Byrne *et al*. 1995, Thamarus *et al*. 2002), Random Amplified Polymorphic DNA (RAPD) markers (Grattapaglia and Sederoff 1994, Vaillancourt *et al.* 1994, Verhaegen and Plomion 1996, Bundock *et al.* 2000, Gan *et al.* 2003) and Amplified Fragment Length Polymorphism (AFLP) markers (Marques *et al*. 1998, Myburg *et al*. 2003). However, the relatively low throughput of of these techniques (e.g. RFLP) and low proportion of polymorphisms shared among different outbred pedigrees (e.g. RAPD and AFLP) have hampered the integration of information from different maps, except where shared parents were used in mapping pedigrees (Myburg *et al.* 2003). More recently, several *Eucalyptus* genetic maps have been constructed using co-dominant microsatellite markers (Byrne *et al*. 1996, Brondani *et al*. 1998, Bundock *et al*. 2000, Thamarus *et al*. 2002, Brondani *et al*. 2002, Brondani *et al*. 2006, Freeman *et al*. 2006, Thumma *et al*.

2010), which proved informative for genetic analysis in outbred eucalypts, but still limited in throughput for rapid genome-wide genetic dissection. Although almost 300 microsatellite markers have already been mapped in eucalypts (Bundock *et al*. 2000, Thamarus *et al*. 2002, Brondani *et al*. 2006), the genus will still benefit from the availability of high-density genetic linkage maps with thousands of DNA markers anchored to a reference genome sequence. This will facilitate the identification of positional candidate genes and the identification of tightly linked QTL markers for molecular breeding.

Diversity Arrays Technology (DArT, Jaccoud *et al*. 2001) offers a rapid and affordable methodology for high-throughput DNA marker analysis. As DArT assays are performed in a highly parallel and automated fashion, the cost per data point is reduced by at least an order of magnitude compared to gel-based marker technologies, which makes it attractive to plant breeders aiming to track genome-wide segregation in large pedigrees. The technology was originally developed for rice (Jaccoud *et al*. 2001) and later validated in barley (Wenzl *et al*. 2006) and *Arabidopsis* (Wittenberg *et al*. 2005). DArT markers are currently being used in more than 55 species (http://www.diversityarrays.com/). A dedicated DArT genotyping array was recently produced for *Eucalyptus* tree species (Sansaloni *et al*. 2010). This array of 7680 markers was enriched for informative, polymorphic DArT markers by generating genomic representations from diverse *Eucalyptus* species and performing segregation analyses of more than 20,000 DArT polymorphisms in *Eucalyptus* mapping populations.

The aim of this study was to generate high-density genetic linkage maps for *E. grandis, E. urophylla* and an F1 hybrid of these species. We describe the use of a pseudo-backcross mapping pedigree to construct linkage maps of the parental genomes using DArT and microsatellite markers. The maps provide a high-resolution framework for future quantitative analysis of traits that differentiate the two species, as well as hybrid fitness traits that segregate in the F2 progeny.

Materials and Methods

Plant material and DNA extraction

A commercially grown F1 hybrid (*E. grandis* x *E. urophylla*) clone (GUSAP1, Sappi, South Africa) was selected for backcrossing to individuals of the parental species. Two F2 backcross (BC) mapping families were established using the F1 hybrid as a pollen parent with unrelated *E. grandis* (GSAP2) and *E. urophylla* (USAP1) individuals as seed parents in both crosses. Unrelated backcross parents were used to avoid potential inbreeding depression. The mapping pedigree consisted of 367 individuals from the *E urophylla* BC family and 180 individuals from the *E. grandis* BC family. DNA was isolated from all of the backcross individuals, the F1 hybrid, the two backcross parents and the original *E. grandis* (GSAP1) seed parent of the F1 hybrid using a BIO101/Savant FastPrep FP120 (MP Biomedicals, Solon, OH) instrument in conjunction with DNeasy 96 Plant kits (QIAGEN, Valencia, CA).

Marker analysis

A total of 71 previously published microsatellite markers were screened for polymorphism in the two backcross families (Table S1). Markers with the prefix "EMBRA" were previously developed from *E. urophylla* and *E. grandis (*Brondani *et al.* 1998*,* Brondani *et al.* 2006*),* "Eg" from *E. globulus* (Thamarus *et al*. 2002), "En" from *E. nitens* (Byrne *et al*. 1996) and "Es" from *E. sieberi* (Glaubitz *et al*. 2001). Two microsatellites (CesA1-MS1, CesA3-MS2) located in the promoters of cellulose synthase genes, *EgCesA1* and *EgCesA3* (Creux *et al*. 2009) were also used.

Multiplexed PCR amplification of the microsatellite markers was performed using the QIAGEN Multiplex PCR kit. The reactions were performed in a total volume of 10 µl containing 12 ng of template DNA, 0.2 µM of 10X primer mix (0.2 µM of each primer in mixes of up to 12 primer pairs each), and 1X QIAGEN Multiplex PCR master mix. PCR amplification was performed in an iCycler thermocycler (Bio-Rad Laboratories, Hercules, CA) with the following cycling conditions: initial denaturing and activation of the enzyme for 15 minutes at 94°C, followed by 35 cycles of denaturing at 94°C for 30 s, annealing at

50-60°C for 45 s, and extension at 72°C for 1 min, followed by final extension of 30 min at 60°C. Microsatellite primers were labeled with phosphoramidite fluorescent labels (6-FAM™, HEX™ or VIC™) for automated fragment analysis on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Life Technologies, Foster City, CA) using ROX™ (Genescan™ 500 ROX™) (Applied Biosystems) as internal standard. Electropherograms were analyzed using GeneMapper® 3.0 software (Applied Biosystems).

DArT marker assays were performed by Diversity Arrays Technology Pty Ltd (DArT P/L, Canberra, Australia) as described previously (Sansaloni *et al*. 2010).

Linkage analysis and parental map construction

Genetic linkage maps were constructed using JoinMap® 4 (Van Ooijen 2006) in combination with a two-way pseudo-testcross mapping strategy (Grattapaglia and Sederoff 1994). DArT and microsatellite markers were separated into three types: testcross markers segregating only in the hybrid parent (expected segregation ratio 1:1); testcross markers segregating only in the backcross parents (1:1); and intercross microsatellite (1:3, 1:2:1 or 1:1:1:1) and DArT (3:1) markers, segregating in both parents of the particular backcross. Four marker parental linkage maps were constructed: a maternal map of the *E. grandis* (GSAP2) backcross parent, a maternal map of the *E. urophylla* (USAP1) backcross parent, and two separate paternal maps of the F1 hybrid (GUSAP1). Segregation ratios were evaluated using the χ^2 test included in JoinMap® 4. For all four maps, linkage groups (LGs) were defined at a logarithm-of-the-odds (LOD) score of 8.0 or above. The marker order in each LG was subsequently determined by calculating the goodness-of-fit criterion and simultaneously calculating the map position corresponding to that order (Stam 1993) with the parameter settings Rec = 0.40, LOD = 3 and Jump = 5. The overall marker order of the linkage group was improved in each round by sequentially removing markers based on high mean chi-square values, nearest neighbour fit, and the genotype probability function as implemented in JoinMap® 4 (Van Ooijen 2006) and then reordering the remaining markers in the linkage group. Recombination

fractions were converted to additive map distances in centiMorgans (cM, Kosambi 1944). Linkage maps were drawn using MapChart© 2.2 (Voorrips 2002) and numbered according to the convention established by Grattapaglia and Sederoff (1994) and Brondani *et al*. (2006). Total genome length and genome coverage were calculated using the method of Lange and Boehnke (1982).

