DOI: 10.1007/s10535-014-0459-9 BIOLOGIA PLANTARUM 59 (2): 389-392, 2015

BRIEF COMMUNICATION

Marker-trait associations for survival, growth, and flowering components in *Eucalyptus cladocalyx* under arid conditions

P. BALLESTA[1], F. MORA[2]*, E. RUIZ[3], and R. CONTRERAS-SOTO[1,4]

Faculty of Forest Sciences, University of Concepción, Victoria 631, Concepción, Chile[1]
Institute of Biological Sciences, University of Talca, 2 Norte 685, Talca, Chile[2]
Department of Botany, University of Concepción, Casilla 160-C, Concepción, Chile[3]
State University of Maringá, Department of Agronomy, Colombo 5790, 87020-900, Maringá, Paraná, Brazil[4]

Abstract

Understanding the basis of the genetic variations responsible for the complex traits found in *Eucalyptus cladocalyx* under arid environmental conditions is crucial for designing genetic architecture studies. Forty-five half-sib families from Australia were used to identify inter-simple sequence repeat (ISSR) markers that are associated with growth (height, diameter at breast height, and stem straightness), flowering traits (flowering intensity, flowering precocity, reproductive capacity, and late flowering) and tree survival under arid conditions in southern Atacama Desert, Chile. Each DNA pellet consisted of a pool of five trees from each family. ISSR markers were associated with all the traits studied and accounted for 9.8 to 23.4 % of the phenotypic variation. Several loci were associated with more than one trait. For example, $UBC_{810(450\text{-}500\ bp)}$, $ISO_{1(600\text{-}610\ bp)}$, and $TGT_{9(780\text{-}800\ bp)}$ were associated with three of the traits studied. These identified genomic regions may contribute to the increase of the efficiency of the conventional tree breeding program for *E. cladocalyx*.

Additional key words: inter-microsatellites, marker-assisted selection, structured populations, sugar gum.

Eucalyptus cladocalyx F. Muell is cultivated for conservation, timber, and honey production in arid and semi-arid regions (McDonald *et al*. 2003, Mora *et al*. 2009, Bush and Thumma 2013). For example, a breeding program was recently initiated in Australia because of its suitability for low rainfall regions where *E. cladocalyx* may be successfully cultivated for the production of timber, poles, and environmental remediation (Bush *et al*. 2011). Some genotypes of this species flower abundantly and for a long period of time, which are highly desirable traits in the honey production industry (Mora *et al*. 2009, Cané-Retamales *et al*. 2011). Various studies have shown that a significant amount of genetic variation is responsible for the complex quantitative traits, such as growth (Mora *et al*. 2009), flowering (Mora *et al*. 2009, Cané-Retamales *et al*. 2011), and timber properties (Bush *et al*. 2011).

Marker-trait association analysis has been proposed to identify polymorphisms involved in phenotypic variations in *Eucalyptus*, including growth traits (Freeman *et al*. 2009, Thumma *et al*. 2010), wood properties (Thamarus *et al.* 2004, Freeman *et al*. 2009), cellulose fiber properties (Thamarus *et al*. 2004), and flowering (Missiaggia *et al*. 2005). Association mapping or linkage disequilibrium (LD) attempts to account for the effects of population structure and incorporates the historical relationships between the phenotypes and molecular markers relative to the LD (Cardon and Bell 2001, Lohwasser *et al.* 2013). The LD method has also been used for various tree species, such as *Pinus taeda* (Gonzalez-Martinez *et al*. 2008), *Eucalyptus spp*. (Thumma *et al.* 2005), and *Populus tremuloides* (Kelleher *et al*. 2012).

Submitted 21 April 2014, *last revision* 7 July 2014, *accepted* 12 August 2014.
Abbreviations: DBH - diameter at breast height; EF - early flowering; FI - flowering intensity; HT - total height; ISSR - inter simple sequence repeat; LF - late flowering; PCoA - principal coordinates analysis; RC - reproductive capacity; ST - stem straightness; TS - survival.
Acknowledgements: This research was supported by the Chilean National Science and Technology Research Fund (FONDECYT), project No. 1130306. The authors thank Mr. Augusto Gomes for providing the samples and evaluating the progeny test.
* Corresponding author; fax: (+56) 71 2200277, e-mail: morapoblete@gmail.com

Eucalyptus cladocalyx F. Muell progeny tests conducted in the arid region of Chile have shown that there is sufficient genetic variability to a low-input breeding program (Mora *et al*. 2009, Cané-Retamales *et al.* 2011, Vargas-Reeve *et al*. 2013). The moderate to high heritability suggests that there is a high probability of finding genomic regions that are related to the phenotypic variations. Thus, the aim of this study was to identify the inter-simple sequence repeat (ISSR) markers that are associated with complex phenotypic traits in *E. cladocalyx* families.

