RESEARCH ARTICLE

A unifying study of phenotypic and molecular genetic variability in natural populations of *Anadenanthera colubrina* var. *cebil* from Yungas and Paranaense biogeographic provinces in Argentina

MARÍA VICTORIA GARCÍA[1,2,3*], KATHLEEN PRINZ[4,5], MARÍA EUGENIA BARRANDEGUY[1,2,3], MARCOS MIRETTI[1,2,3] and REINER FINKELDEY[4]

[1]*Facultad de Ciencias Exactas, Departamento de Genética, Químicas y Naturales, Universidad Nacional de Misiones, Posadas 3300, Argentina*
[2]*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Posadas 3300, Argentina*
[3]*Instituto de Biología Subtropical (UNaM – CONICET) Posadas 3300, Argentina*
[4]*Forest Genetics and Forest Tree Breeding, Georg-August-University Göttingen D-37077 Göttingen, Germany*
[5]*Institute for Systematic Botany with Herbarium Haussknecht and Botanical Gardens, Friedrich-Schiller-University Jena, D-07743 Jena, Germany*

Abstract

Anadenanthera colubrina var. *cebil* is a discontinuously distributed native tree species in South American subtropical forests. Thirteen quantitative traits and eight nuclear microsatellite loci were examined in individuals from two biogeographic provinces of Argentina to determine the number and composition of genetically distinguishable groups of individuals and explore possible spatial patterns of the phenotypic and genetic variability. Means of reproductive traits were higher in the Yungas than in the Paranaense biogeographic province, whereas five out of eight nonreproductive quantitative traits showed higher mean values in the latter. Variance coefficients were moderate, and there were significant differences between and within provinces. Three clusters were defined based on spatial model for cluster membership for quantitative traits. One cluster grouped the individuals from the Paranaense biogeographic province whereas the individuals from the Yungas biogeographic province grouped regarding its population of origin. Parameters of molecular genetic variability showed higher values in the Yungas than in the Paranaense biogeographic province. Observed heterozygosity was lower than expected heterozygosity in both biogeographic provinces, indicating an excess of homozygosity. The homozygosity test by Watterson and the exact test by Slatkin suggested diversifying selection for locus *Ac41.1*. Bayesian clustering spatial model for microsatellites loci data were performed for both all loci and for all loci excluding locus *Ac41.1*. In both analyses two clusters were inferred. Analysis of molecular variance revealed similar results for all genotypes and for all genotypes defined excluding locus *Ac41.1*. Most of the total variance is attributable to genetic variation within clusters. The presence of homogeneous clusters was detected for both the phenotypic and molecular genetic variability. Two Bayesian clustering analyses were performed according to molecular genetic data, and two clusters were inferred. Individuals were assigned to their provinces of origin. Genetic molecular variation was higher in the populations of the Yungas biogeographic province which translates into highly qualified populations for conservation. Populations from the Paranaense biogeographic province showed the highest mean value of number of seeds per fruit making them valuable as well with regard to the exploitation of management strategies as a means to recover the impacted areas where these populations are located.

Introduction

Anadenanthera colubrina (Vell.) Brenan var. *cebil* (Fabaceae, Mimosoideae), locally known as curupay, is a

tree species native to subtropical forests of South America. These trees can reach up to 35 m in height, with flowers arranged in hermaphrodite inflorescences and long legume fruits with narrow and flattened seeds (Justiniano and Fredericksen 1998; Cialdella 2000). Regarding its mating system, it has been suggested that *A. colubrina* var. *cebil* is

a predominantly outcrossing species (Cialdella 2000). Bees are the main pollinators, and seeds are dispersed by autochory or anemochory after pod dehiscence (Justiniano and Fredericksen 1998; Abraham de Noir *et al.* 2002).

Valuable traits for this species, such as high germination, rapid growth, regeneration rate in disturbed areas and ability to adapt to environmental disturbances make it an ideal species for sustainable forest management, playing a key role in the recovery of degraded forest areas (Justiniano and Fredericksen 1998).

A. colubrina var. *cebil* is discontinuously distributed following the distribution of seasonally dry tropical forests (SDTFs) (Prado and Gibbs 1993; Pennington *et al.* 2000). Three nuclei can be identified for this species and two of them are located in Argentina: the Misiones nucleus includes the Paranaense biogeographic province in northeastern Argentina and the Sub-Andean Piedmont nucleus where *A. colubrina* var. *cebil* is important in the upland subtropical rainforest (the 'Yungas' s. str.) (Prado and Gibbs 1993). These biogeographic provinces still keep high levels of biodiversity. However, as a consequence of the occurrence of SDTFs on highly favourable soils for agriculture, a massive destruction of this forest has been reported in South America (Prado 2000). Consequently, the area corresponding to natural *A. colubrina* var. *cebil* forests has drastically diminished over the last 30 years and the remaining forest are still under threat from different forms of forest destruction. At present, the Paranaense biogeographic province has strongly been affected by human activities, while populations in the Yungas have slightly been affected by these activities since they are located in protected areas.

Phenotypic characters are species-specific and may vary strongly depending on species distribution, habitat conditions as well as human interference (Holderegger *et al.* 2006). Phenotypic characters allow for adaptation analysis and selection of quantitative traits to a particular environment (Holderegger *et al.* 2006). Variation of quantitative traits is the basis for the development of management strategies and resource conservation (Zobel and Talbert 1991).

On the other hand, molecular markers such as microsatellites, also known as simple sequence repeats (SSR), are gaining popularity in genetic diversity analysis due to their high level of polymorphism, codominance and neutral Mendelian inheritance (Wee *et al.* 2012) indicating that the pattern of genetic differentiation is completely determined by the demographic history of the populations, the mutation rates at the loci concerned and the rates and patterns of migration among the populations (Holsinger and Weir 2009). Several studies showed some evidence about the nonneutrality for several SSR loci (Nielsen *et al.* 2006; Lazrek *et al.* 2009; Shi *et al.* 2011). Hence, the processes underlying population differentiation might be inferred incorrectly if the presence of selection in putatively neutral markers is not tested (Languth and Balkenhol 2012). Loci showing unusually large amounts of differentiation may indicate regions of the genome that have been subject to diversifying selection, whereas loci showing unusually small amounts of differentiation may indicate regions of the genome that have been subject to stabilizing selection (Holsinger and Weir 2009). Despite this, patterns of genetic differentiation among populations analysed by microsatellites have been explained almost exclusively by genetic drift and migration while the potential role of selection has been ignored (Nielsen *et al.* 2006).