The parental origin of the testcross markers in the map of the F1 hybrid was inferred from genotypes obtained for the *E. grandis* (GSAP1) seed parent of the F1 hybrid (GUSAP1) since the two linkage phases in the maps of the F1 hybrid represent the markers amplified from either the *E. grandis* or the *E. urophylla* chromosome of each homologous pair.

Comparative mapping

The two maps of the F1 hybrid were aligned using shared testcross DArT (1:1) and shared microsatellite markers. Intercross DArT (3:1) and shared microsatellite markers were then used to align the backcross parent maps to that of the F1 hybrid. The parental maps were aligned using MapChart© 2.2 (Voorrips 2002). Where marker order differed between individual maps, markers were classified as non-colinear only when the difference in order involved markers that were spaced more than 1 cM apart.

Consensus map construction

An integrated (consensus) map for the entire pedigree was constructed using the 'combine groups for map integration' module in JoinMap® 4. The locus order was calculated using the regression mapping module and the following parameters: LOD $\geq$ 3.0, REC frequency $\leq$ 0.4, goodness- of-fit Jump threshold for the removal of loci = 5.0, the number of added loci after which to perform a ripple = 1, and third round = Yes. The heterogeneity test in JoinMap was used to exclude pairs of markers with significantly different recombination fractions in individual datasets. The overall marker order was improved iteratively as described earlier for parental map construction.

DNA sequence analysis of cloned DArT fragments

All of the cloned DArT fragments printed on the array were re-arrayed from plasmid stocks and Sanger sequenced in both directions (Genbank accessions HR865291-HR872186). To identify potential protein-coding regions mapped in the present study, the DArT fragment sequences were compared with all non-redundant GenBank CDS translations, RefSeq proteins, PDB, SwissProt, PIR, and PRF (http://www.ncbi.nlm.nih.gov) using BLASTX at a threshold of $1e^{-10}$. Customized scripts (Coetzer *et al.* 2010) were used to group redundant DArT fragments and assign functional annotations derived from BLASTX and BLAST2GO to each group. The DArT fragment sequences were also compared to the 8X draft assembly of the *E. grandis* genome sequence (DOE-JGI) using BLAST (http://eucalyptusdb.bi.up.ac.za/blast) at a threshold of $1e^{-10.}$ Marker sequences with more than 90% identity to the draft genome sequence were used to align the consensus linkage map with the corresponding superscaffolds in the V1.0 assembly of the *E. grandis* genome (DOE-JGI, www.phytozome.net).

Genome-wide distribution of genetic recombination

To investigate the genome-wide correlation of physical and recombination distances (bp vs cM), 153 genomic regions each corresponding to an approximately 1 cM interval were selected throughout the 11 linkage groups where both flanking markers were located on the same *de novo* assembled scaffold of the *E. grandis* 8X genome assembly (http://eucalyptusdb.bi.up.ac.za).

Results

Microsatellite polymorphism

A total of 68 (96%) microsatellite markers (Table S1), primarily from the EMBRA (Brondani *et al*. 2006) and CSIRO (Thamarus *et al*. 2002) sets, were found to be polymorphic in at least one of the backcross

families and were used for linkage mapping. Of the 63 markers polymorphic in the *E. grandis* backcross, 35 (55%) were informative in both parents and segregated with three to four alleles, 22 (35%) were only informative in the F1 hybrid (GUSAP1) and 6 (9.5%) were only informative in the *E. grandis* BC parent (GSAP2). Of the 64 markers in the *E. urophylla* backcross, 46 (72%) were informative in both parents, 14 (22%) were only informative in the F1 hybrid (GUSAP1) and four (6%) were only informative in the *E. urophylla* BC parent (USAP1). As expected, a higher proportion of microsatellite markers were polymorphic and segregated from the F1 hybrid, than from the backcross parent in each backcross family (90.4% vs 65.0% and 93.8% vs 78.1%, respectively).

DArT polymorphism

Of the 7680 markers on the DArT array, 3297 (43%) segregated in one or both backcrosses. Of these, 680 were excluded from the final mapping dataset based on filtering using three quality parameters (<90% reproducibility, <75% call rate and a Q-value <60%) and removal of markers for which the parental source could not be determined. The remaining 2617 markers were used for linkage map construction (Table 1). Of these, 1743 (66.6%) segregated in the *E. grandis* backcross pedigree and 1757 (67.1%) in the *E. urophylla* backcross pedigree, with 883 (33.7%) common between the two families. A higher proportion of testcross (1:1) DArT markers segregated out of the F1 hybrid than out of either backcross parent (37.5% vs 24.6% and 40.8% vs 22.8%, respectively, Table 1) consistent with the higher expected heterozygosity of the F1 hybrid.

Linkage analysis and parental linkage maps

The 68 microsatellite markers and 2617 DArT markers were used for the construction of four single-tree genetic linkage maps, one for each of the backcross parents and two for the F1 hybrid (Figure S1). All of the parental marker data sets separated into 11 main linkage groups (LG) corresponding to the haploid chromosome number of *Eucalyptus*. The final parental linkage maps contained a total of 2440 DArT and 67 microsatellite markers (Table 2). Total map lengths ranged from 924.7 cM for the *E. grandis* BC parent to 1107.3 cM for the *E. urophylla* BC parent with the F1 hybrid maps intermediate in size.

The genotypic ratios of a relatively large proportion of testcross and intercross markers deviated significantly from the expected Mendelian ratios in both backcross families (Table S2). Distorted markers were not excluded from the mapping analysis, because segregation distortion is expected to be prevalent in interspecific crosses and omitting such markers would result in low coverage in many regions of the genetic map (Myburg *et al*. 2003, Brondani *et al.* 2006). Chi-square testing revealed that 31.1% and 35.7% of the DArT markers showed significant ($\alpha = 0.05$) segregation distortion in the *E. grandis* and *E. urophylla* BC families, respectively (Table S2). Similar proportions of markers were distorted in the backcross parent maps and the two F1 hybrid maps (27.5% and 36.3% vs 32.1 and 32.3%, Table S2). Clusters of distorted markers that were observed throughout the four parental maps most likely represent true cases of genomic segregation distortion linked to postzygotic isolation barriers segregating in the F2 backcross progeny (Myburg *et al*. 2004). Some chromosomal regions exhibited segregation distortion in four parental maps, e.g. almost the entire length of LG5 and the distal end of LG7 showed distorted marker segregation in all four maps.

The large number of markers mapped resulted in high map coverage. On average, 80-91% of the loci in the BC parent and F1 hybrid maps were within 1 cM of a marker and 99.9% of loci in the four parental maps were within 5 cM of a marker.

