

## Research Paper

# Thirteen years under arid conditions: exploring marker-trait associations in *Eucalyptus cladocalyx* for complex traits related to flowering, stem form and growth

Osvin Arriagada<sup>1)</sup>, Antonio Teixeira do Amaral Junior<sup>2)</sup> and Freddy Mora\*<sup>1)</sup>

<sup>1)</sup> Institute of Biological Sciences, University of Talca, 2 Norte 685, 3460000 Talca, Chile

<sup>2)</sup> Laboratório de Melhoramento Genético Vegetal, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Av. Alberto Lamego 2000, 28013-602 Campos dos Goytacazes, Brazil

We present an association analysis for seven key traits related to flowering, stem form and growth in *Eucalyptus cladocalyx*, a tree species suitable for low rainfall sites, using a long-term progeny trial with 49 open-pollinated maternal families in the southern Atacama Desert, Chile. The progeny trial was carried out in an arid environment with a mean annual rainfall of 152 mm. Simple sequence repeats (SSR) from a full consensus map of *Eucalyptus* were used for genotyping 245 individual trees. Twenty-three significant marker-trait associations were identified, explaining between 5.9 and 23.7% of the phenotypic variance. The marker EMBRA101 located on LG10 at 56.5 cM was concomitantly associated with diameter at breast height and tree height. Nine SSR were significantly associated with stem forking and stem straightness, explaining between 5.9 and 14.8% of the phenotypic variation. To our knowledge, this is the first study reporting a SSR-based association mapping analysis for stem form traits in *Eucalyptus*. These results provide novel and valuable information for understanding the genetic base of key traits in *E. cladocalyx* for breeding purposes under arid conditions.

**Key Words:** EMBRA, flowering, forking, quantitative trait loci, tree growth, stem form.

## Introduction

*Eucalyptus* is the most widely planted hardwood genus in the world because of its broad adaptability, rapid growth and wood properties (Rockwood *et al.* 2008). In this context, extensive genome mapping studies using bi-parental and natural populations have identified hundreds of quantitative trait loci (QTL) for several complex traits in *Eucalyptus*, in particular those related to growth and wood properties (Kullan *et al.* 2012).

In *Eucalyptus*, association genetic studies have been relatively limited (Külheim *et al.* 2011, Thavamanikumar *et al.* 2014, Thumma *et al.* 2005, 2009). However, as these studies are carried out in a large part of the whole population, marker-trait associations have been successfully validated in independent populations. For example, Thumma *et al.* (2005), using trees from a natural population of *E. nitens*, found two haplotypes significantly associated with microfibril angle (MFA) in the cinnamoyl CoA reductase (CCR) gene, which

explained between 3.4 and 5.9% of the total variation in MFA; results that were confirmed in two populations of *E. nitens* and *E. globulus*. In addition, Thavamanikumar *et al.* (2014) identified nine stable single nucleotide polymorphisms (SNP) associated with wood quality and growth traits in two populations (discovery and validation populations) of *E. globulus* growing on different sites. Overall, these results demonstrate the utility of association mapping studies to identify molecular markers closely associated with target traits, allowing their utilization and validation in different populations for which they were developed. Nevertheless, these studies are focused on a limited number of species belonging to the subgenus *Symphyomyrtus* section *Maidenaria*, such as *E. grandis* and *E. globulus* (Mora and Serra 2014, Song *et al.* 2016) whose performance and/or tree growth are influenced by water deficit.

The availability of water is the main environmental factor limiting tree growth and productivity in drylands, which cover about 41% of the global land surface and are expanding due to global warming (Reynolds *et al.* 2007). Low annual rainfalls and warm temperatures characterize these areas; climate conditions that lead to water scarcity (Plaza-Bonilla *et al.* 2015).

Breeding programs of *Eucalyptus cladocalyx* F. Muell

Communicated by Atsushi Watanabe

Received October 27, 2017. Accepted April 1, 2018.

First Published Online in J-STAGE on June 29, 2018.

\*Corresponding author (e-mail: morapoblete@gmail.com)

have been undertaken in Australia (Bush *et al.* 2011) and Chile (Mora *et al.* 2009) because the species is well suited to dryland areas where it might be planted for the production of sawn timber, naturally durable posts and honey (Bush *et al.* 2015). In Chile, breeding populations of *E. cladocalyx* have been established in the southern Atacama Desert, in which water scarcity and soil erosion are extreme and generate socio-economic impacts on its inhabitants (Jorquera-Jaramillo *et al.* 2015). Therefore, identifying molecular markers associated with target traits related to drought adaptation would be useful for the genetic improvement of *E. cladocalyx* in drylands.

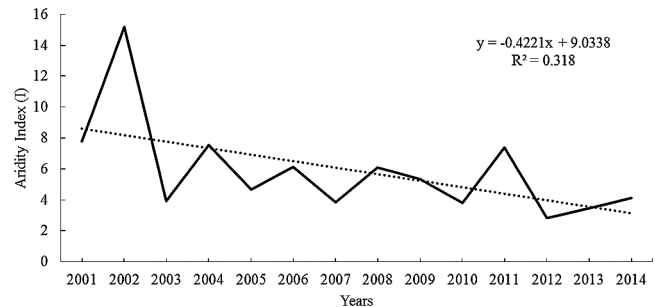
High levels of genetic divergence and moderate levels of genetic variability have been found among natural populations of *E. cladocalyx* (Bush *et al.* 2011, Mora *et al.* 2017). Moreover, significant phenotypic variation in straightness, height, diameter, precocity and intensity of flowering have been reported in arid environments (Cané-Retamales *et al.* 2011, Mora *et al.* 2009, Vargas-Reeve *et al.* 2013), suggesting a potential for the genetic improvement through the identification of molecular markers associated with these traits.