The marker-trait association analysis was conducted in a progeny test of *E. cladocalyx* in the Province of Choapa, Chile, according to Cané-Retamales *et al*. (2011). This trial included 47 families from 5 Australian regions (Table 1 Suppl.). The following eight quantitative traits were assessed according to Mora *et al*. (2009): 1 - total height (HT), 2 - diameter at breast height (DBH), and 3 - stem straightness (ST) of 10-year-old trees; 4 - flowering intensity (FI), 5 - late flowering (LF), 6 - reproductive capacity (RC), 7 - tree survival (TS) of 12-year-old trees, and 8 - early flowering 3-year-old trees (EF). The variables ST, FI, EF, LF, RC, and TS were considered as discrete variables. Stem straightness was classified using the four categories described by Vargas-Reeve *et al.* (2013). FI was classified using the method described by Cané-Retamales *et al.* (2011). LF was measured at the end of the 2013 blooming season (May-June) and scored as binary variable: the value 0 was assigned if the tree did not flower by the end of the season, and the value 1 was assigned if it did. RC was assessed on a binary scale: the value 0 was assigned if the tree never flowered during a 12-year period, and the value 1 was assigned if it did. Finally, TS was also assessed on a binary scale as described by Mora and Serra (2014): the value 0 indicated death, and the value 1 indicated the survival of the tree to May 2013.

Leaves were collected from each tree of Australian origin (n = 5 trees per family × 47 families). DNA was extracted using CTAB according to a modified version of the method described by Doyle and Doyle (1987). Each DNA pellet consisted of a pool of five trees from each family. Family bulks were prepared by pooling aliquots, containing equivalent amounts of total DNA, according to Barakat *et al.* (2013). Eight inter-simple sequence repeat (ISSR) primers were used: $(RCA)_7$, $T(GT)_9$, $TA(CAG)_4$, $RA(GCT)_6$, GA_8, UBC_{810}, ISO_1, and ISO_2 (Balasaravanan *et al*. 2005, Okun *et al*. 2008, Chezhian *et al*. 2010). The reactions occurred in a total volume of 0.02 cm^3 containing 20 - 50 ng of DNA template, 1× reaction buffer (5× *Green Go Taq Flexi* buffer), 0.6 µM primers, 200 µM dNTPs, 1.5 mM $MgCl_2$, and 0.625 U cm^{-3} *Taq Pol* (all chemicals from *Promega*, Madison, USA). Two different programs in an *Axygen Maxygen* PCR *Therm-1000* thermocycler were used for amplification. The following amplification program was used for the primers $(RCA)_7$, $T(GT)_9$, $TA(CAG)_4$, $RA(GCT)_6$, GA_8, and UBC_{810}: an initial denaturation at 94 °C for 3 min, 35 cycles of denaturation at 94 °C for 30 s, annealing at a specific temperature for each primer for 30 s, an extension at 72 °C for 1 min, and a final extension at 72 °C for 10 min. The following program was used for the primers ISO_1 and ISO_2: an initial denaturation at 94 °C for 7 min, 45 cycles of denaturation at 94 °C for 30 s, annealing at a specific temperature for each primer for 45 s, an extension at 72 °C for 1 min, and a final extension at 72 °C for 7 min.

The program *ARLEQUIN* 3.5 (Excoffier and Lischer 2010) was used to discriminate the ISSR loci that did not fulfill the criteria of neutrality (outliers) according to Excoffier *et al*. (2009). Neutral loci were then analyzed by the Bayesian procedure using the program *STRUCTURE 2.2* (Pritchard *et al*. 2000). Based on a no-admixture and the independent allele frequency population model, we hypothesized that the number of clusters (K) would vary between 1 and 5, considering the putative Australian subpopulations. The K value was calculated from 20 independent repetitions for each possible K value, a burn-in value of 500 and the number of Gibbs chains estimated at 20 000 using a variation of the Markov Chain Monte Carlo method. The optimal K value was calculated using a method described by Evanno *et al*. (2005). Principal coordinates analysis (PCoA) was also carried out to confirm the population genetic structure using *GENALEX* (Peakall and Smouse 2006).