Comparative and consensus maps

The two-way pseudo-backcross design, as well as the inclusion of multi-allelic microsatellite markers, allowed robust identification of homologous pairs of linkage groups representing the *E. grandis*, *E. urophylla* and F1 hybrid genomes (Figure S1). The large number of shared testcross and/or intercross (612) DArT markers and 46 microsatellite markers in the two maps of the F1 hybrid facilitated the alignment of these two maps. The linkage groups of the backcross parent maps were aligned to the two F1 hybrid maps with the use of 538 (23.4%) and 545 (23.7%) common markers in the *E. grandis and E. urophylla* BC families, respectively. The linkage maps of the two backcross parents were aligned with 251 (10.9%) common makers. Comparison of marker orders and map positions in the parental maps

(Figure 1) revealed only two non-syntenic marker placements between the *E. grandis* and *E. urophylla* BC parent maps. DArT marker ePT_636534 mapped to LG5 in the *E. grandis* BC parent map and LG1 in the *E. urophylla* BC parent map. Similarly, ePT_637292 mapped to LG2 and LG8 in the *E. grandis* and *E. urophylla* BC parent maps, respectively (Figure 1A). Apart from a small proportion of markers with different local orders (indicated by crossed lines, Figure S1), the locus order was largely conserved among the four parental maps. Excluding markers closer than 1.0 cM, 93.2%, 93.3%, and 95.1% of the markers were mapped with the same linear order in the *E. grandis* and *E. urophylla* BC parent maps, the *E. grandis* BC parent and F1 hybrid, and the *E. urophylla* BC parent and F1 hybrid maps, respectively.

The consistent ordering of markers in the four parental maps (Figure S1) allowed the construction of a high-density consensus linkage map for the *E. grandis x E. urophylla* backcross pedigree (Figure 2). The integrated linkage map comprised 2229 DArT and 61 microsatellite loci (Table 3). The total length of the consensus map was 1107.6 cM with an average marker spacing of 0.48 cM. Large numbers of perfectly co-segregating markers were also observed. Potential redundancy of DArT markers in the consensus map was evaluated by collapsing perfectly co-segregating loci into bins. A total of 1640 non-redundant bins were formed revealing that 28.3% of the mapped DArT markers were potentially redundant (i.e. possibly duplicate copies of the same cloned DArT fragment, or tightly linked). Besides co-segregation, regions of apparent DArT marker clustering was observed in all linkage groups, particularly in LG2, LG3, LG5, LG7 and LG9 (Figure 2). Clustering of markers in LG2, LG5 and LG7 has been reported in previous studies (Brondani *et. al*. 2006), supporting the possible biological basis for this occurrence. The locus order was well conserved between the consensus map and single-tree parental maps for all linkage groups (Figure S2). Only *E. grandis* LG1 and LG7 exhibited substantially shifted marker positions relative to the consensus map. This was also visible in the alignment of the parental maps (Figure S1) and may be the result of difference in map coverage at the ends of linkage groups (e.g. LG1) or due to different local rates of recombination in regions of the *E. grandis* homologs (e.g. LG7).

DNA sequence analysis of DArT fragments and alignment to the *E. grandis* genome sequence

DNA sequences were obtained for 6895 of the 7680 cloned DArT fragments on the array (Genbank accessions HR865291-HR872186). Of the sequenced markers, 2030 were polymorphic and could be mapped in this study (Table S3). Consistent with the previously reported enrichment of DArT markers in single copy DNA (Tinker *et al*. 2010), a comparison of the DArT fragment sequences to the non-redundant protein database (NCBI) using BLASTX ($<1e^{-10}$) revealed that 865 (42.6%, Table S3) of the marker fragments potentially contained protein coding sequences. Annotation of the putative protein coding sequences revealed a broad range of functional categories. Sequence analysis also revealed that 477 marker fragments (mapped to 305 loci) exhibited similarity to the same or similar protein sequences. Those mapping to different loci may represent duplicated gene loci or different gene family members in *Eucalyptus*, while those mapping to the same locus could be cloned copies of the same amplified DArT fragment (marker redundancy).

Mapping of the DArT marker sequences to the draft *E. grandis* genome sequence assembly (V1.0, DOE-JGI, http://eucalyptusdb.bi.up.ac.za/) identified 1836 (90.3%) marker sequences that could be placed in the genome (at an identity greater than 90% over the length of the sequence). The DArT markers placed in the genome cover approximately 600 Mbp (87%) of the sequenced genome space (690 Mbp) in the V1.0 *E. grandis* genome assembly (www.phytozome.net). The remaining 9.7% of the markers that could not be placed in the genome could have originated from unassembled parts of the *E. grandis* genome (gaps), or they may represent allelic variants of *E. grandis* or other *Eucalyptus* species, since the DArT array was constructed with DNA from a variety of species mainly *E. grandis, E. urophylla, E. globulus* and *E. nitens*, some of which are very distantly related to *E. grandis* (Sansaloni et al. 2010, Steane et al. 2011). The overall marker order was highly conserved between the consensus map and the *Eucalyptus* genome scaffolds in the draft 8X (V1.0) assembly of the *E. grandis* genome (Figure S3).

Genetic recombination

Comparison of marker intervals on the consensus genetic map to marker positions on *de novo* assembled scaffolds of the *E. grandis* genome (http://eucalyptusdb.bi.up.ac.za) enabled us to compare genetic distance and physical distance in the *Eucalyptus* genome, an important property for future map-based cloning efforts. Due to the early stage of the DOE-JGI *E. grandis* genome assembly, we expected the sequence to contain many gaps and some errors in assembly. We therefore selected 153 genomic intervals throughout the 11 linkage groups, each corresponding to an approximately 1 cM interval in the genetic map with both flanking markers placed in the same *de novo* assembled genomic scaffold. The average physical distance per centiMorgan in the 153 intervals was 633 kb with a range of 100 kb to 2.4 Mbp (Figure S4, Table S4).

Discussion

Dense genetic linkage maps are useful for genome-wide identification of molecular markers closely linked to genes or QTLs, the isolation of genes via map-based cloning, detailed comparative mapping, and genome evolution studies (Varshney *et al*. 2007). To develop resources for such investigations, we used DArT and microsatellite markers to construct high-density genetic linkage maps of *E. grandis*, *E. urophylla* and the fast-growing interspecific F1 hybrid of these two species. This is the first genetic linkage map of the F1 hybrid genome representing one of the most widely used hybrid combinations in commercial plantation forestry in tropical and subtropical areas. The consensus map of the pedigree provides a valuable resource for genetic analysis in *Eucalyptus* based on 2229 DArT and 61 microsatellite loci with excellent genome coverage for targeted marker saturation of economically important traits and new anchor points for evaluation of genome colinearity among *Eucalyptus* species.