To implement a breeding program, it is essential to know the genetic basis of target traits, but to date, there has been little information published on the molecular control of complex traits in *E. cladocalyx*. Therefore, the objectives of the present study were to: (1) describe the pattern of genetic structure among natural populations of *E. cladocalyx*, (2) provide a first insight on the LD within the genome of *E. cladocalyx*, and (3) analyze the association of SSR markers with seven key complex traits using a long-term progeny trial in the southern Atacama Desert, Chile.

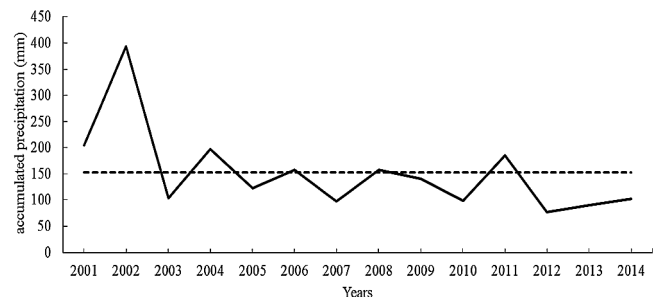
## Materials and Methods

### Plant material and phenotypic evaluation

The association analysis was performed in a long-term provenance-progeny trial comprising 49 open-pollinated maternal families of *E. cladocalyx* established in 2001. The trial was situated in the administrative region of Coquimbo, Choapa Province (31°38' S Latitude; 71°19' W Longitude; and altitude of 297 m) in the south of the Atacama Desert, Chile (Mora *et al.* 2009) under a randomized complete block design (with 30 blocks and single-tree-plots). The climate was classified as arid, according to the De Martonne aridity index, during the period 2001–2014 (Fig. 1). The mean annual rainfall of the study site was 152 mm, varying from 77 mm (2012) to 394 mm (2002) (Fig. 2), according to the nearest meteorological station in the city of Illapel, Choapa Province (Center for Climate and Resilience Research <http://explorador.cr2.cl>). The target population consisted of open-pollinated maternal families (49 families with five individuals per family,  $n = 245$  genotyped trees), of which 47 are from five Australian provenances and two from a local seed source (Illapel, Choapa Province: 31°40' S Latitude; 71°14' W Longitude). For more details of the trial imple-



**Fig. 1.** Changes in the De Martonne aridity index ( $I$ ) in the study period (2001–2014), based on data from the nearest meteorology station (city of Illapel, Province of Choapa, northern Chile). When the value of  $I$  is lower than 10 the area is characterized by a dry climate.



**Fig. 2.** Annual precipitation from 2001 to 2014 based on data from the nearest meteorology station (city of Illapel, Province of Choapa, northern Chile). The horizontal dotted line indicates the mean cumulative precipitation for the study period.

mentation, see Mora *et al.* (2009).

The following seven traits related to flowering, stem form and growth were evaluated: Early Flowering (EF), which was assessed as the presence/absence of capsules and/or flower buds at 3 years of age (Mora *et al.* 2009), Flowering Intensity (FI) of 13-year-old trees, which was ranked from no flowering (0) to heavy flowering (3) according to the method of Cané-Retamales *et al.* (2011), Reproductive Capacity (RC) of 13-year-old trees, which was evaluated on a binary scale; 0 if the tree never flowered during the period and 1 if it bloomed. Stem Straightness (STR) of 9-year-old trees (Vargas-Reeve *et al.* 2013), Stem Forking (SF) of 13-year-old trees, which was recorded as 1 when forked or 0 if single stemmed, and growth traits: Diameter at Breast Height (DBH) and Total Tree Height (HT) of 13-year-old trees.

A mixed modelling approach was used to examine phenotypic differences among provenances (and the local seed source) using SAS 9.2 (SAS Institute, Cary, NC), procedures MIXED and GENMOD for continuous and binary/multinomial traits, respectively. The Tukey-Kramer multiple comparison procedure was applied to determine significant differences between means of each provenance for continuous variables.

### DNA extraction and SSR analysis

Total genomic DNA was isolated from juvenile leaves according to Mora *et al.* (2017). One hundred and thirty SSR markers obtained from the consensus linkage map of *Eucalyptus* developed from an F<sub>1</sub> population derived from a cross between *E. grandis* and *E. urophylla* (Brondani *et al.* 2006), were used for genotyping the individuals. These markers were previously proven to be polymorphic for *E. cladocalyx* and are distributed across the eleven linkage groups (Mora *et al.* 2017). Polymerase chain reaction (PCR) was performed in 20  $\mu$ L of final reaction volume containing 40 ng of genomic DNA, 0.3  $\mu$ M of forward and reverse primers, 1 U of Taq DNA polymerase, 0.2 mM of each dNTP, 10 mM of Tris-HCl pH 8.3, 50 mM of KCl and 1.5 mM of MgCl<sub>2</sub>. PCR amplifications were performed with the following conditions: 95°C for 5 min, followed by 40 cycles of 95°C for 1 min, annealing temperature of each primer for 1 min, and 72°C for 1 min, followed by a final extension step of 72°C for 5 min. Then, the PCR products were separated on 10% (w/v) denaturing polyacrylamide gel in a run of 18 h at 80v with 1X TBE, and finally were stained and visualized according to the methods described by Mora *et al.* (2017).

### Population structure and kinship analysis

A Bayesian clustering approach was performed to infer the most probable number of genetic groups using STRUCTURE 2.3.2 (Pritchard *et al.* 2000). The number of subpopulations ( $K$ ) was set to vary between one and six based on admixture and correlated allele frequencies models, and ten runs per  $K$  were conducted separately. Each run was carried out with 100,000 Monte Carlo Markov Chain (MCMC) replicates and a burn-in period of 10,000 iterations. The true  $K$  value was inferred using the method proposed by Evanno *et al.* (2005).