Considering that variability is essential to ensure evolutionary adaptation processes of forest genetic resources to changing environmental conditions, biodiversity programmes for resource conservation must be formulated to resist its loss. However, despite the value of *A. colubrina* var. *cebil* as a native resource, no information is still available regarding its phenotypic and molecular genetic variation jointly. In this way, we aimed to provide a starting point to understand and evaluate variability in the SDTFs of Argentina.

In this study, we examined the phenotypic variability using quantitative traits and genetic variability using SSR markers in four natural populations of *A. colubrina* var. *cebil* from two different biogeographic provinces in Argentina to determine the number and composition of genetically distinguishable groups of individuals and to explore possible spatial patterns of the phenotypic and genetic variability. We aimed to find trends in genetic variability within and among these biogeographic provinces to identify provinces of high and low genetic diversity, which in turn facilitate genetic conservation activities.

Materials and methods

Plant material

Individuals were sampled in two biogeographic provinces, Paranaense and Yungas, both located in the northern part of Argentina (figure 1). Individuals studied were located in four populations: Candelaria (Cand) and Santa Ana (SA) both in the Parananense biogeographic province, and Tucumán (T) and Jujuy (J) in the Yungas. Nineteen individuals were sampled in Candelaria, 16 in Santa Ana, 13 in Tucumán and 18 in Jujuy.

Fresh and healthy leaves were collected from branches located in the middle of the crown of each tree at the time of fruiting. According to availability, three to five leaves were dried and stored in the laboratory of Departamento de Genética (Facultad de Ciencias Exactas, Químicas y Naturales, Universidad Nacional de Misiones. Argentina) while two individuals from Tucumán and two individuals from Jujuy are stored in the herbarium of Universidad de Salta (Argentina) (García and Barrandeguy 38 – MCNS11943, García and Barrandeguy 37 – MCNS11946, García and Barrandeguy s.n. – MCNS11941 and García and Barrandeguy s.n. – MCNS11942). All samples are available for future studies from the authors. The remaining leaves were
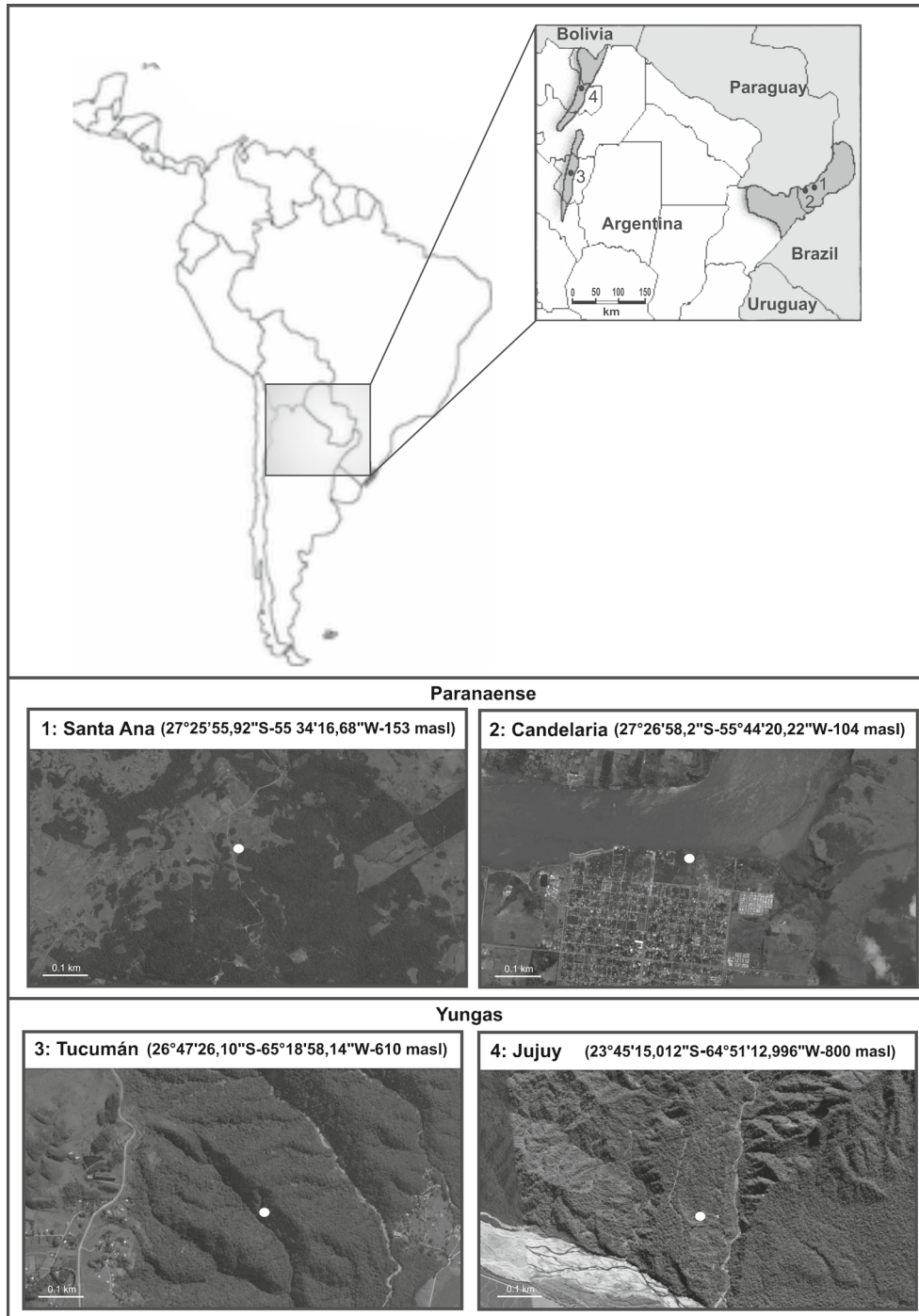
Figure 1. Geographic origin of the populations studied in the Paranaense and Yungas biogeographic provinces. Satellite pictures of both areas are shown (adapted from Google earth – Image© 2012 Digital Globe). Date and time of the retrieval picture 06/04/2012 at 14:00h (UTC-03:00) Buenos Aires.

dried with silica gel and stored at room temperature until DNA extraction. Mature fruits available were collected from investigated trees.