Genetic maps previously reported for *Eucalyptus* species ranged from 919 cM to 1814 cM in length (Brondani *et al*. 2006). The parental maps constructed here ranged from 924.7 cM (*E. grandis* BC parent) to 1107.3 cM (*E. urophylla* BC parent) and 1107.6 cM for the consensus map. Despite high map

coverage, the *E. grandis* BC parent map (924.7 cM) was substantially shorter than maps reported earlier for this species (1552 cM - Grattapaglia and Sederoff 1994, 1415 cM -Verhaegen and Plomion 1996, 1335 cM - Myburg *et al.* 2003, 1814 cM - Brondani *et. al* 2006). Similarly, the *E. urophylla* BC parent map (1107 cM) was shorter than previously reported for the species (1331 cM - Verhaegen and Plomion 1996, 1505 cM - Gan *et al*. 2003), except for the map reported by Brondani *et.al* (2006 - 1133 cM). The difference in map lengths could be explained by the different mapping software used for linkage analysis. The maps reported previously were mostly constructed using MAPMAKER (MM, Lander *et al*. 1987), whereas JoinMap (Van Ooijen 2006) was used in this study. The multilocus likelihood method used by MM assumes the absence of crossover interference, while JoinMap accounts for a level of interference even though both programmes use the Kosambi function. This difference was also observed in other crop plants (Vuylsteke 1999, Liebhard *et al*. 2003, Hong *et al*. 2008). Due to these differences in estimation, JoinMap produces shorter maps than MM (Stam 1993, Vuylsteke 1999, Liebhard *et al*. 2003, Hong *et al*. 2008), especially when large numbers of markers are mapped. The *E. urophylla* parental linkage map reported by Brondani *et al*. (2006) was constructed with MM, but had low genome coverage, which explains the smaller map length. The two F1 hybrid maps (1021 cM and 1067 cM) were intermediate in size compared to the pure-species maps, despite higher numbers of segregating markers. This suggests that (paternal) recombination rates were overall very similar in the F1 hybrid and the pure-species parents, although local differences in recombination rates were apparent in the comparative maps of the F1 hybrid and the backcross parents (Figure S1).

For a comparison of genome coverage achieved in different studies, marker density and distribution should be considered. Past DArT mapping studies in plants (Wenzl *et al*. 2006, Tinker *et al*. 2009) suggested that DArT markers have a reasonably uniform genomic distribution. We observed apparent clustering of DArT markers in several linkage groups of the parental maps (Figure S1) and the consensus map (Figure 2). In addition, more than 25% of the DArT markers in the consensus map co-segregated perfectly with one or more other markers. This may simply be a feature of the large number of

markers mapped in this study, which would by chance lead to higher marker density in some regions of the map. However, some genomic regions may indeed be more polymorphic than others, especially in the F1 hybrid genome where regions that are rapidly diverging between the parental species could give rise to higher marker density in the F1 hybrid maps than the pure-species maps. Clustering of DArT markers has also been reported in mapping studies in wheat (Akbari *et al*. 2006, Semagn *et al*. 2006), barley (Wenzl *et al*.2004) and oat (Tinker *et al*.2009) and may be the result of reduced recombination in regions such as centromeres or regions with an excess of repeats (Vuylsteke *et al*. 1999, Young *et al*. 1999, Van Os *et al*. 2006). Despite the apparent clustering and redundancy of many DArT markers, the average marker interval (Table 1) in our maps was smaller than that of previous *Eucalyptus* genetic maps (Marques *et al*. 1998, Myburg *et al*. 2003, Brondani *et al*. 2006). Only four map intervals greater than 10 cM were observed for the *E. grandis* and *E. urophylla* BC parent maps. The consensus map had no intervals larger than 10 cM and only 10 intervals ranging 5 cM to 10 cM, with the largest gap (9.6 cM) on the distal end of LG5 (Figure 2). It is known that DArT genomic representations obtained with *PstI* reflect the methylation status of the genomic DNA and produce markers preferentially situated in hypomethylated, gene rich regions (van Os *et al*. 2006). Therefore, regions with lower marker density may be heterochromatin rich, or simply regions with lower genetic variability. Nevertheless, the high genome coverage achieved (c > 99.9% at 5 cM) makes these maps particularly useful for genome-assisted breeding.

In *Eucalyptus,* segregation distortion is normally higher in interspecific crosses (Grattapaglia *et al.* 1994, Verhaegen and Plomion 1996, Marques *et al*. 1998, Myburg *et al*. 2003) than in intraspecific crosses (Byrne *et al*. 1995, Thamarus *et al.* 2002). The observed segregation distortion in eucalypts is most likely caused by linkage between genetic markers and genes with recessive deleterious alleles or by hybrid incompatibility (Potts and Wiltshire 1997). Markers with significant deviation from the expected Mendelian ratios occurred throughout the F1 hybrid and BC parent maps (Table S2) suggesting the presence of multiple segregation distorting loci (SDLs) as previously reported for *Eucalyptus* (Myburg *et*

al. 2004). Approximately the same proportion of DArT markers were distorted in the two backcross parents than in the F1 hybrid which suggests that genetic factors affecting hybrid fitness may also be segregating in the two pure-species parents. This may be a feature of F2 pseudo-backcrosses where the two alleles segregating from the backcross parent can exhibit different (positive or negative) heterospecific interactions with the alleles segregating from the F1 hybrid (Myburg et al. 2004). The distorted markers often occurred as clusters (>10 markers/5 cM) or in some cases spanning the entire chromosome in the parental and hybrid maps (LG5). Clustering of loci showing segregation distortion has been reported before in *Eucalyptus* (Byrne *et al*. 1995, Verhaegen and Plomion 1996, Marques *et al*. 1998, Bundock *et al*. 2000, Brondani *et al.* 2006). These regions may contain genetic factors influencing the viability of F1 gametes, or fitness of F2 progeny (Lorieux *et al*. 2000, Cervera *et al*. 2001, Myburg *et al*. 2004, Liebhard *et al*. 2003, Bundock *et al*. 2000).

The reliability of consensus mapping was questioned by Beavis and Grant (1991) who cited the variability of recombination frequency in different populations or crosses. However, where marker order is conserved among individual maps, consensus mapping is a robust approach (Lespinasse *et al*. 2000). Only a small number of markers exhibited a change in order in the consensus map relative to the parental maps, specifically in LG1 and LG7 of the *E. grandis* BC parent (Figure S1, Figure S2). Changes in marker order during map integration have been reported in *Eucalyptus* (Brondani *et al.* 2006) and other species (Doligez *et al.* 2006, Lombard and Delourme 2001, Mace *et al*. 2009) and could be caused by heterogeneity in recombination, incorrect ordering in individual parental maps and missing or poor quality marker data (Lombard and Delourme 2001). Despite the fact that the markers in the parental maps were ordered with high statistical support and the order of markers in the consensus map was highly similar to that in the *E. grandis* genome scaffolds (Figure S3) users of this map should be aware of the mentioned limitations of consensus mapping when interpreting consensus marker order, as well as total map length and spacing (Table 3).

The high marker density of the consensus map allowed selection of more than 150 pairs of markers that are both located on the same *de novo* assembled *E. grandis* genome scaffold. The ratio of physical to genetic distance (Figure S4) will determine the feasibility of future map-based cloning efforts in *Eucalyptus*. The average physical distance observed per centiMorgan (633 kb/cM) was substantially larger than that reported before in *Populus* (200 kb/cM, Yin *et al*. 2004), and rice (244 kb/cM, Chen *et al*. 2002). The first JGI annotation of the *E. grandis* genome (V1.0, www.phytozome.net) predicted a total of 41,204 protein-coding loci in the 11 chromosome assemblies, which correspond to the 11 linkage groups in our map (Figure S3). Based on the cumulative size of the 11 chromosome assemblies (605.8 Mbp), the average gene density in the *E. grandis* genome is predicted to be 68 per Mbp. This is lower than the gene density in *Arabidopsis* (218 per Mbp, www.phytozome.net) and *Populus* (100 per Mbp, www.phytozome.net). However, considering genetic distance, the gene density in *Eucalyptus,* 43 per cM (633 kb), is predicted to be the same as in *Populus* (43.6 per cM, 200kb). This means that a QTL interval of 20 cM would on average contain approximately 860 genes. In this context, genetical genomics (eQTL mapping) approaches (e.g. Kirst *et al*. 2004) would be valuable to further dissect candidate genes underlying trait QTLs. The high-density of the genetic maps that can be achieved with the *Eucalyptus* DArT array (up to an average spacing of 0.48 cM, Table 3) will ensure many (~ 40) sequence-anchored marker loci per QTL (assuming a confidence interval of 20 cM), which will increase the accuracy of QTL tagging. A total of 1836 DArT markers were placed in the genome sequence assembly (Figure S3). These markers and additional markers developed from the genome sequence in tagged QTL intervals will support fine-scale mapping of QTL regions of interest. Most QTLs underlying economically important traits in *Eucalyptus* have not been characterized at this scale. We expect that the sequence-anchored genetic maps reported here and others to follow will accelerate the tagging of QTLs and cloning of positional candidate genes, and enhance *Eucalyptus* breeding through marker-assisted selection.