Linkage disequilibrium (LD) was estimated using *ARLEQUIN 3.5* (Excoffier and Lischer 2010) by means of the pairwise linkage test with 100 000 steps in Markov chain and 1 000 steps of dememorisation. The Bonferroni correction test was performed to correct for multiple testing. Pairwise LD was estimated for the entire set of eucalypt samples as well as for each cluster from the structure analysis. The random family effect was predicted using the best linear unbiased predictor (BLUP) in *SAS* (*SAS Institute Inc*. 2007) and were used as phenotypic inputs for the subsequent marker-trait association analysis. Marker-trait associations were carried out by using a unified mixed-model method (Yu *et al*. 2006) with terms to account for genetic relatedness due to the historical population structure. This analysis was performed using *TASSEL v. 3.0.137* (Bradbury *et al*. 2007). Associations were initially determined at the nominal $P \leq 0.05$ level. Additionally, for a multiple comparison adjustment, false discovery rate (FDR)-adjusted P values were calculated using *PROC MULTTEST* in *SAS* 2007.

The ISSR primers generated a total of 93 loci that varied between 100 and 2 500 bp. The genetic structure analysis separated the *E. cladocalyx* subpopulations sampled into two potentially homogenous genetic groups (Fig. 1 Suppl.). Cluster 1 contained 12 families, over 46 % of which were collected from Flinders Chase. Cluster 2 contained 33 families (73.3 %) which are from Mount Remarkable (12/33), Cowell (9/33), Marble Range (4/33), Wirrabara (5/33), and Flinders Chase (3/33) (Fig. 1). In concordance with *STRUCTURE* results, the PCoA analysis revealed the presence of two principal

groups (Fig. 2) in which most families from Flinders Chase were assigned in the same cluster. This finding is partially consistent with a study conducted by Bush and Thumma (2013) as they did not study the Eyre Peninsula provenances. The LD analysis in each cluster revealed 44 (about 5 %: 44/903) and 23 (about 2 %: 23/1128) highly correlated loci pairs (significant LD, after a Bonferroni correction) within the clusters 1 and 2, respectively. The r^2 values ranged from 0.08 and 1.0 within cluster 1, and from 0.04 to 0.47 within cluster 2. The number of loci combinations in LD was low, which is expected for self-incompatibility species (Ellis and Sedgley 1992), although this result is not common in populations with high levels of inbreeding (McDonald *et al*. 2003). In accordance with previous studies (Bush and Thumma 2013), the provenance Flinders Chase (Kangaroo Island) had a high number of allelic combinations in LD. Bush and Thumma (2013) found that LD decayed rapidly in an *E. cladocalyx* breeding population, and was more extensive in the Kangaroo Island subpopulations; a result consistent with elevated levels of inbreeding observed in this subpopulation.

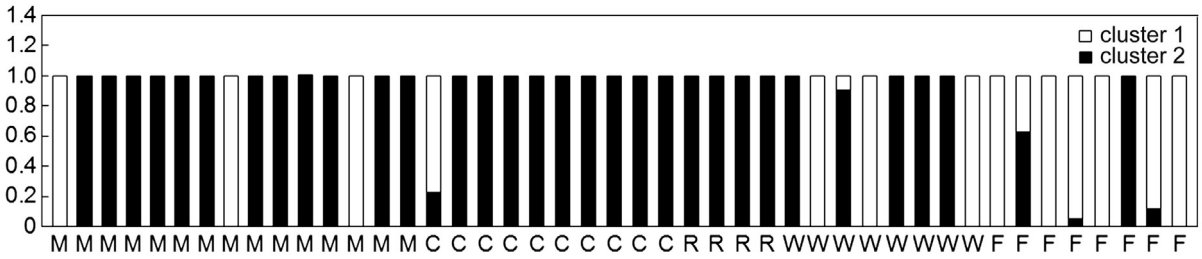

Fig. 1. A bar plot showing the probability of membership for 5 provenances and 45 families assessed using ISSR markers. Each family is represented by a vertical column. *White* and *black bars* correspond to cluster 1 and cluster 2, respectively. The abbreviations M, C, R, W, and F represent Mount Remarkable, Cowell, Marble Range, Wirrabara, and Flinders Chase, respectively.