Quantitative traits

Variability of quantitative traits was evaluated by means of eight vegetative and five reproductive traits. The evaluated vegetative traits were: NPL, number of pairs of leaflets; MDBL, mean of distance between leaflets; LML, length of medium leaflet; WML, width of medium leaflet; L/WML, length/width medium leaflet relation; LL, length of leaf; WL, width of leaf; L/WL, length/width of leaf relation; while reproductive traits were: LF, length of fruit; WF, width of fruit; NSF, number of seeds per fruit; LS, length of seed; WS,

width of seed. The traits, length of seed and width of seed were measured in all seeds of fruits analysed. According to availability, traits were measured in three to five leaves and five to eight fruits per individual.

Molecular markers and genotyping

Genomic DNA from leaves was extracted with the DNeasy® - Plant mini kit (Qiagen, Hilden, Germany). Eight nuclear microsatellite markers developed for *A. colubrina* var. *cebil* (Barrandeguy *et al.* 2012) were used for genotyping of all individuals. PCR amplifications were performed in a final volume of 15 μL using 0.5 ng/μL of genomic DNA, 1× Hot Start Buffer (0.8 M Tris-HCl pH 9.0, 0.2 M (NH$_4$)$_2$SO$_4$, 0.2% w/v Tween-20 (Solis BioDyne, Tartu, Estonia)), 2.5 mM MgCl$_2$, 0.2 mM of each dNTP, 1 U Hot Start DNA polymerase (5 U/μL Hot FirePol Solis Bio-Dyne, Tartu, Estonia) and 0.33 pmol of each primer. Forward primers were fluorescently labelled (FAM=blue or HEX=green) on the 5′-end. PCR was performed in a gradient cycler (Biometra, Göttingen, Germany) using a touchdown program. PCR conditions were: 94°C for 15 min, 10 cycles of 1 min at 94°C, annealing for 1 min at temperatures decreasing from 60°C to 50°C (with 1°C decremental steps in each cycle) and 1 min at 72°C followed by 29 similar annealing cycles at 50°C, and a final elongation at 72°C for 20 min. Two loci (*Ac28.3* and *Ac157.1*) were amplified without the touchdown program, and the annealing temperature was 65°C in 30 cycles. Electrophoresis was carried out in an ABI Prism® 3100 (Applied Biosystems, Foster City, USA), and fragment sizes were scored with GeneScan™ analysis software using GS 500 ROX™ (Applied Biosystems) as internal size standard.

Data analysis

Phenotypic variability: Mean values, standard deviation and coefficient of variance were calculated for quantitative traits. A nested analysis of variance with unequal sample sizes (analysis of variance, ANOVA) (Sokal and Rohl 1980) was performed between biogeographic provinces and among individuals within the provinces using the following linear model:

$$Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk},$$

where Y$_{ijk}$ is the *k*th observation for the *j*th individual of the *i*th biogeographic province, μ denotes the grand mean, α_i the random contribution for the *i*th biogeographic province, β_{ij} the random contribution for the *j*th individual of the *i*th biogeographic province and ε_{ijk} is the residual error term.

The presence of homogeneous clusters was detected by a spatial model for cluster membership using the R package Geneland ver. 4.0.0 software (Guillot *et al.* 2012). This model develops a Bayesian classification to investigate genetic population structure. Prior model for cluster membership is based on statistical model known as coloured

Poisson–Voronoi tessellation which assumes that each cluster area in the geographic domain can be approximated by the union of a few polygons. The polygons are assumed to be centred by a homogeneous Poisson process, which means points located completely at random in the geographic domain (Guillot *et al.* 2012). The method was applied to the phenotypic data transformed into log-shape measures (Guillot *et al.* 2012). Different clusters (K) numbers ranged from 1 to 6 were tested. Markov chain Monte Carlo (MCMC) simulation was used to obtain the joint posterior distribution of the parameters. The estimation of these parameters was obtained using 500,000 MCMC iterations, thinning at every 100th iteration, with a burn-in of 5000 iterations. The model was launched 10 independent MCMC runs.

Molecular genetic variability: The SSR loci were tested for null alleles using the Micro-Checker ver. 2.2.3 software (van Oosterhout *et al.* 2004). This software constructs random genotypes by randomizing the observed alleles for each locus within samples and subsequently comparing the observed genotypes with the distribution of randomized genotypes. The frequency of null alleles (NA) for each locus was estimated using the method of Brookfield (1996). Deviations from Hardy–Weinberg equilibrium (HWE) were determined for each locus in each biogeographic province. The exact *P* value associated with the null hypothesis of HWE was estimated using a Markov chain algorithm implementing the GenePop 4.0.10 software (Rousset 2008).

Number of alleles per locus was registered by mere counting. Genetic variability within the biogeographic provinces was characterized by allelic richness (*R*), observed heterozygosity (*H*$_o$) and expected heterozygosity (*H*$_e$) under HWE expectation. Allelic richness is independent from sample size (Gillet *et al.* 2005) and was calculated using the ADZE ver. 1.0 software (Szpiech *et al.* 2008), whereas heterozygosities were estimated using the GenAlEx ver. 6 software (Peakall and Smouse 2006).

The *F*$_{ST}$ index was calculated locus by locus and over all loci to analyse the structure of genetic variability. Statistical significance was calculated based on 1023 permutations. This analysis was performed using the Arlequin ver. 3.5 software (Excoffier and Lischer 2010). The estimation of *F*$_{ST}$ was performed by excluding null alleles (ENA) in order to analyse their influence (Chapuis and Estoup 2007). The analyses with and without null alleles (INA and ENA respectively) were performed using the F$_{REE}$NA software (Chapuis and Estoup 2007).

To test the neutral behaviour of the SSR loci, i.e. markers potentially under selection, a Ewens–Watterson test on allele frequency distribution was performed for each locus to detect deviations from a neutral equilibrium model using the Arlequin ver. 3.5 software (Excoffier and Lischer 2010). The probability of neutrality for each SSR locus was assessed using both the homozygosity test-statistic by Watterson (1978) and the exact test by Slatkin (1994, 1996).