Acknowledgements

The authors are grateful to Sappi (South Africa) for making the crosses, maintaining the mapping pedigree and providing plant materials for DNA isolation and, in particular, wish to acknowledge Geoff Galloway (Sappi) for assistance in this regards. Minique de Castro (University of Pretoria) provided technical assistance with microsatellite marker analysis and Nanette Coetzer (University of Pretoria) assisted with bioinformatics and statistical analyses. Diversity Arrays Technology Pty Ltd (DArT P/L, Canberra, Australia) are acknowledged for excellent technical assistance and service. Jeremy Schmutz and Jerry Jenkins (HudsonAlpha Genome Sequencing Center, Huntsville, AL) kindly provided genomic scaffold positions for the sequenced DArT markers in the draft *E. grandis* genome sequence (DOE-JGI). This work was funded by Sappi through the Forest Molecular Genetics Programme and by 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

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A (2006) Diversity arrays technology (DArT) for high throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113:1409–1420

Beavis WD, Grant D (1991) A linkage map based on information from four F2 populations of maize (*Zea mays* L.). Theor Appl Genet 82: 632-644

Bison O, Ramalho MAP, Rezende GDSP, Aguiar AM, De Resende MDV (2006) Comparison between open pollinated progenies and hybrid performance in *Eucalyptus grandis* and *Eucalyptus urophylla*. Silv Gen 55:192–196

Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. Theor Appl Genet 97: 816–827

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

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 Bioliogy 6: 20

Bundock PC, Hayden M, Vaillancourt RE (2000) Linkage maps of *Eucalyptus globulus* using RAPD and microsatellite markers. Silv Genet 49:223–232

Byrne M, Murrell JC, Allen B, Moran GF (1995) An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers. Theor Appl Genet 91:869–875

Byrne M, Marques-Garcia MI, Uren T, Smith DS, Moran GF (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus.* Aust J Bot 44:331–341

Campinhos EJ, Ikemori YK (1989) Selection and management of the basic population *Eucalyptus grandis* and *E. urophylla* established at Aracruz for the long term breeding programme. In: Gibson GI, Griffin AR, Matheson AC (eds) Breeding Tropical Trees: Population Structure and Genetic Improvement Strategies in Clonal and Seedling Forestry. [Proc. IUFRO Conference]. Pattaya, Thailand. Oxford Forestry Institute, Oxford pp. 169–175

Cervera MT, Storme V, Ivens B, Gusmao J, Liu BH, Hostyn V, Van Slycken M, Van Montague M, Boerjan W (2001) Dense genetic linkage maps of three *Populus* species (*Populus deltoides, P. nigra and P. trichocarpa*) based on AFLP and microsatellite markers. Genetics 158:787–809

Chen M, Presting G, Barbazuk W, Goicoechea J, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S, Higingbottom S, Phimphilai J, Phimphilai D, Thurmond S, Gaudette B, Li P, Liu J, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman M, Tomkins J, Luo M, Bancroft I, Salse J, Regad F, Mohapatra T, Singh N, Tyagi A, Soderlundn C, Dean R, Wing R (2002) An integrated physical and genetic map of the rice genome. The Plant Cell 14:537-45

Coetzer N, Gazendam I, Oelofse D, Berger DK (2010) SSHscreen and SSHdb, generic software for microarray based gene discovery: application to the stress response in cowpea. *Plant Methods* 6:10

Creux NM, Martin R, Berger DK, Myburg AA (2009) Comparative analysis of orthologous cellulose synthase promoters from *Arabidopsis*, *Populus* and *Eucalyptus*: evidence of conserved regulatory elements in angiosperms. New Phytologist 179:722-737

Darrow WK (1995) A study of eucalypt clonal and hybrid material planted in the sub-humid zones of Kwazulu-Natal: Results of the first six years. Inst Commer For Res Bull Ser 2(95):1–23

Doligez A, Adam-Blondon AF, Cipriani G, Di Gaspero G, Laucou V, Merdinoglu D, Meredith CP, Riaz S, Roux C, This P (2006) An integrated SSR map of grapevine based on five mapping populations. Theor Appl Genet 113(3):369–382

Eldridge K, Davidson J, Harwood C, Van Wyk G (1993) Eucalypt Domestication and Breeding. Clarendon Press, Oxford.

Freeman J, 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

Gan S, Shi J, Li M, Wu K, Wu J, Bai J (2003) Moderate-density molecular maps of *Eucalyptus urophylla* S.T. Blake and *E. tereticornis* Smith genomes based on RAPD markers. Genetica 118:59–67

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudotestcross: mapping strategy and RAPD markers. Genetics 137:1121–1137

Grattapaglia D, Bertolucci FLG, Penchel R, Sederoff R (1996) Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. Genet 144:1205–1214

Griffin AR, Burgess IP, Wolf L (1988) Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'Herit: A Review. Aust J Bot. 36:41–66

Hong Y, Liang X, Chen X, Liu H, Zhou G, Li S, Wen S (2008) Construction of genetic linkage map based on SSR markers in peanut (Arachis hypogaea L.) Agricultural Sciences in China 7:915-9

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

Khurana DK, Khosla PK (1998) Hybrids in forest tree improvement. In: Mandal AK, Gibson GL (eds) Forest genetics and tree breeding. CBS Publishers and distributors, New Delhi, India pp. 86–102

Kirst M, Myburg AA, De Leon JP, Kirst ME, Scott J, Sederoff R (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of *Eucalyptus*. Plant Physiol 135:2368–2378

Kosambi D (1944) The estimation of map distances from recombination values. Ann Eugen 12:172-175

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lange K, Boehnke M (1982) How many polymorphic genes will it take to span the human genome? Am J Hum Genet 34:842–845

Lespinasse D, Rodier-Goud M, Grivet L, Leconte A, Legnate H, Seguin M (2000) A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite, and isozyme markers. Theor Appl Genet 100(1):127–138

Liebhard R, Koller B, Gianfranceschi L, Gessler C (2003) Creating a saturated reference map for the apple (*Malus* × *domestica* Borkh.) genome. Theor Appl Genet 106:1497-1508

Lombard V, Delourme R (2001) A consensus linkage map for rapeseed (*Brassica napus* L.): construction and integration of three individual maps from DH populations. Theor Appl Genet 103:491–507

Lorieux M, Ndjiondjop MN, Ghesquiere A (2000) A first interspecific *Oryza sativa x Oryza glaberrima* microsatellite based genetic linkage map. Theor Appl Genet 100:593–601