A total of 18 loci amplified were associated with at least one trait studied (Table 1) and represented 19.4 % of the loci tested. The LD analysis between these loci revealed that they were in linkage equilibrium. Marker-trait associations accounted for 9.8 to 23.4 % of the phenotypic variation. Three loci were associated with total height, six loci were associated with diameter (DHB), and five loci were associated with stem straightness. Two loci were associated with flowering intensity, one locus was associated with late flowering, one locus was associated with early flowering, and two loci were associated with reproductive capacity. Flowering intensity was associated with two loci, one of them was also associated with reproductive capacity ($TGT_{9(270\text{-}280\ bp)}$). Cané-Retamales *et al*. (2011) found a positive correlation between the intensity and precocity, which highlights the positive effect of selecting young trees. This result suggests that the loci responsible for a greater flowering intensity may also be responsible for the reproductive initiation state of the tree.

Table 1. Summary of marker-trait associations detected for growth and flowering traits in *Eucalyptus cladocalyx*.

Trait	Associations	Phenotypic variation [%]
Diameter at breast height	6	9.8 - 23.4
Early flowering	1	12.0
Total height	3	12.0 - 21.9
Flowering intensity	2	11.5 - 11.9
Late flowering	1	19.5
Reproductive capacity	2	10.3 - 11.6
Stem straightness	5	10.0 - 18.7
Tree survival	6	10.2 - 20.2

Three loci were associated with total height and diameter at breast height ($ISO_{1(600\text{-}610)}$, $TGT_{9(780\text{-}800)}$, $UBC_{810(450\text{-}500)}$). This result was expected because many studies have shown a positive genetic correlation between both variables in *Eucalyptus*; *e.g*., Vargas-Reeve *et al.* (2013) in *E. cladocalyx*, which suggests that these loci could have pleiotropic effects. Only one locus was associated with the stem straightness and also with the diameter ($ISO_{2(300)}$). Callister *et al.* (2008) showed that there is a positive correlation between growth and straightness in the same species. Cappa *et al.* (2010) described an important localization effect that establishes a correlation between growth and straightness in *Eucalyptus viminalis*. Therefore, the association between these two variables remains unclear. Although the majority of the loci were associated with one trait, the markers $UBC_{810(450\text{-}500)}$, $TGT_{9(780\text{-}800)}$, and $ISO_{1(600\text{-}610)}$ were associated with three traits: $UBC_{810(450\text{-}500)}$ was associated with DHB, HT, and TS; $T(GT)_{9(780\text{-}800)}$ was associated with HT, DHB, and TS; and $ISO_{1(600\text{-}610)}$ was associated with RC, HT, and DHB.

In this study, six loci were associated with tree survival, and one of them was also associated with early flowering. The flowering process is particularly sensitive to abiotic stress; therefore, plants will modify their flowering schedule in response to stress. Several criteria have been identified that measure water deficit tolerance/resistance capacity. Rönnberg-Wästljung *et al.* (2005) identified four regions that are related to efficient water use in *Salix dasyclados* × *Salix viminalis* and account for 11 % of the total variation. Tschaplinski *et al.* (2006) identified 12 genomic regions that are associated

with an efficient water use in *Populus*, and several of these regions are conserved at different sites. In this study, survival was measured according to the capacity of establishment (death or survival) under the arid conditions in northern Chile, and therefore, the detected loci may be related to any of the criteria used in the previously mentioned studies.

In summary, genomic regions associated with complex traits that may contribute to increase the efficiency of the small-scale breeding program have been identified. Six significant genomic regions may be used as selection criteria for a genetic improvement of more than one trait (growth traits, flowering traits, and survival) under arid conditions.