A Bayesian classification model for microsatellite data was used to investigate the genetic population structure. The spatial Bayesian clustering model is appropriate for quantifying genetic structure accounting for the complexities of spatial correlation in genotype and in allele frequencies when determining genetic population structure (Wheeleer and Waller 2012). Different numbers of clusters (K) ranging from 1 to 6 were tested using the uncorrelated frequency model. A MCMC simulation was used to obtain the joint posterior distribution of the parameters. The estimation of these parameters was obtained using 500,000 MCMC iterations, thinning at every 100th iteration, with a burn-in of 5000 iterations. The model was launched 10 independent MCMC runs. The spatial model analysis for cluster membership was performed using the R package Geneland ver. 4.0.0 software (Guillot *et al.* 2012).

A hierarchical analysis of molecular variance (AMOVA) was performed to examine the distribution of genetic variability between and within previously defined clusters. This analysis was performed using the Arlequin 3.5 software (Excoffier and Lischer 2010).

Phenotypic and molecular genetic variability jointly: Combination between phenotypic and genetic data can improve the accuracy of inferences carried out from one type of data only. In this way, the model proposed by Guillot *et al.* (2012) was assayed to detect the presence of homogeneous clusters. This model mimics the expectation from a population: homogeneity in terms of genetic and phenotypic variation and some geographical continuity. The existence of homogeneous clusters corresponds to the fact that some individuals have shared some aspects of their recent ecological or evolutionary history. This shared history is summarized by cluster-specific parameters that represent allele frequencies and means and variances of phenotypic traits. Spatial model for cluster membership was performed using the R package Geneland ver. $\geq$ 4.0.0 software (Guillot *et al.* 2012). Different numbers of clusters (K) ranging from 1 to 6 were tested. A MCMC simulation was used to obtain the joint posterior distribution of the parameters. The estimation of these parameters was obtained using 500,000 MCMC it erations, thinning at every 100th iteration, with a burn-in of 5000 iterations. The model was launched 10 independent MCMC runs.

Results and discussion

Phenotypic variability

Table 1 shows the mean values, standard deviation and coefficients of variance for all measured quantitative traits. Means of reproductive traits showed higher values in the individuals from Yungas biogeographic province except for WS, whereas five out of eight nonreproductive quantitative traits showed higher mean values in the Paranaense biogeographic province.

Table 1. Means, standard deviations (SD) and coefficients of variance (CV) of quantitative traits in the biogeographic provinces. Asterisks indicate reproductive traits. Traits are given in cm. CV is given in percentage.

| | Biogeographic provinces | | | |
| | Paranaense | | Yungas | |
Traits	Mean (SD)	CV	Mean (SD)	CV
NPL	15.43 (2.37)	15.0	18.35 (1.92)	10.5
MDBL	0.95 (0.12)	13.0	0.70 (0.08)	11.7
LML	7.49 (0.86)	12.0	5.36 (0.56)	10.4
WML	0.80 (0.10)	13.0	0.74 (0.08)	10.6
L/WML	9.37 (1.16)	12.0	6.83 (0.62)	9.1
LL	13.86 (1.91)	14.0	15.51 (1.94)	12.5
WL	14.88 (1.75)	12.0	10.18 (1.18)	11.6
L/WL	0.97 (0.16)	16.0	1.51 (0.17)	11.2
LF*	14.77 (3.75)	25.0	18.46 (3.95)	21.4
WF*	1.92 (0.13)	7.0	2.36 (0.38)	16.2
NSF*	8.23 (2.91)	35.0	10.07 (2.37)	23.5
LS*	1.18 (0.14)	12.0	1.30 (0.20)	15.0
WS*	1.19 (0.14)	12.0	1.09 (0.19)	17.3

NPL, number of pairs of leaflets; MDBL, mean distance between leaflets; LML, length of medium leaflet; WML, width of medium leaflet; L/WML, length/width of medium leaflet; LL, length of leaf; WL, width of leaf; L/WL, length/width of leaf; LF, length of fruit; WF, width of fruit; NSF, number of seeds per fruit; LS, length of seed; WS, width of seed.

Coefficients of variance showed moderate values, and NSF was highest in both biogeographic provinces reaching 35% in the Paranaense biogeographic province. ANOVA resulted in significant differences for analysed traits between and within the biogeographic provinces. Ten out of 13 traits showed high statistically significant differences between biogeographic provinces, whereas 12 ten out of 13 traits showed high statistically significant differences among individuals within each biogeographic provinces (table 2).

Patterns of quantitative trait variation were different in each biogeographic province indicating differences in plant adaptation for each region. Means values of four out of five reproductive traits were higher in the Yungas biogeographic province than in the Paranaense biogeographic province, whereas the mean values of five out eight vegetative traits and coefficients of variance of seven out eight vegetative traits were higher in the latter. Thus, populations in the Yungas biogeographic province are assumed to be younger than their counterparts in the Paranaense biogeographic province. However, variation of the reproductive trait NSF (35%) was higher in the Paranaense biogeographic province suggesting high rates of regeneration in disturbed areas. This finding indicates a high ability of *A. colubrina* var. *cebil* to adapt to environmental disturbances and rapid growth that is essential to survive in the highly threatened SDTFs.

Populations in the Paranaense biogeographic province are surrounded by human settlements, and this proximity indicates a source of fitness reduction in populations of *A. colubrina* var. *cebil* in this biogeographic region.

Table 2. Analyses of variance for quantitative traits measured in all individuals of *A. colubrina* var. *cebil*.

	F	
Trait	Between provinces	Within provinces
NPL	30.12**	3.73**
MDBL	23.88**	11.19**
LML	41.56**	14.21**
WML	2.84ns	12.70**
L/WML	49.60**	8.98**
LL	6.04*	24.48**
WL	58.79**	13.71**
L/WL	75.81**	0.01ns
LF*	1.04ns	15.16**
WF*	24.57**	4.41**
NSF*	65.55**	10.13**
LS*	0.55ns	51.10**
WS*	5.69*	44.60**

NPL, number of pairs of leaflets; MDBL, mean distance between leaflets; LML, length of medium leaflet; WML, width of medium leaflet; L/WML, length/width of medium leaflet; LL, length of leaf; WL, width of leaf; L/WL, length/width of leaf; LF, length of fruit; WF, width of fruit; NSF, number of seeds per fruit; LS, length of seed; WS, width of seed.
ns, Not significant; $*P < 0.05$, $**P < 0.01$.