A kinship analysis was performed using the software SPAGeDi (Hardy and Vekemans 2002) to define the degree of genetic covariance between each pair of individuals. The Loiselle's kinship coefficient was applied with 10,000 permutation tests. The theoretical minimum kinship is zero (i.e. individuals are not related), and that estimates below this were truncated at the zero boundary (Hardy and Vekemans 2002).

### Linkage disequilibrium (LD) analysis

The extent of LD between pairs of polymorphic loci was calculated by the squared allele frequency correlation coefficient ( $r^2$ ) implemented in the software TASSEL 2.1 (Bradbury *et al.* 2007). The significance of pairwise LD ( $p < 0.05$ ) among all possible SSR loci was determined with 1,000 permutations.

### Association analysis

Adjusted entry means (AEM) were calculated for each individual according to Stich *et al.* (2008), which were used as adjusted phenotypes for the subsequent association anal-

ysis (Mora *et al.* 2016). The marker-trait association analysis was carried out in TASSEL 2.1 (Bradbury *et al.* 2007) with the following mixed linear model proposed by Yu *et al.* (2006):

$$y = Sa + Qv + Zu + \varepsilon$$

where  $y$ ,  $\alpha$ ,  $v$ ,  $u$ , and  $\varepsilon$  are vectors of adjusted phenotypic observations, SSR effects (fixed), population structure effects (fixed), polygene background effects (random), and residual effects, respectively.  $S$ ,  $Q$  and  $Z$  are incidence matrices relating  $y$  to  $\alpha$ ,  $v$  and  $u$ , respectively.

Multiple hypothesis testing adjustment was performed using the "q-value" package in R software (Storey and Tibshirani 2003), however, none of the associations survived the false discovery rate (FDR). Therefore, a marker-trait association was declared significant when the  $p$ -value was less than 0.01, without FDR correction. The amount of phenotypic variation explained by each marker was estimated as the coefficient of determination ( $R^2$ ) (Contreras-Soto *et al.* 2017).

## Results

Growth, flowering and stem form traits differed significantly among provenances (Tables 1, 2). Trees from Wirrabara State Forest had the higher mean in HT (7 m) and DBH (8.8 cm), while Cowell had the lowest growth 13 years after planting. The worst provenance for early flowering (EF), flowering intensity (FI) and reproductive capacity (RC) was Flinders Chase, which had only 8% of the individuals blossom early, a mean for FI of 0.67 (multinomial scale) and 30% of individuals presented reproductive capacity. For stem traits, the population from Illapel had the best stem straightness (mean = 2.5), while the trees from Marble Range had the lowest stem straightness value with a mean of 0.3. Cowell had the highest number of trees with SF (56%), and interestingly, the individuals from Wirrabara State Forest did not present SF. These results are similar to those obtained by Mora *et al.* (2009), Cané-Retamales *et al.* (2011), and Vargas-Reeve *et al.* (2013), indicating that the sampling population under study is representative for the trial.

Based on the Evanno method, the wild Australian trees were divided into three genetically-differentiated groups (Fig. 3), which coincide with the three geographical regions from where these F<sub>1</sub> families were derived. 97.5% (39 of 40) of the trees from Flinders Chase were grouped in the genetic group 1. 90% (18 of 20) and 94% (47 of 50) of the individuals from Marble Range and Cowell were respectively grouped in the genetic group 2. Finally, 86.7% (39 of 45) and 92.5% (74 of 80) of the trees from Wirrabara State Forest and Mount Remarkable, were respectively grouped in the genetic group 3. The ten individuals from the local seed source were mainly grouped in the genetic group 3 (80%).

The majority (80.91%) of pairwise comparisons (29,892)

**Table 1.** Flowering intensity (FI) and stem straightness (STR; multinomial scale) variation among provenances (and the local seed source), and percentage in each category for the target population of *Eucalyptus cladocalyx* evaluated under low rainfall conditions

	Mount Remarkable <sup>a</sup>	Cowell <sup>a</sup>	Marble Range <sup>a</sup>	Wirrabara State Forest <sup>a</sup>	Flinders Chase <sup>a</sup>	Illapel <sup>b</sup>	Total
Families (N)	16	10	4	9	8	2	49
Trees (N)	80	50	20	45	40	10	245
FI (%)							
0	29	21	33	22	69	40	34
1	31	21	27	27	13	20	24
2	21	15	27	24	0	0	16
3	19	44	13	27	18	40	25
MV	1.3 B	1.8 A	1.2 B	1.6 AB	0.67 C	1.4 AB	
STR (%)							
0	1	45	76	0	0	0	14
1	16	47	18	11	33	0	23
2	49	8	6	47	46	50	38
3	33	0	0	42	21	50	25
MV	2.1 A	0.6 C	0.3 D	2.3 A	1.9 B	2.5 A	

<sup>a</sup> Australian provenances. <sup>b</sup> Chilean landrace.

FI: Flowering intensity; scored on a scale from 0 (no flowering) to 3 (most intense), STR: Stem straightness; scored from 0 (least straight) to 3 (straightest).

MV: mean on the multinomial scale; values with the same letter indicate that the populations are not significantly different according to the Tukey-Kramer test.