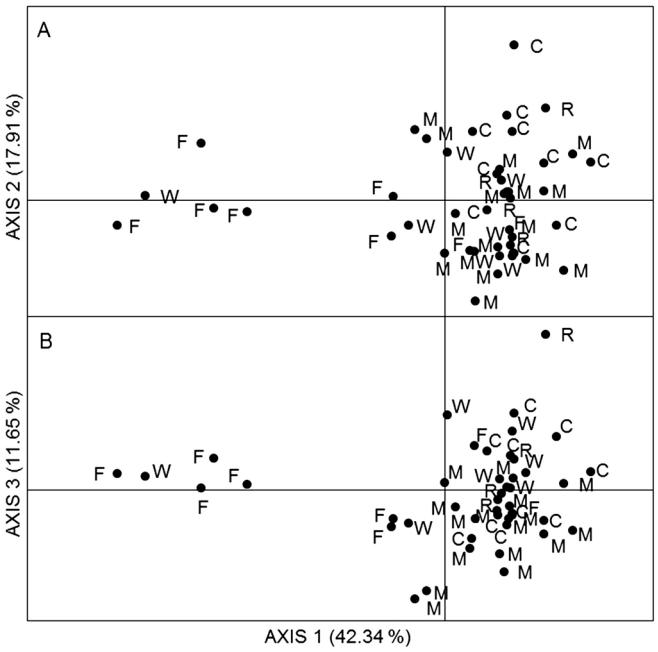

Fig. 2. A principal coordinate analysis (PCoA) of ISSR data using the covariance analysis. In both PCoA graphs, the first axis supports the highest percent of the total variance, which separates most samples of Flinders Chase from the rest. The second and third axes represent 17.91 and 11.65 % of the total variance, respectively. Each family is represented by a *black circle* and the abbreviations F, W, M, R, and C correspond to Flinders Chase, Wirrabara, Mount Remarkable, Marble Range, and Cowell, respectively.

References

Balasaravanan, T., Chezhian, P., Kamalakannan, R., Ghosh, M., Yasodha, R., Varghese, M., Gurumurthi, K.: Determination of inter- and intra-species genetic relationships among six *Eucalyptus* species based on inter-simple sequence repeats (ISSR). - Tree Physiol. **25**: 1295-1302, 2005.

Barakat, M.N., Wahba, L.E., Milad, S.I.: Molecular mapping of QTLs for wheat flag leaf senescence under water-stress. - Biol. Plant. **57**: 79-84, 2013.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. Buckler, E.S.: TASSEL: software for association mapping of complex traits in diverse samples. - Bioinformatics **23**: 2633-2635, 2007.

Bush, D., McCarthy, K., Meder, R.: Genetic variation of natural durability traits in *Eucalyptus cladocalyx* (sugar gum). - Ann. Forest Sci. **68**: 1057-1066, 2011.

Bush, D., Thumma, B.: Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers. - Tree Genet. Genomes **9**: 741-752, 2013.

Callister, A., Bush, D.J., Collins, S., Davis, W.: Prospects for genetic improvement of *Eucalyptus cladocalyx* in Western Australia. - New Zeal. J. Forest. Sci. **38**: 211-226, 2008.

Cané-Retamales, C., Mora, F., Vargas-Reeve, F., Contreras-Soto, R.: Bayesian threshold analysis of breeding values, genetic correlation and heritability of flowering intensity in *Eucalyptus cladocalyx* under arid conditions. - Euphytica **178**: 177-183, 2011.

Cappa, E.P., Pathauer, P.S., López, G.A.: Provenance variation and genetic parameters of *Eucalyptus viminalis* in Argentina. - Tree Genet. Genomes **6**: 981-994, 2010.

Cardon, L.R., Bell, J.I.: Association study designs for complex diseases. - Nature Rev. Genet. **2**: 91-99, 2001.

Chezhian, P., Yasodha, R., Gosh, M.: Genetic diversity analysis in a seed orchard of *Eucalyptus tereticornis*. - New Forests **40**: 85-99, 2010.

Doyle, J.J., Doyle, J.L.: Isolation of DNA from fresh plant tissue. - Focus **12**: 13-15, 1987.

Ellis M.F., Sedgley, M.: Floral morphology and breeding system of three species of *Eucalyptus*, section *bisectaria* (*Myrtaceae*). - Aust. J. Bot. **40**: 249-262, 1992.

Evanno, G., Regnaut, S., Goudet, J.: Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. - Mol. Ecol. **14**: 2611-262, 2005.

Excoffier, L., Hofer, T., Foll, F.: Detecting loci under selection in a hierarchically structured population. - Heredity **103**:

285-298, 2009.

Excoffier, L., Lischer, H.E.L.: *Arlequin suite ver 3.5*: a new series of programs to perform population genetics analyses under *Linux* and *Windows*. - Mol. Ecol. Resources **10**: 564-567, 2010.