However, our contradicting results highlights that the investigated variability is interesting for management strategies.

In the analysis based on a spatial model for cluster membership, three clusters were inferred (figure 2). Estimated cluster membership is similar to spatial clusters simulated from prior model. One cluster exclusively grouped the individuals from the Paranaense biogeographic province, whereas in the Yungas biogeographic province the individuals grouped regarding its population Tucumán and Jujuy, respectively. Thus, individuals from the Paranaense biogeographic province belong to cluster 1 with a 45% of posterior probability, individuals from Tucumán belong to cluster 2 with a 44% of posterior probability and individuals from Jujuy belong to cluster 3 with a 48% of posterior probability.

Many studies pointed out the difficulty to correlate quantitative variation with geographical descriptors (Sagnard *et al.* 2002). However, several hypotheses have been proposed to explain this difficulty. One of them postulates the presence of an important amount of gene flow among populations, which counterbalances the effect of differential selection (Sagnard *et al.* 2002). In the present study, gene flow may explain the spatial pattern of phenotypic variability due to the moderate assignment of the individuals to their geographic origins. Especially, this is most likely for the populations from the Paranaense biogeographic province are geographically close to each other rather than Tucuman and Jujuy from the Yungas biogeographic province. In addition, these populations are subject to similar management as they are located close to human settlements. On the other hand, Tucuman and Jujuy are located in protected areas, and the populations are separated by a long distance.

Molecular genetic variability

Null alleles were observed in loci *Ac28.3* and *Ac172.1* in the Paranaense biogeographic province. Also, locus *Ac48.1* showed null alleles in the Yungas biogeographic province while locus *Ac11.2* showed it in both biogeographic provinces (table 3). Feres *et al.* (2012) developed 20 SSR specific loci for *A. colubrina* using 51 individuals from two Brazilian populations. They reported 14 out 20 polymorphic loci while total number of polymorphic loci showed null alleles. Three out of eight loci showed a significant departure from the expected proportions of HW equilibrium at the level of 95% in the Paranaense biogeographic province while four out of eight loci showed departure from expected proportions in the Yungas biogeographic province (table 3). Deviations from HWE were primarily due to the presence of null alleles since all loci with null alleles were not in HWE. On the other hand, Barrandeguy *et al.* (2012) analysed the same loci that in the present work but these authors did not detected null alleles and all loci did not significantly differ from HWE in their analysed samples.

For each locus, effective number of alleles (N_e) was lower than number of alleles (N_a) (table 3). The discrepancy is explained by uneven frequencies among alleles
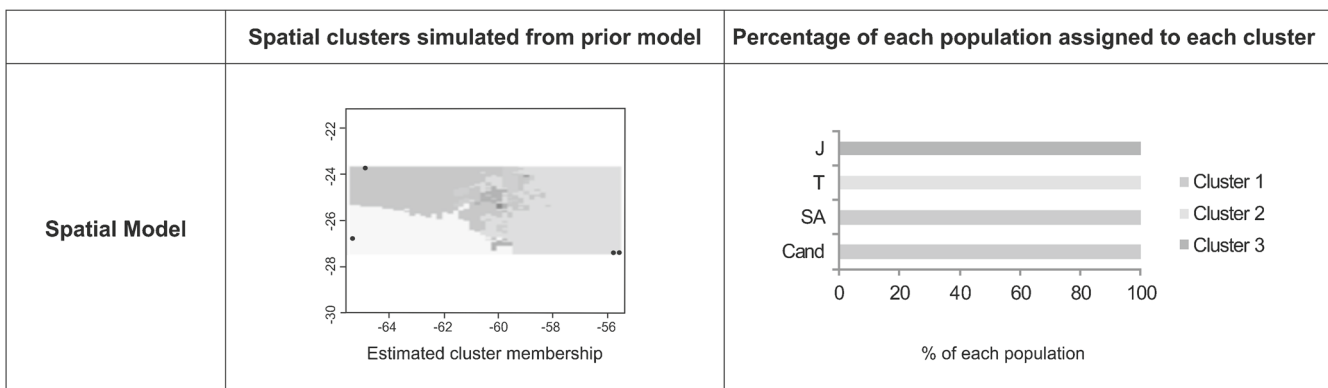

Figure 2. Bayesian clustering for phenotypic variability among individuals of *A. colubrina* var. *cebil* of the Yungas and Paranaense biogeographic provinces.

Table 3. Characterization of the eight microsatellite loci in all individuals of *A. colubrina* var. *cebil*.

| | Biogeographic province | | | | | | | | | | |
| | Paranaense | | | | | Yungas | | | | | |
Locus	N	N_a	N_e	NA	P HWE	N	N_a	N_e	NA	P HWE	F_{ST}
Ac34.3	36	13.00	9.35	0.00	0.658	32	13.00	9.89	0.03	0.211	0.025*
Ac48.1	36	12.00	5.82	0.05	0.418	27	15.00	9.22	0.26	0.000	0.028*
Ac11.2	36	6.00	3.80	0.25	0.000	33	6.00	4.18	0.19	0.002	0.045*
Ac28.3	36	13.00	7.37	0.32	0.000	26	20.00	11.75	0.04	0.021	0.072*
Ac157.1	36	22.00	9.15	0.00	0.080	32	15.00	5.68	0.03	0.790	0.038*
Ac41.1	35	14.00	3.20	0.00	0.555	33	11.00	3.65	0.00	0.372	0.215*
Ac172.1	36	12.00	3.94	0.20	0.003	33	15.00	10.52	0.03	0.024	0.091*
Ac162.1	36	14.00	5.67	0.00	0.363	28	16.00	8.76	0.02	0.195	0.013ns

N, total number of samples analysed for each locus; N_a, number of alleles per locus; N_e, number of effective alleles per locus; NA, frequency of null alleles; P HWE, exact P value associated with the null hypothesis of HWE; F_{ST}, fixation index.
ns, not significant, *$P < 0.05$.

indicating the presence of several alleles at low frequency. Locus *Ac41.1* showed the highest discrepancy between N_e and N_a, and accordingly, this locus also displayed the highest value of F_{ST}. However, all loci showed statistically significant F_{ST} values (table 3). Twelve out 14 polymorphic loci analysed by Feres *et al.* (2012) showed statistically significant F_{ST} values. The parameters of genetic variability (N_a, N_e, R) showed higher values in the Yungas biogeographic province than in the Paranaense biogeographic province, despite N being lower in the former populations (table 4).