**Table 2.** Growth, stem forking and flowering variation among provenances (and the local seed source), for the target population of *Eucalyptus cladocalyx* evaluated under low rainfall conditions

	Mount Remarkable <sup>a</sup>	Cowell <sup>a</sup>	Marble Range <sup>a</sup>	Wirrabara State Forest <sup>a</sup>	Flinders Chase <sup>a</sup>	Illapel <sup>b</sup>
Families (N)	16	10	4	9	8	2
Trees (N)	80	50	20	45	40	10
EF	0.27 A	0.31 A	0.18 AB	0.29 A	0.08 B	0.40 A
RC	0.71 A	0.74 A	0.61 AB	0.8 A	0.3 B	0.7 AB
SF	0.3 C	0.56 A	0.41 AB	0 C	0.18 B	0.10 B
HT (m)	6.5 A	4.2 B	4.5 B	7.0 A	6.9 A	6.4 A
DBH (cm)	7.8 AB	4.0 D	5.1 CD	8.8 A	7.1 BC	7.2 ABC

<sup>a</sup> Australian provenances. <sup>b</sup> Chilean landrace.

EF: Early flowering, RC: Reproductive capacity, SF: Stem forking, HT: Total tree height, DBH: Diameter at breast height.

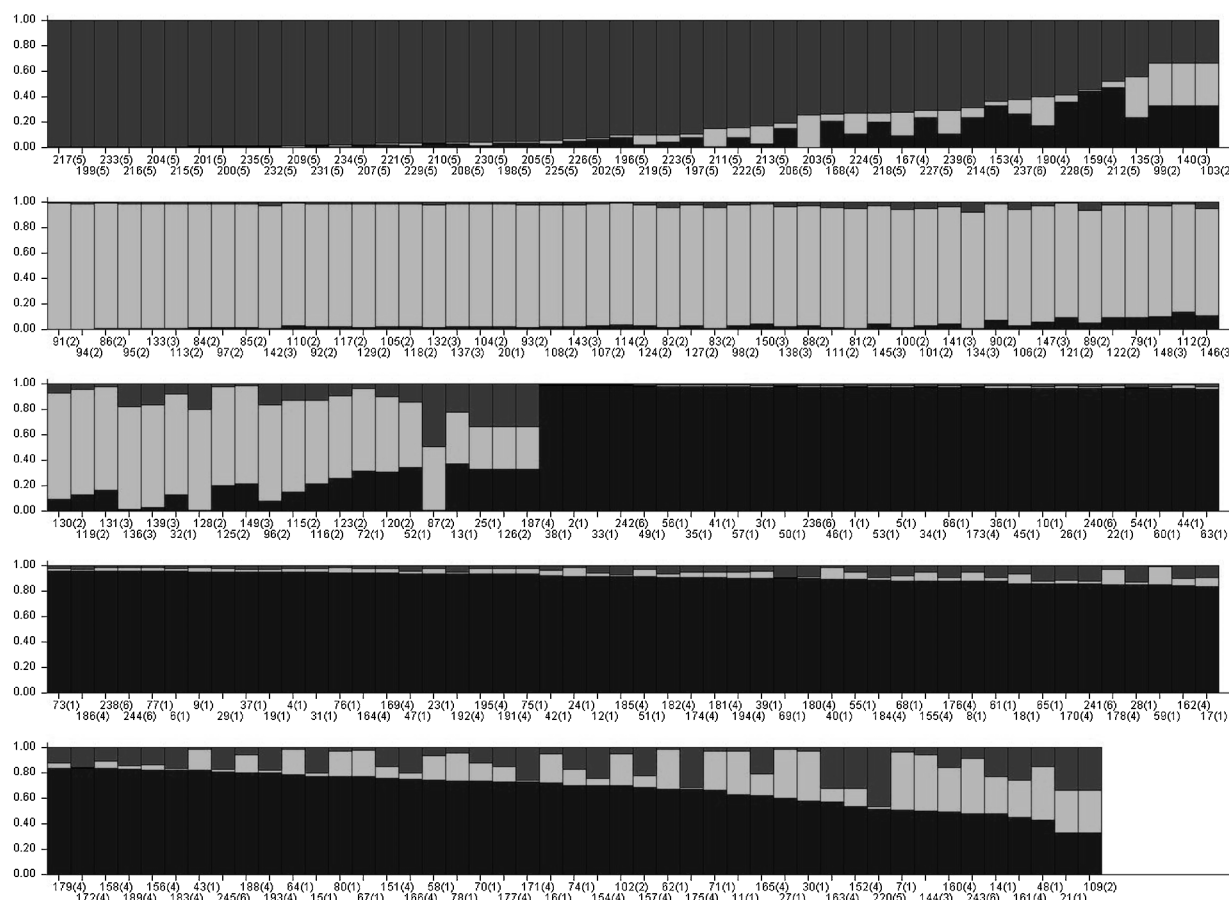
Values with the same letter in the same row indicate that the populations are not significantly different according to the Tukey-Kramer test.

between 245 individuals gave <0.05 kinship values, while 12.12% of the pairwise kinship estimates ranged from 0.05 to 0.10. The coefficient of relatedness expected for open-pollinated families is equal or higher than 0.25, and only a small proportion of pairwise kinship coefficients were larger than this value. This was essentially due to the majority of analyses were performed between individuals belonging to different half-sib families. Moreover, when the coefficients of relatedness are estimated using molecular markers, it can have lower values due to statistical error (Hansen and Nielsen 2010).

A total of 4,593 pairs of comparison were performed to investigate linkage disequilibrium (LD) in the entire set of *E. cladocalyx* genotypes. Based on  $r^2$  estimates, the average  $r^2$  of global marker pairs was 0.0101 and only 1,090 pairs of loci were significant ( $p < 0.05$ ). Moreover, the mean  $r^2$  of linked marker pairs was 0.0103, ranging from 0.0005 to 0.0518, and the 24.15% of linked marker pairs in LD was significant. LD statistics for the *E. cladocalyx* genotypes

sampled and by genetic groups are shown in the **Table 3**.