Lyttle TW (1991) Segregation distorters. Annu Rev Genet 25:511–557

Mace ES, Rami JF, Bouchet S, Klein PE, Klein RR, Kilian A, Wenzl P, Xia L, Halloran K, Jordan DR (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. BMC Plant Biology 9:13

Marques CM, Araujo JA, Ferreira JG, Whetten R, O'Malley DM, Liu BH, Sederoff R (1998) AFLP genetic maps of *Eucalyptus globulus* and *Eucalyptus tereticornis*. Theor Appl Genet 96:727–737

Myburg AA, Griffin AR, Sederoff RR, Whetten RW (2003) Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. Theor Appl Genet 107:1028–1042

Myburg AA, Vogl C, Griffin AR, Sederoff RR, Whetten RW (2004) Genetics of postzygotic isolation in *Eucalyptus*: whole-genome analysis of barriers to introgression in a wide interspecific cross of *Eucalyptus grandis* and *E. globulus*. Genetics 166:1405–1418

Potts BM, Wiltshire RJE (1997) Eucalypt genetics and genecology. In *Eucalypt Ecology: Individuals to cosystems* Edited by: Wiliams J, Woinarski J. Cambridge, Cambridge University Press

Potts BM, Dungey HS (2004) Interspecific hybridization of eucalypts: key issues for breeders and geneticists. New For. 27: 115–138

Rezende GDSP, de Resende MDV (2000) Dominance effects in *Eucalyptus grandis*, *Eucalyptus urophylla* and hybrids. In: Dungey H.S., Dieters M.J. and Nikles D.G. (eds), Hybrid Breeding and Genetics of Forest Trees. Proceedings of QFRI/CRC-SPF Symposium, 9–14th April 2000 Noosa, Queensland, Australia. Department of Primary Industries, Brisbane, pp. 93–100

Sansaloni CP, Petroli CD, Carling J, Hudson J, Steane DA, Myburg AA, Grattapaglia D, Vaillancourt RE, 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, Skinnes H, Maroy AG, Tarkegne Y, William M (2006) Distribution of DArT, AFLP and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. Genome 49:545–555

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 5:739–744

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

Thamarus K, Groom K, Murrell J, Byrne M, Moran G (2002) A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre and floral traits. Theor Appl Genet 104:379–387

Thumma BR, Southerton SG, Bell JC, Owen JV, Henery ML, Moran GF (2010) Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens*. Tree Genetics & Genomes 6:305–317

Tinker NA, Kilian A, Wight CP, Heller-Uszynska K, Wenzl P, Rines HW, Bjørnstad A, Howarth CJ, Jannink J, Anderson JM, Rossnagel BG, Stuthman DD, Sorrells ME, Jackson EW, Tuvesson S, Kolb

FL, Olsson O, Federizzi LC, Marty LC, Ohm HW, Molnar SJ, Scoles GJ, Eckstein PE, Bonman JM, Ceplitis A, Langdon T (2009) New DArT markers for oat provide enhanced map coverage and global germplasm characterization. BMC Genomics 10:39

Vaillancourt RE, Potts BM, Manson A, Reid JB (1994) Detection of QTLs in a *Eucalyptus gunni* x *E. globulus* F2 using a RAPD linkage map. In: Proc Int Wood Biotechnol Symp, 31 Aug-1 Sept, Hokutopia (Convention Hall) Tokyo, pp 63–70

Van Ooijen JW (2006) JoinMap®. Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands

Van Os HV, Andrzejewski S, Bakker E, Barrena I, Bryan GJ, Caromel B, Ghareeb B, Isidore E, De Jong W, Van Koert P, Lefebvre V, Milbourne D, Ritter E, Van der Voort JNAMR, Rousselle-Bourgeois F, Van Vliet J, Waugh R, Visser RGF, Bakker J, Van Eck HJ (2006) Construction of a 10,000 marker ultra-dense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genome-wide physical map. Genetics 173: 1075–1087

Varshney RK, Tuberosa R (2007) Genomics-assisted crop improvement: An overview. In: Varshney RK, Tuberosa R (eds) Genomics-assisted crop improvement, vol 1: genomics approaches and platforms. Springer, Dordrecht, the Netherlands, pp 1–12

Verhaegen D, Plomion C (1996) Genetic mapping in *Eucalyptus urophylla* and *E. grandis* using RAPD markers. Genome 39:1051–1061

Verhaegen D, Plomion C, Gion JM, Poitel M, Costa P, Kremer A (1997) Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers. 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. Theor Appl Genet 95:597–608

Vigneron P, Bouvet J (2000) Eucalypt hybrid breeding in Congo. In: Dungey HS, Dieters MJ, Nikles DG (eds) Hybrid Breeding and Genetics of Forest Trees. Proceedings of QFRI/CRC-SPF Symposium, 9–

14th April 2000 Noosa, Queensland, Australia. Department of Primary Industries, Brisbane, pp. 14–26

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93:77–78

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Luebberstedt T, Xia XC, Stam P, Zabeau M, Kuiper M (1999) Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet 99:921-935

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Jaroslava O, Cakir M, Poulsen D, Wang J, Raman R, Smith KP, Muehlbauer GJ, Chalmers KJ, Kleinhofs A, Huttner E, Kilian A (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. *BMC Genomics* 7:206-228

Wingfield MJ, Swart WJ, Abear BJ (1989) First record of Cryphonectria canker of *Eucalyptus* in South Africa. *Phytophylactica* 21: 311-313

Wittenberg AHJ, Van der Lee T, Cayla C, Kilian A, Visser RGF, Schouten HJ (2005) Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. Mol Gen Genomics 274:30-39

Wright JA (1997) A review of the worldwide activities in tree improvement for *Eucalyptus grandis*, *Eucalyptus urophylla* and the hybrid urograndis. In: White T, Huber D, Powell G (eds) 24th Biennial Southern Forest Tree Improvement Conference. Orlando, Florida, USA. Southern Forest Tree Improvement Committee, Orlando, pp. 96–102

Yin TM, DiFazio SP, Gunter LE, Jawdy SS, Boerjan W, Tuskan GA (2004) Genetic and physical mapping of *Melampsora* rust resistance genes in *Populus* and characterization of linkage disequilibrium and flanking genomic sequence. New Phytologist 164: 95–105

Young WP, Schupp JM, Keim P (1999) DNA methylation and AFLP marker distribution in the soybean genome. Theor Appl Genet 99:785–792

Tables

Table 1. Summary of the 2617 DArT markers that segregated and were used for linkage analysis in the F2 backcross pedigree. Of these, 2440 markers were included in the final parental linkage maps (Table 2) and 2229 in the consensus linkage map (Table 3).

Segregation type	***E. grandis* BC family**	***E. urophylla* BC family**
F1 hybrid	655 (37.5%)	718 (40.8%)
Testcross markers (1:1)		
Backcross parent	429 (24.6%)	401 (22.8%)
Testcross markers (1:1)		
Both parents		
Intercross markers (3:1)	659 (37.8%)	638 (36.3%)
Total markers[a]	**1743**	**1757**

[a] A total of 883 markers (33.7%) were shared between the two backcross families bringing the total for the two families to 2617.