In both provinces, effective number of alleles was lower than number of alleles, also indicating the presence of several alleles in low frequency over all loci. Values of F_{ST} with and without the ENA correction were similar (mean: 0.076, CI: 0.054–0.107 and mean: 0.078, CI: 0.053–0.110, respectively). Observed heterozygosity was lower than expected heterozygosity in both biogeographic provinces (table 4), and an excess of homozygosity may be explained as consequence of null alleles in some loci. In addition, historic and demographic changes are considered for these populations highly influencing the genetic variability.

The Yungas biogeographic province showed higher molecular genetic variability than the Paranaense biogeographic province. Genetic differences according to species distribution was also observed in populations of *Astronium urundeuva* (Caetano *et al.* 2008), a species distributed similarly to *A. colubrina* var *cebil*. In agreement with our results, these authors found high genetic diversity for *A. urundeuva*

in the Yungas biogeographic province proposing some weak recent isolation of this province.

The homozygosity test by Watterson and the exact test by Slatkin resulted in a P value of 0.05 suggesting diversifying selection for locus *Ac41.1*, whereas the rest of the loci can be considered neutral after performing both neutrality tests. Assumed nonneutrality of locus *Ac41.1* might be owing to the spatial distribution of one particular allele. This allele is present in 27 out of 35 individuals in the Paranaense biogeographic province while it is present in only three individuals in the Yungas province.

Bayesian clustering models for microsatellites loci data were applied for all loci and for all excluding locus *Ac41.1* as a consequence of the putative nonneutral behaviour. In both analyses based on a spatial model for cluster membership simulated from prior model, two clusters were inferred grouping the individuals corresponding to their geographic provinces of origin (figure 3). Even though each individual showed relative moderate posterior probability for belonging to different cluster, the estimation of cluster membership was similar to spatial clusters simulated from prior model.

Distribution of genetic variability between and within these clusters was analysed. An AMOVA for all haplotypes defined by all loci attributed most of the total variance to genetic variation within clusters (91.67%) while variation between clusters was only 8.33%. An AMOVA for all haplotypes defined by all loci excluding locus *Ac41.1*

Table 4. Estimates of genetic parameters in populations of *A. colubrina* var. *cebil* of the Paranaense and Yungas biogeographic provinces averaged over all loci.
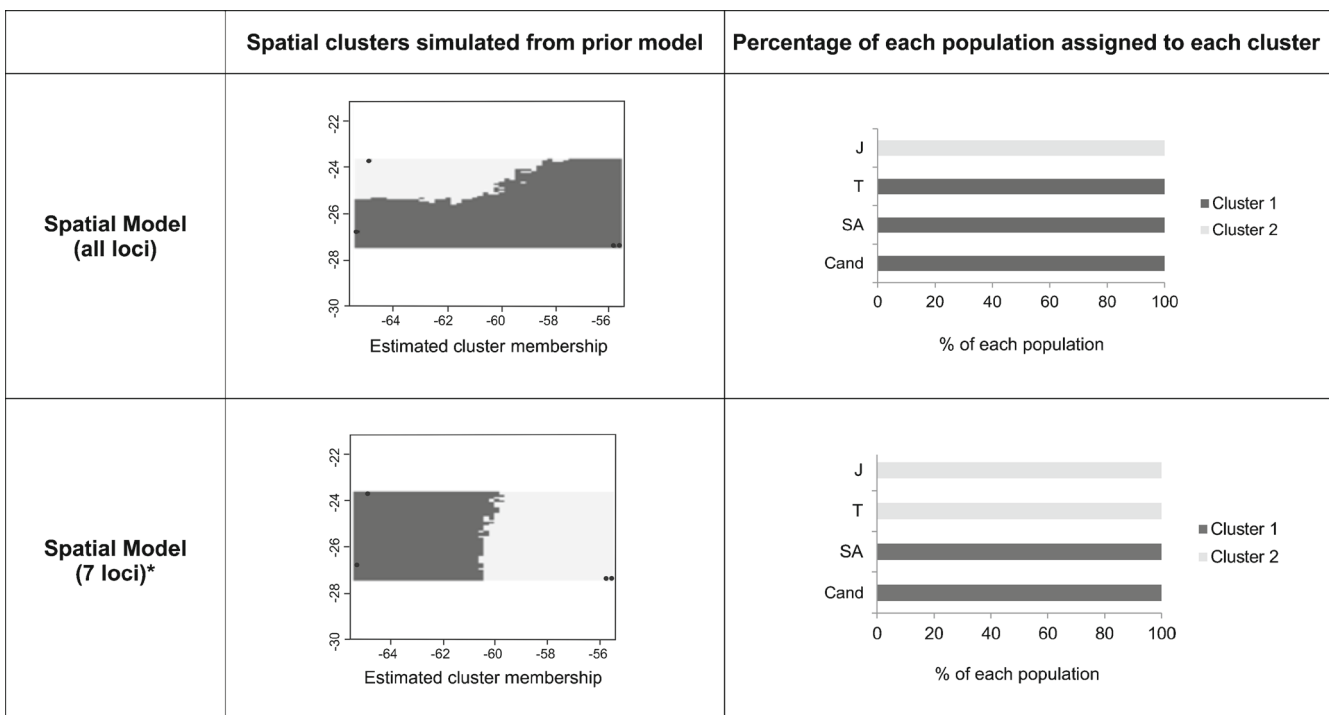
Biogeographic provinces	N*	N_a*	N_e*	H_o*	H_e*	R*
Paranaense	35.87 (0.12)	13.25 (1.54)	6.08 (0.85)	0.70 (0.07)	0.81 (0.02)	4.15 (0.96)
Yungas	30.50 (1.05)	13.87 (1.44)	7.96 (1.07)	0.75 (0.05)	0.85 (0.02)	5.65 (1.01)