Twenty-three significant associations ( $p < 0.01$ ) were identified for the seven target traits, involving 19 SSR markers. On LG10, a relatively major association was detected (marker EMBRA101), explaining the highest proportion of phenotypic variance for HT and DBH. Flowering traits were associated with four markers, which explained between 11.7% (EMBRA115) and 18.9% (EMBRA45) of the phenotypic variation. Ten associations were identified for stem form traits on LG 2, 4, 6, 9, 10 and 11, suggesting that the associations for stem form have a genome wide distribution. The phenotypic variation explained by each marker-trait association ranged from 5.9% (EMBRA142) to 23.7% (EMBRA101) for SF and HT, respectively. The number of associations detected for each trait, including the LG number, position and the proportion of phenotypic variance explained by each significant marker are summarized in **Table 4**.



**Fig. 3.** The Bayesian clustering approach implemented in STRUCTURE suggested  $K = 3$  as the most likely number of genetically-differentiated groups. A vertical bar represents each individual. The numbers in parentheses correspond to Mount Remarkable (1), Cowell (2), Marble Range (3), Wirrabara State Forest (4), Flinders Chase (5) and Illapel (6).

**Table 3.** Linkage disequilibrium (LD) statistics for the entire set of *E. cladocalyx* genotypes and by genetic groups

	Total			Significant linked		Confidence interval (95%)		Significant unlinked		Confidence interval (95%)	
	N	Pairs of loci	Significant	Pairs of loci	Mean of $r^2$	Lower	Upper	Pairs of loci	Mean of $r^2$	Lower	Upper
Sample	245	4593	1090 (23.73%)	114	0.017	0.010	0.021	976	0.016	0.010	0.018
Cluster 1	48	836	60 (7.18%)	9	0.070	0.046	0.088	51	0.073	0.041	0.085
Cluster 2	73	3072	432 (14.06%)	44	0.045	0.030	0.046	388	0.044	0.029	0.053
Cluster 3	124	4256	685 (16.09%)	80	0.021	0.007	0.032	605	0.024	0.017	0.029

N = number of genotypes.

## Discussion

In the present study, the genetic structure analysis revealed that the population was composed of three genetically-differentiated groups ( $K = 3$ ), which coincided with the geographical origin in the state of South of Australia. This result agrees with those obtained by McDonald *et al.* (2003), who sampled natural populations from Eyre Peninsula (Marble Range), Flinders Ranges (Wirrabara State Forest and Mount Remarkable) and Kangaroo Island (Flinders Chase), reporting three homogeneous genetic groups, consistent with their geographical origin. According to Mora *et*

*al.* (2009), the origin of the trees from Illapel-Chile (local seed source) is unknown. However, our results indicated that these trees probably came from Flinders Ranges and/or Kangaroo Island. This result is in line with the Australian Low Rainfall Tree Improvement Group who mentioned that Flinders Ranges and Kangaroo Island provenances are recommended for farm forestry purposes, and therefore, it is possible that these provenances has been much more dispersed than other (Clarke *et al.* 2009).

In the studies carried out by Ballesta *et al.* (2015) and Contreras-Soto *et al.* (2016), a sample of *E. cladocalyx* was genotyped by inter-microsatellite markers to estimate the

**Table 4.** Summary of the SSR markers associated with early flowering (EF), flowering intensity (FI), reproductive capacity (RC), diameter at breast height (DBH), total tree height (HT), stem forking (SF) and stem straightness (STR) in *Eucalyptus cladocalyx*

Category	Trait	Locus	LG <sup>a</sup>	Position <sup>a</sup> (cM)	p-value	R <sup>2</sup>
Flowering	EF	EMBRA115	3	2.5	0.0097	0.1170
		EMBRA45	5	76.6	8.34E-4	0.1892
		EMBRA217	9	198	2.20E-4	0.1584
	FI	EMBRA50	6	93.2	0.0068	0.1226
	RC	EMBRA33	10	0	0.0037	0.1895
		EMBRA50	6	93.2	0.0052	0.1311
Growth	DBH	EMBRA139	8	80.1	0.0095	0.0721
		EMBRA174	7	88.3	0.0089	0.0884
		EMBRA101	10	56.5	1.02E-6	0.2331
		EMBRA32	6	93.2	0.0091	0.0782
	HT	EMBRA139	8	80.1	0.0042	0.0805
		EMBRA145	7	7.7	0.0085	0.0804
		EMBRA101	10	56.5	5.33E-7	0.2371
Form	SF	EMBRA183	9	102.7	0.0045	0.1112
		EMBRA213	4	115	6.68E-5	0.1186
		EMBRA142	2	23.7	0.0018	0.0595
		EMBRA136	10	185.2	3.81E-5	0.1174
		EMBRA51	6	101.8	0.0092	0.1021
	STR	EMBRA25	6	112.9	0.0017	0.0651
		EMBRA213	4	115	6.46E-4	0.1175
		EMBRA154	6	14.6	4.53E-5	0.098
		EMBRA8	6	25	5.80E-4	0.0983
		EMBRA165	11	9.1	0.0024	0.1418

R<sup>2</sup>: proportion of phenotypic variance explained by markers; <sup>a</sup> Linkage group number and position according to Brondani *et al.* (2006).

linkage disequilibrium (LD). In both studies, the number of loci combinations in significant LD was low. Similarly, Bush and Thumma (2013) genotyped a breeding population of *E. cladocalyx* with 79 SNP markers selected from putative genes related to important traits in *E. nitens* and obtained an average LD ( $r^2$ ) of 0.0100. These results are in accordance with our findings, where a limited number of SSR pairs showed significant  $r^2$  (0.23) and the average level of LD was low ( $r^2 = 0.0101$ ). In general, the LD found in *E. cladocalyx* is low compared with most other *Eucalyptus* species. For example, in *E. tereticornis* and *E. camaldulensis* an average LD ( $r^2$ ) equal to 0.038 and 0.039 was obtained using 62 SSR markers distributed on the eleven linkage groups, respectively (Arumugasundaram *et al.* 2011). In *E. globulus*, Cappa *et al.* (2013) estimated an average LD ( $r^2$ ) of 0.09 using 1,909 DArT markers.