Freeman, J.S., Whittock, S.P., Potts, B.M., Vaillancourt, R.E.: QTL influencing growth and wood properties in *Eucalyptus globulus*. - Tree Genet. Genomes **5**: 713-722, 2009.

Gonzalez-Martinez, S.C., Huber, D., Ersoz, E., Davis, J.M., Neale, D.B.: Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. - Heredity **101**: 19-26, 2008.

Kelleher, C.T., Wilkin, J., Zhuang, J., Cortés, A.J., Quintero, A.L.P., Gallagher, T.F., Bohlmann, J., Douglas, C.J., Ellis, B.E., Ritland, K.: SNP discovery, gene diversity, and linkage disequilibrium in wild populations of *Populus tremuloides*. - Tree Genet. Genomes **8**: 821-829, 2012.

Lohwasser, U., Rehman Arif, M.A., Börner, A.: Discovery of loci determining pre-harvest sprouting and dormancy in wheat and barley applying segregation and association mapping. - Biol. Plant. **57**: 663-674, 2013.

McDonald, M.W., Rawlins, M., Butchet, P.A., Bell, J.C.: Regional divergence and inbreeding in *Eucalyptus cladocalyx* (*Myrtaceae*). - Aust. J. Bot. **51**: 393-403, 2003.

Missiaggia, A.A., Piacezzi, A., Grattapaglia, D.: Genetic mapping of Eef1, a major effect QTL for early flowering in *Eucalyptus grandis*. - Tree Genet. Genomes **1**: 79-84, 2005.

Mora, F., Gleadow, R., Perret, S., Scapim, C.A.: Genetic variation for early flowering, survival and growth in sugar gum (*Eucalyptus cladocalyx* F. Muell) in southern Atacama Desert. - Euphytica **169**: 335-344, 2009.

Mora, F., Serra, N.: Bayesian estimation of genetic parameters for growth, stem straightness, and survival in *Eucalyptus globulus* on an Andean Foothill site. - Tree Genet. Genomes **10**: 711-719, 2014.

Okun, D.O., Kenya, E.U., Oballa, P.O., Odee, D.W., Muluvi, G.M.: Analysis of genetic diversity in *Eucalyptus grandis* (Hill *ex* Maiden) seed sources using inter simple sequence repeats (ISSR) molecular markers. - Afr. J. Biotechnol. **7**: 2119-2123, 2008.

Peakall, R., Smouse, P.: *GENALEX 6*: genetic analysis in *Excel*. Population genetic software for teaching and research. - Mol. Ecol. Notes **6**: 288-295, 2006.

Pritchard, J.K., Stephens, M., Donnelly, P.: Inference of population structure using multilocus genotype data. - Genet. **155**: 945-959, 2000.

Rönnberg-Wästljung, A.C., Glynn, C., Weih, M.: QTL analyses of drought tolerance and growth for *Salix dasyclados* × *Salix viminalis* hybrid in contrasting water regimes. - Theor. appl. Genet. **110**: 537-549, 2005.

Thamarus, K.A., Groom, K., Bradley, A., Raymond, C.A., Schimleck, L.R., Williams, E.R., Moran, G.F.: Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of *Eucalyptus globulus*. - Theor. appl. Genet. **109**: 856-864, 2004.

Thumma, B.R., Baltunis, B.S., Bell, J.C., Emebiri, L.C., Moran, G.F., Southerton, S.G.: Quantitative trait locus (QTL) analysis of growth and vegetative propagation traits in *Eucalyptus nitens* full-sib families. - Tree Genet. Genomes **6**: 877-889, 2010.

Thumma, B.R., Nolan, M.F., Evans, R., Moran, G.: Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus spp*. - Genetics **171**: 1257-1265, 2005.

Tschaplinski, T.J., Tuskan, G.A., Sewell, M.M., Gebre, G.M., Todd, D.E., Pendley, C.D.: Phenotypic variation and quantitative trait locus identification for osmotic potential in an interspecific hybrid inbred F2 poplar pedigree grown in contrasting environments. - Tree physiol. **26**: 595-604, 2006.

Vargas-Reeve, F., Mora, F., Perret, S., Scapim, C.A.: Heritability of stem straightness and genetic correlations in *Eucalyptus cladocalyx* in the semi-arid region of Chile. - Crop Breed. appl. Biotechnol. **13**: 107-112, 2013.

Yu, J.M., Pressoir. G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., Buckler, E.S.: A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. - Nature Genet. **38**: 203-208, 2006.