N, total number of samples analysed for each locus; N_a, number of alleles per locus; N_e, number of effective allele per locus; H_o, observed heterozygosity; H_e, expected heterozygosity; R, allelic richness. *Mean (±SE) average across all loci with standard error.
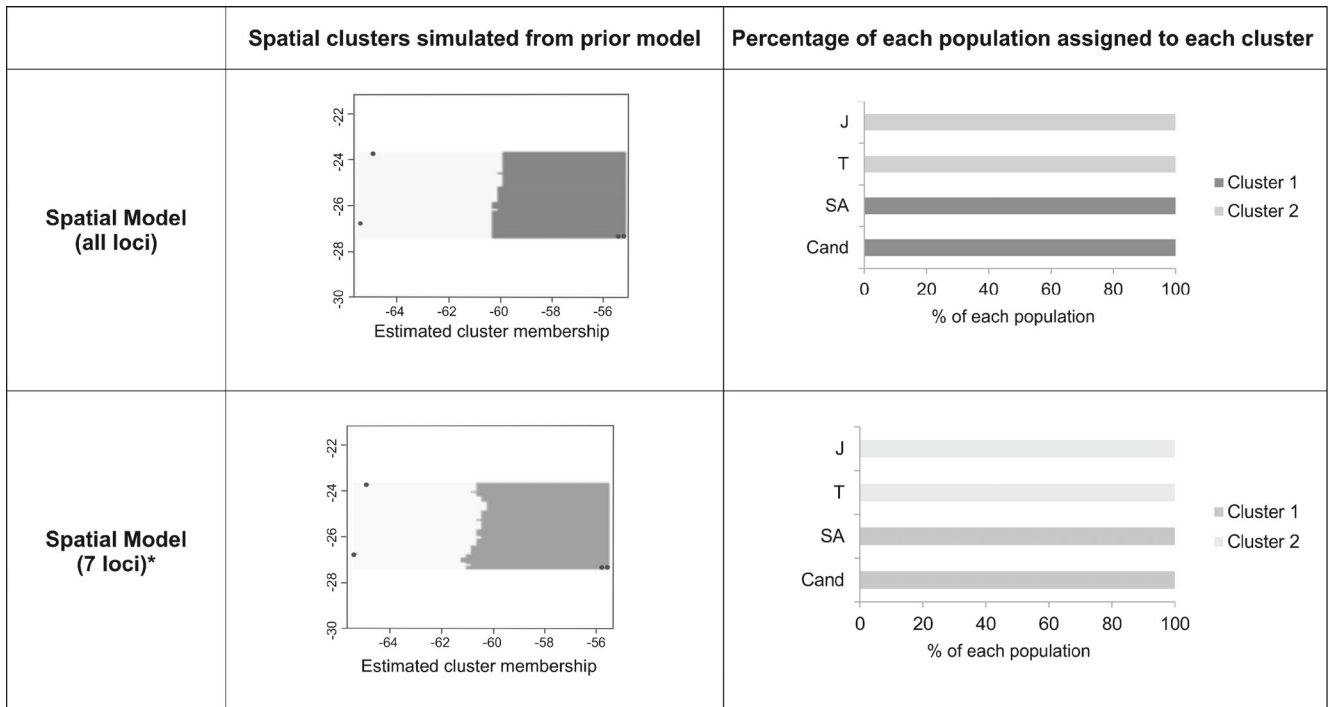
Locus *Ac41.1* was excluded from the analysis

Figure 3. Bayesian clustering for molecular genetic variability among individuals of *A. colubrina* var. *cebil* of the Yungas and Paranaense biogeographic provinces.

Locus *Ac41.1* was excluded from the analysis

Figure 4. Bayesian clustering for combined phenotypic and molecular genetic variability among individuals of *A. colubrina* var. *cebil* of the Yungas and Paranaense biogeographic provinces.

also attributed most of the total variance to genetic variation within clusters (94.99%) while variation between clusters was only 5.91%. For all loci, F_{ST} was 0.083 indicating moderate genetic differentiation of populations (Wright 1978), and similarly, F_{ST} was 0.050 for all loci excluding locus *Ac41.1*. Each biogeographic province retained a considerable amount of genetic variation, and they are differentiated from each other among individuals. Although not high, differentiation among the provinces is obviously as all individuals group to their provinces of origin. In addition, putative nonneutral behaviour of locus *Ac41.1* not affected the distribution of genetic variation.

Combination of phenotypic and molecular genetic variability

The presence of homogeneous clusters was detected for each of the phenotypic and molecular genetic variability. Bayesian clustering analyses based on a spatial model for cluster membership revealed clusters of individuals assigned to their provinces of origin. A combined Bayesian clustering analysis of both the phenotypic and genetic data for all loci revealed two groups comprising in one cluster the individuals from Paranaense biogeographic province and Tucumán whereas the other cluster grouped the individuals from Jujuy (figure 4). Individuals from Jujuy grouped in a similar way than in Bayesian clustering analysis of quantitative group. Even this individuals grouping could suggested certain relationship between locus *Ac41.1* profile and phenotypic clusters we cannot prove a specific linkage of the marker to some phenotypic trait, and several genes as well as environmental impacts may be involved in phenotypic expression (e.g., Vieira *et al.* 2007).

Conclusion

An assessment of variability is essential to understand the current gene pool and to design appropriate conservation strategies. A maximum of populations should be preserved prioritizing those which are more diverse and divergent through different environments (Inza *et al.* 2012). In our study, the genetic variability was high either for quantitative traits or molecular markers. Even though the analysis showed the phenotypic and genetic variability within provinces, they differed from each other at the individual level. Populations in the Yungas biogeographic province showed higher molecular genetic variation than populations in the Paranaense biogeographic province making them especially important for conservation. The populations in the Yungas biogeographic province are currently located in protected areas. However, these protected areas may be insufficient and additional areas are necessary for conservation, also in this biogeographic province. For example, *Cedrela lilloi* populations not included in protected areas showed high diversity and genetic variation that is absent in the rest of the region (Inza *et al.* 2012). Populations in the Paranaense

biogeographic province are more affected by human activities. However, these populations also contain a huge amount of phenotypic variability and showed the highest mean value of NSF. This is equally important for the creation of management strategies to recover the adversely affected areas in this biogeographic province.