In a previous report, Ballesta *et al.* (2015) found six loci associated with diameter at breast height (DBH) in *E. cladocalyx* using ISSR markers, and the proportion of phenotypic variation explained by these markers varied from 9.8% to 23.4%. This is in accordance with our findings, where four SSR markers were significantly associated with DBH on linkage groups (LG) 6, 7, 8 and 10, explaining between 7.2% and 23.3% of the phenotypic variance. Among these, the marker EMBRA101 located on LG 10 at 56.5 cM explained the greatest DBH variation. Similar to our findings, Kullán *et al.* (2012) identified five QTL for

DBH on LG 6, 9 and 10 in a hybrid population of *E. grandis* × *E. urophylla*, and the QTL located on LG10 at 89.0 cM had the greatest effect (8%).

In molecular breeding studies, frequently, two or more different traits are observed to be associated with a particular QTL. These co-localizations of QTL have been reported in *Eucalyptus*, and are consistent with significant phenotypic correlations found between the traits (van den Berg *et al.* 2016). In *E. cladocalyx*, Mora *et al.* (2009) reported a positive and moderate correlation between HT and DBH ( $r = 0.49$ ). Consistently, Ballesta *et al.* (2015) found three ISSR concomitantly associated with HT and DBH in *E. cladocalyx*. In the present study, the markers EMBRA101 and EMBRA139 were concomitantly associated with HT and DBH. These results confirm that DBH is a sufficient growth measure to use in *Eucalyptus* breeding programs because the diameter is easier and quicker to measure than height, and it has high genetic correlations with different growth traits (van den Berg *et al.* 2016).

The long reproductive cycles of *Eucalyptus* species is a limiting factor for tree improvement. Therefore, the date of first flowering and flowering intensity are target traits in breeding programs of *Eucalyptus* (Contreras-Soto *et al.* 2016). In *E. cladocalyx*, Ballesta *et al.* (2015) reported three ISSR associated with EF and FI, explaining 11.5–12% of the phenotypic variance. More recently, Contreras-Soto *et al.* (2016) identified three ISSR associated with EF each explaining between 10% and 16.4% of the phenotypic variance, and two loci associated with FI that explained 24% of total phenotypic variance. Moreover, the locus ISO1–500 bp was associated with both flowering traits, which is in agreement with the positive correlation ( $r = 0.45$ ) between EF and FI reported by Cané-Retamales *et al.* (2011) in *E. cladocalyx*. In the present study, three SSR markers were associated with EF on LG 3, 5 and 9, explaining between 11.7 and 18.9% of the phenotypic variability. However, no marker was concomitantly associated with both flowering traits. These results may be useful to identify individuals of *E. cladocalyx* that flower early and intensively for reducing the breeding cycles and increasing honey production.

Stem straightness (STR) and absence of forking in the main bole are desirable traits in forest tree breeding programs (Isik *et al.* 2015). Although, stem form traits have not been commonly studied in *Eucalyptus* because the production of pulp is one of the main uses of its wood; there is increasing interest in producing high quality plantation-grown timber for solid wood products (Hamilton *et al.* 2015). To our knowledge, no SSR-based association mapping analysis has yet been reported for SRT and SF in *Eucalyptus*. Given that *E. cladocalyx* is mainly used for the production of sawn timber and posts (Bush *et al.* 2015), it is important to understand the genetic architecture of stem form in this species in order to develop good quality trees for timber production. A previous study reported that the heritability of STR was moderate ( $h^2 = 0.40$ ) in *E. cladocalyx*, indicating a high potential for genetic improvement (Vargas-Reeve *et al.* 2013). This result

is in agreement with our findings, in which five QTL were identified for STR, explaining between 6.5% and 14.1% of the phenotypic variation. Moreover, five SSR markers were associated with SF, and one marker (EMBRA213) was concomitantly associated with STR. This is in line with previous studies, where STR had a moderate correlation with SF (Weng *et al.* 2015, Xiong *et al.* 2014), indicating that both traits may be partially controlled by the same genes and could be improved simultaneously (Xiong *et al.* 2014).

Germplasm of *E. cladocalyx* introduced into Chile presents a moderate level of genetic diversity, which would be sufficient for practical applications (Mora *et al.* 2017). Therefore, the molecular markers associated with target traits in this study are useful for the development of breeding strategies in *E. cladocalyx* by using marker-assisted selection. Moreover, the cultivation of this species is a valuable alternative for dryland farmers of northern Chile.