Table 2. Summary of DArT and microsatellite (SSR) markers mapped in each linkage group of the two backcross families

	***E. grandis* BC parent**				**F1 hybrid (*E. grandis* BC)**				**F1 hybrid (*E. urophylla* BC)**				***E. urophylla* BC parent**			
Linkage group	No of DArT markers	No of SSR markers	Size in cM	Mean distance between markers	No of DArT markers	No of SSR markers	Size in cM	Mean distance between markers	No of DArT markers	No of SSR markers	Size in cM	Mean distance between markers	No of DArT markers	No of SSR markers	Size in cM	Mean distance between markers
1	70	5	89.9	1.19	86	7	87.0	0.93	90	7	92.3	0.95	61	6	85.1	1.27
2	103	6	92.3	0.84	146	8	102.2	0.66	139	8	106.0	0.72	101	6	107.1	1.00
3	103	4	71.6	0.68	112	4	98.4	0.84	133	4	108.6	0.78	99	5	113.4	1.09
4	75	2	69.4	0.90	88	5	68.9	0.74	73	6	69.8	0.88	54	4	87.4	1.50
5	107	2	82.0	0.75	132	2	103.8	0.77	107	4	90.8	0.81	86	4	98.5	1.09
6	102	3	90.5	0.78	116	4	107.4	0.89	119	3	121.6	0.99	98	3	124.8	1.23
7	86	3	82.0	0.93	85	4	81.2	0.91	109	6	78.8	0.68	78	5	87.1	1.04
8	118	3	98.7	0.81	149	7	109.8	0.70	167	9	115.3	0.64	126	6	123.1	0.93
9	80	3	79.6	0.95	97	6	71.5	0.69	119	6	80.5	0.64	59	4	79.0	1.25
10	41	1	82.5	1.96	75	1	95.2	1.25	81	2	95.0	1.14	61	1	93.7	1.51
11	72	2	86.2	1.16	81	2	95.8	1.15	115	4	108.6	0.92	89	2	108.2	1.28
Total	**957**	**34**	**924.7**	**0.99**	**1167**	**50**	**1021.2**	**0.86**	**1252**	**59**	**1067.3**	**0.83**	**912**	**46**	**1107.3**	**1.19**

Table 3. Summary of markers integrated into the consensus map for the interspecific F2 backcross pedigree of *E. grandis* x *E. urophylla*

Consensus linkage group	No of DArT markers	No of microsatellite markers	Map length (cM)	Mean marker spacing (cM)
1	173	7	88.8	0.49
2	228	7	102.1	0.43
3	251	6	105.5	0.41
4	157	6	79.8	0.48
5	218	2	110.4	0.50
6	232	4	136.9	0.58
7	163	6	83.5	0.49
8	263	11	119.1	0.45
9	203	7	88.5	0.42
10	155	1	97.7	0.62
11	186	4	95.3	0.50
Total	**2229**	**61**	**1107.6**	**0.48**