Importantly, the phenotypic variation explained by each marker-trait association ranged from 5.9% to 23.7%, which is considered relatively high as a single marker usually explains less than 10% of the phenotypic variance (Du *et al.* 2016). The replication of marker-trait associations in one or more independent populations is crucial for separating true from false positives and to provide less biased estimates of allelic effect sizes (Hall *et al.* 2010). According to Rockman (2008), the allelic effects generally decline in replication studies, a phenomenon known as the “Beavis effect”, and it occurs because significant associations are reported only when test statistics exceed a predetermined critical threshold. Hall *et al.* (2010) stated that the estimated effects of significant associations are sampled from a truncated distribution, and the weaker the initial effect, the more serious is this overestimation. For example in *Populus tremula*, two non-synonymous SNPs in the photoreceptor gene *PHYB2* were associated with the bud set, and each explained ~8% of the phenotypic variation. However, after correcting for the possible upward bias in the effect size, these two SNPs accounted for 1.4% and 5.9% of the variation in bud set (Ingvarsson *et al.* 2008). An important next step would therefore be to validate the associations detected in this study in other populations of *E. cladocalyx* to remove spurious associations and to provide less biased estimates of allelic effect sizes.

In conclusion, the breeding population of *E. cladocalyx* used in this study is genetically structured corresponding to the three geographical regions where these samples have been derived. This study provides a first insight into LD within the genome of *E. cladocalyx*. An important number of significant associations were identified for the seven target traits, in which four genomic regions could be used as selection criteria for more than one trait. Our findings provide novel and valuable information for understanding the genetic architecture of complex traits in *E. cladocalyx* for breeding purposes under arid conditions. However, the confirmation of these results in other genetic backgrounds and environments is required.

## Acknowledgements

This study was funded by FONDECYT (Grant Numbers 1130306 and 1170695). The authors thank Mr. Augusto Gomes for providing the samples of *E. cladocalyx*. Osvin Arriagada thanks CONICYT for a doctoral fellowship (CONICYT-PCHA/Doctorado Nacional/año 2013-folio 21130812).

## Literature Cited

- Arumugasundaram, S., M. Ghosh, S. Veerasamy and Y. Ramasamy (2011) Species discrimination, population structure and linkage disequilibrium in *Eucalyptus camaldulensis* and *Eucalyptus tereticornis* using SSR markers. *PLoS ONE* 6: e28252.
- Ballesta, P., F. Mora, E. Ruiz and R. Contreras-Soto (2015) Marker-trait associations for survival, growth, and flowering components in *Eucalyptus cladocalyx* under arid conditions. *Biol. Plant.* 59: 389–393.
- Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633–2635.
- Brondani, R.P.V., E.R. Williams, C. Brondani and D. Grattapaglia (2006) A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. *BMC Plant Biol.* 6: 20.
- Bush, D., D. Kain, C. Matheson and P. Kanowski (2011) Marker-based adjustment of the additive relationship matrix for estimation of genetic parameters—an example using *Eucalyptus cladocalyx*. *Tree Genet. Genomes* 7: 23–35.
- Bush, D. and B. Thumma (2013) Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers. *Tree Genet. Genomes* 9: 741–752.
- Bush, D., D. Kain, P. Kanowski and C. Matheson (2015) Genetic parameter estimates informed by a marker-based pedigree: a case study with *Eucalyptus cladocalyx* in southern Australia. *Tree Genet. Genomes* 11: 798.
- Cané-Retamales, C., F. Mora, F. Vargas-Reeve, S. Perret and R. Contreras-Soto (2011) Bayesian threshold analysis of breeding values, genetic correlation and heritability of flowering intensity in *Eucalyptus cladocalyx* under arid conditions. *Euphytica* 178: 177–183.
- Cappa, E.P., Y.A. El-Kassaby, M.N. Garcia, C. Acuña, N.M. Borralho, D. Grattapaglia and S.N. Marcucci-Poltri (2013) Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. *PLoS ONE* 8: e81267.
- Clarke, B., I. Leod and T. Vercoe (2009) Trees for farm forestry—22 promising species. *Rural Industries Research and Development Corporation*, 232 p.
- Contreras-Soto, R., P. Ballesta, E. Ruiz and F. Mora (2016) Identification of ISSR markers linked to flowering traits in a representative sample of *Eucalyptus cladocalyx*. *J. Forest. Res.* 27: 239–245.
- Contreras-Soto, R.I., F. Mora, F. Lazzari, M.A.R. Oliveira, C.A. Scapim and I. Schuster (2017) Genome-wide association mapping for flowering and maturity in tropical soybean: implications for breeding strategies. *Breed. Sci.* 67: 435–449.
- Du, Q., C. Gong, Q. Wang, D. Zhou, H. Yang, W. Pan, B. Li and D. Zhang (2016) Genetic architecture of growth traits in *Populus* revealed by integrated quantitative trait locus (QTL) analysis and