Figures

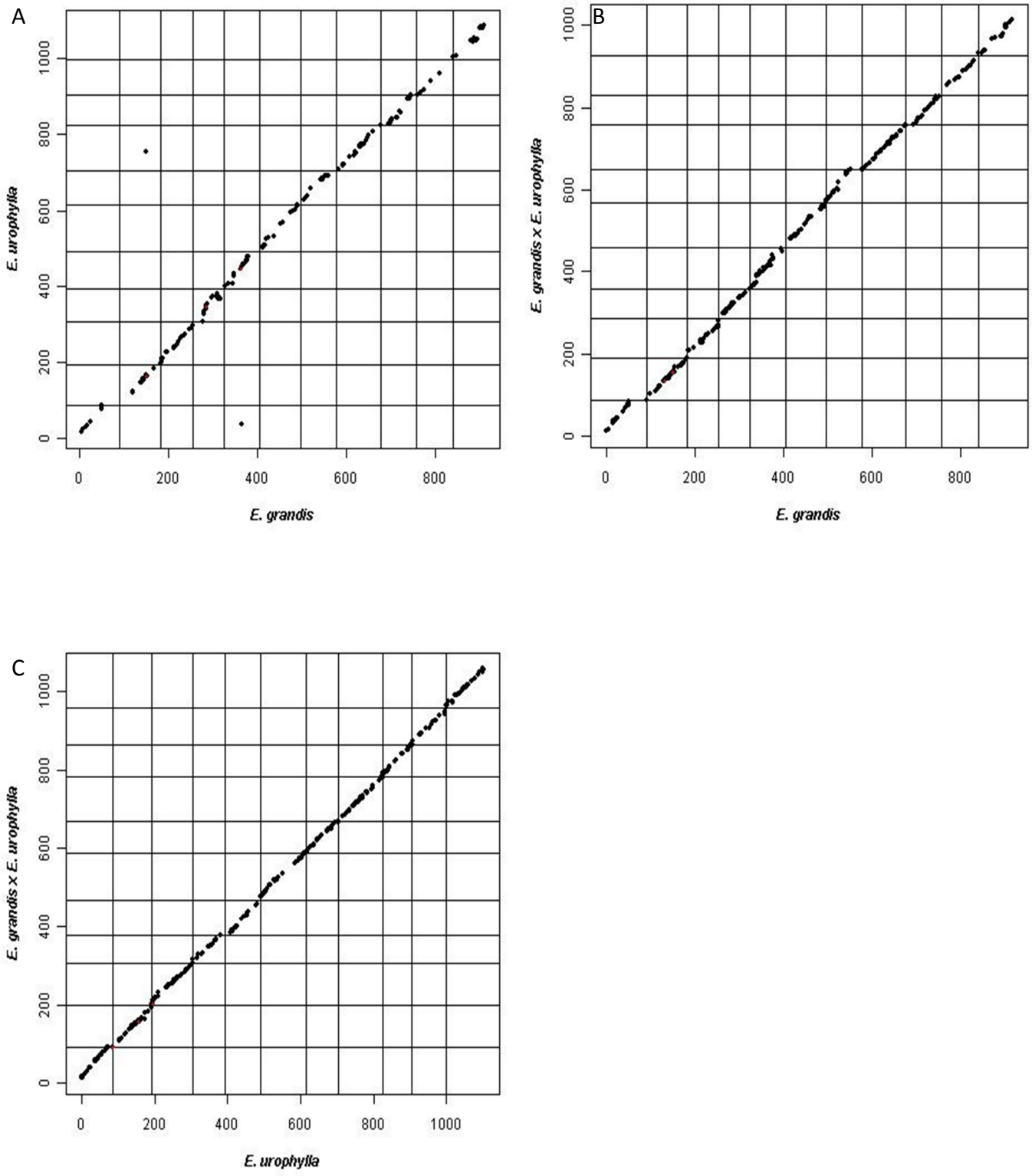

Figure 1

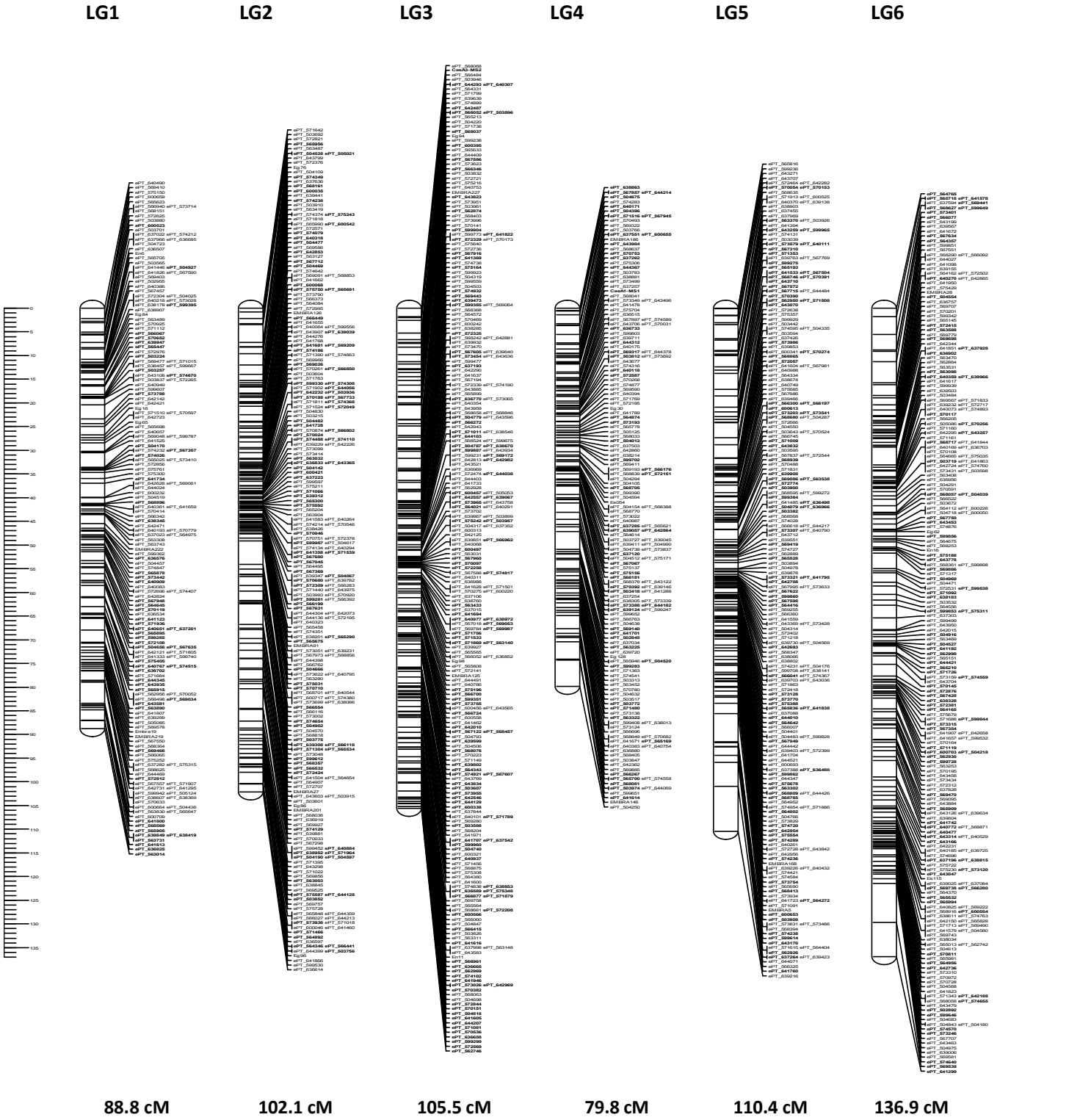
LG1
LG2
LG3
LG4
LG5
LG6
88.8 cM
102.1 cM
105.5 cM
79.8 cM
110.4 cM
136.9 cM

Figure 2

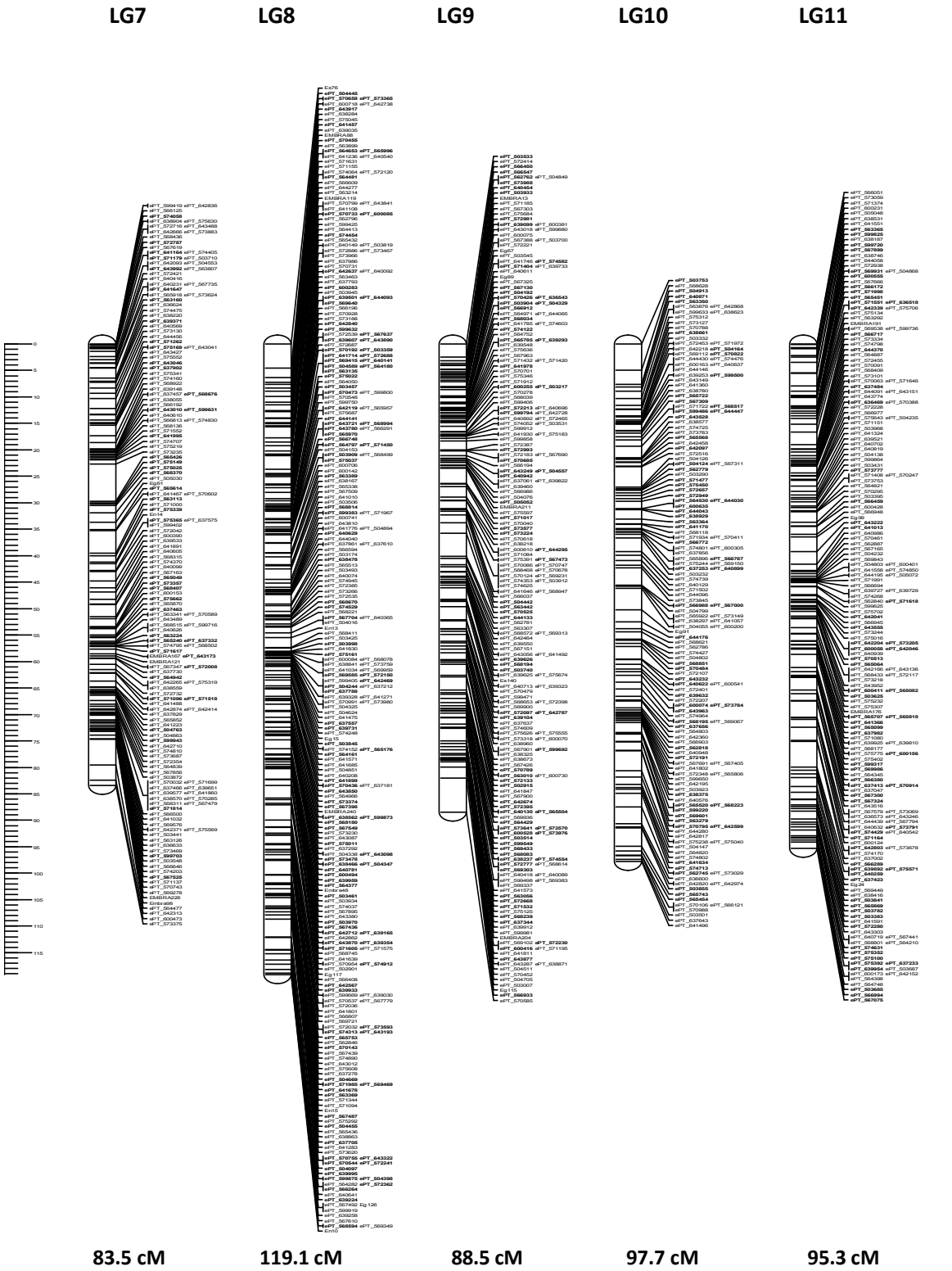

Figure 2 (cont.)