- association studies. *New Phytol.* 209: 1067–1082.
- Evanno, G., S. Regnaut and J. Goudet (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611–2620.
- Hall, D., C. Tegström and P.K. Ingvarsson (2010) Using association mapping to dissect the genetic basis of complex traits in plants. *Brief. Funct. Genomics* 9: 157–165.
- Hamilton, M.G., M. Acuna, J.C. Wiedemann, R. Mitchell, D.J. Pilbeam, M.W. Brown and B.M. Potts (2015) Genetic control of *Eucalyptus globulus* harvest traits. *Can. J. For. Res.* 45: 615–624.
- Hansen, O.K. and U.B. Nielsen (2010) Microsatellites used to establish full pedigree in a half-sib trial and correlation between number of male strobili and paternal success. *Ann. For. Sci.* 67: 703.
- Hardy, O.J. and X. Vekemans (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Resour.* 2: 618–620.
- Ingvarsson, P.K., M.V. Garcia, V. Luquez, D. Hall and S. Jansson (2008) Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* 178: 2217–2226.
- Isik, F., S. Kumar, P.J. Martínez-García, H. Iwata and T. Yamamoto (2015) Chapter Three—Acceleration of forest and fruit tree domestication by genomic selection. *Adv. Bot. Res.* 74: 93–124.
- Jorquera-Jaramillo, C., M. Yáñez-Acevedo, J.R. Gutiérrez, J.L. Cortés-Bugueño, V. Pastén-Marambio and C. Barraza-Cepeda (2015) Restoration of degraded drylands in northern Chile: The need of local stakeholders' participation to prevent and combat desertification. In EGU General Assembly Conference Abstracts (Vol. 17).
- Külheim, C., S.H. Yeoh, I.R. Wallis, S. Laffan, G.F. Moran and W.J. Foley (2011) The molecular basis of quantitative variation in foliar secondary metabolites in *Eucalyptus globulus*. *New Phytol.* 191: 1041–1053.
- Kullan, A.R.K., M.M. Van Dyk, C.A. Hefer, N. Jones, A. Kanzler and A.A. Myburg (2012) Genetic dissection of growth, wood basic density and gene expression in interspecific backcrosses of *Eucalyptus grandis* and *E. urophylla*. *BMC Genet.* 13: 60.
- McDonald, M.W., M. Rawlings, P.A. Butcher and J.C. Bell (2003) Regional divergence and inbreeding in *Eucalyptus cladocalyx* (Myrtaceae). *Aust. J. Bot.* 51: 393–403.
- Mora, F., R. Gleadow, S. Perret and C.A. Scapim (2009) Genetic variation for early flowering, survival and growth in sugar gum (*Eucalyptus cladocalyx* F. Muell) in southern Atacama Desert. *Euphytica* 169: 335–344.
- Mora, F. and N. Serra (2014) 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.
- Mora, F., Y.A. Quitral, I. Matus, J. Russell, R. Waugh and A. del Pozo (2016) SNP-based QTL mapping of 15 complex traits in barley under rain-fed and well-watered conditions by a mixed modeling approach. *Front. Plant Sci.* 7: 909.
- Mora, F., O. Arriagada, P. Ballesta and E. Ruiz (2017) Genetic diversity and population structure of a drought-tolerant species of *Eucalyptus*, using microsatellite markers. *J. Plant Biochem. Biotechnol.* 26: 274–281.
- Plaza-Bonilla, D., J.L. Arrúe, C. Cantero-Martínez, R. Fanlo, A. Iglesias and J. Álvaro-Fuentes (2015) Carbon management in dryland agricultural systems. A review. *Agron. Sustain. Dev.* 35: 1319–1334.
- Pritchard, J.K., M. Stephens and P. Donnelly (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Reynolds, J.F., D.M.S. Smith, E.F. Lambin, B.L. Turner II, M. Mortimore, S.P.J. Batterbury, T.E. Downing, H. Dowlatabadi, R.J. Fernández, J.E. Herrick *et al.* (2007) Global desertification: building a science for dryland development. *Science* 316: 847–851.
- Rockman, M.V. (2008) Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature* 456: 738–744.
- Rockwood, D.L., A.W. Rudie, S.A. Ralph, J.Y. Zhu and J.E. Winandy (2008) Energy product options for eucalyptus species grown as short rotation woody crops. *Int. J. Mol. Sci.* 9: 1361–1378.
- Song, Z., M. Zhang, F. Li, Q. Weng, C. Zhou, M. Li, J. Li, H. Huang, X. Mo and S. Gan (2016) Genome scans for divergent selection in natural populations of the widespread hardwood species *Eucalyptus grandis* (Myrtaceae) using microsatellites. *Sci. Rep.* 6: 34941.
- Stich, B., J. Möhring, H.P. Piepho, M. Heckenberger, E.S. Buckler and A.E. Melchinger (2008) Comparison of mixed-model approaches for association mapping. *Genetics* 178: 1745–1754.
- Storey, J.D. and R. Tibshirani (2003) Statistical significance for genome-wide studies. *Proc. Natl. Acad. Sci. USA* 100: 9440–9445.
- Thavamanikumar, S., L.J. McManus, P.K. Ades, G. Bossinger, D.J. Stackpole, R. Kerr, S. Hadjigol, J.S. Freeman, R.E. Vaillancourt, P. Zhu *et al.* (2014) Association mapping for wood quality and growth traits in *Eucalyptus globulus* ssp. *globulus* Labill identifies nine stable marker-trait associations for seven traits. *Tree Genet. Genomes* 10: 1661–1678.
- Thumma, B.R., M.F. Nolan, R. Evans and G.F. Moran (2005) Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* 171: 1257–1265.
- Thumma, B.R., B.A. Matheson, D. Zhang, C. Meeske, R. Meder, G.M. Downes and S.G. Southerton (2009) Identification of a *cis*-acting regulatory polymorphism in a eucalypt *COBRA*-like gene affecting cellulose content. *Genetics* 183: 1153–1164.
- Van den Berg, G.J., S.D. Verry, P.W. Chirwa and F. van Deventer (2016) Estimates of genetic parameters and genetic gains for growth traits of two *Eucalyptus urophylla* populations in Zululand, South Africa. *South. For.* 78: 209–216.
- Vargas-Reeve, F., F. Mora, S. Perret and C.A. Scapim (2013) Heritability of stem straightness and genetic correlations in *Eucalyptus cladocalyx* in the semi-arid region of Chile. *Crop Breed. Appl. Biot.* 13: 107–112.
- Weng, Y.H., P. Lu, G.W. Adams, M.S. Fullarton and K.J. Tosh (2015) Genetic parameters of growth and stem quality traits for jack pine second-generation progeny tested in New Brunswick. *Can. J. For. Res.* 45: 36–43.
- Xiong, J.S., S.E. McKeand, R.W. Whetten and F.T. Isik (2014) Genetics of stem forking and ramicorn branches in a cloned loblolly pine family. *Forest Sci.* 60: 360–366.
- Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland *et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38: 203–208.