To our knowledge, selection tests have never been applied in microsatellites studies of native forest populations from SDTF. This is explained by the missing SSR marker availability for nonmodel species such as *Anadenanthera colubrina* var. *cebil*. However, an increasing number of SSR loci allowing for tests of neutrality will be necessary for future microsatellite studies of native forest trees.

References

Abraham de Noir F., Bravo S. and Abdala R. 2002 Mecanismos de dispersión de algunas especies leñosas nativas del chaco occidental y serrano. *Que. Rev. Ciencias For.* **9**, 140–150.

Barrandeguy M. E., Prinz K., García M. V. and Finkeldey R. 2012 Development of microsatellite markers for *Anadenanthera colubrina* var. *cebil* (Fabaceae) a native tree from South America. *Am. J. Bot.* e372–e374.

Brookfield J. F. Y. 1996 A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol. Ecol.* **5**, 453–455.

Caetano S., Prado D., Pennington T., Beck S., Oliveira-Filho A., Spichiger R. and Naciri Y. 2008 The history of seasonally dry tropical forests in eastern South America: inferences from the genetic structure of the tree *Astronium unrundeuva* (Anacardiaceae). *Mol. Ecol.* **17**, 3147–3199.

Chapuis M. and Estoup A. 2007 Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* **24**, 621–631.

Cialdella A. M. 2000 Flora Fanerogámica Argentina, Fasciculo 67: Fabaceae subfamilia Mimosoideae, pp. 1–10. Proflora-CONICET, Córdoba, Argentina.

Feres J. M., Monteiro M., Zucchi M. I., Pinheiro J. B., Mestriner M. A. and Alzate-Marin A. L. 2012 Development of microsatellite markers for Anadenanthera colubrina (Leguminosae), a neotropical tree species. *Am. J. Bot.*, e154–e156.

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

Gillet E., Gömöry D. and Paule L. 2005 Measuring genetic variation within and among populations at marker loci. In *Conservation and management of forest genetic resource in Europe* (ed. T. Geburek and J. Turok) pp. 237–270. Arbora Publishers, Zvolen, Slovakia.

Guillot G., Renaud S., Ledevin R., Michaux J. and Claude J. 2012 A unifying model for the analysis of phenotypic, genetic and geographic data. *Syst. Biol.* **61**, 897–911.

Holderegger R., Kamm U. and Gugerli F. 2006 Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Genet.* **21**, 797–807.

Holsinger K. E. and Weir B. S. 2009 Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nat. Rev.* **10**, 639–650.

Inza M. V., Zelener N., Fornes L. and Gallo L. A. 2012 Effect of latitudinal gradient and impact of logging in genetic divertisy of *Cedrel lilloi* along the Argentina Yungas Rainforest. *Ecol. Evol.* **2**, 2722–2736.

Justiniano M. J. and Fredericksen T. S. 1998 *Ecología y Silvicultura de especies menos conocidas: Curupaú. Anadenanthera colubrina* (Vell.) Benth. Mimosoideae. Proyecto de Manejo Forestal Sostenible. pp. 1–30, Santa Cruz, Bolivia.

Languth E. L. and Balkenhol N. 2012 Relative sensitivity of neutral versus adaptive genetic data for assessing population differentiation. *Conserv. Genet.* **13**, 1421–1426.

Lazrek F., Roussel V., Ronfort J., Cardinet G., Chardon F., Aouani M. E. and Huguet T. 2009 The use of neutral and non-neutral SSRs to analyse the genetic structure of a Tunisian collection of *Medicago truncatula* lines and to reveal associations with eco-environmental variables. *Genetica* **135**, 391–402.

Nielsen E. A., Hansen M. M. and Meldrup D. 2006 Evidence of microsatellite hitch-hiking selection in Atlantic cod (*Gadus morhua* L.): implications for inferring population structure in nonmodel organisms. *Mol. Ecol.* **15**, 3219–3322.

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

Pennington R. T., Prado D. E. and Pendry C. A. 2000 Neotropical seasonally dry forests and Quaternary vegetation changes. *J. Biogeogr.* **27**, 261–273.

Prado D. E. 2000 Seasonally Dry Forest of Tropical South America: from forgotten ecosystems to a new phytogeography unit. *Edinburgh J. Bot.* **57**, 437–461.

Prado D. E. and Gibbs P. E. 1993 Patterns of species distributions in the dry seasonal forests of South America. *Ann. Mo. Bot. Gard.* **80**, 902–927.

Rousset F. 2008 Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–106.

Sagnard F., Barberot C. and Fady B. 2002 Structure of genetic diversity in *Abies alba* Mill. from southwestern Alps: multivariate analysis of adaptive and non-adaptive traits for conservation in France. *For. Ecol. Manage.* **157**, 175–189.

Shi M. M., Michalski S. G., Chen X. Y. and Durka W. 2011 Isolation by elevation: Genetic structure at neutral and putatively non-neutral loci in a dominant tree of subtropical foresta, *Castamopsis eyrei*. *PLoS One* **6**, e21302.

Slatkin M. 1994 An exact test for neutrality based on the Ewens sampling distribution. *Genet. Res.* **64**, 71–74.

Slatkin M. 1996 A correction to the exact test based on the Ewens sampling distribution. *Genet. Res.* **68**, 259–260.

Sokal R. R. and Rohl F. J. 1980 *Biometry*. Freeman, New York, USA.

Szpiech Z. A., Jakobsson M. and Rosenberg N. A. 2008 ADZE: A rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**, 2498–2504.

Van Oosterhout C., Hutchinson W. F., Wills D. P. M. and Shipley P. 2004 MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**, 535–538.

Vieira E. A., Irajá F, de Carvalho F., Bertan I., Kopp M. M., Zimmer P. D. *et al.* 2007 Association between genetic distances in wheat (*Triticum aestivum* L.) as estimated by AFLP and morphological markers. *Genet. Mol. Biol.* **30**, 392–399.

Watterson G. 1978 The homozygosity test of neutrality. *Genetics* **88**, 405–417.

Wee A. K. S., Li C. H., Dvorak W. S. and Hong Y. 2012 Genetic diversity in natural populations of *Gmelina arborea*: implications for breeding and conservation. *New For.* **43**, 411–428.

Wheeleer D. C. and Waller L. A. 2012 Spatial analysis linking landscape features and genetic population structure in cougars (Puma concolor) in the northern Rocky mountains. *Stat. Interface* **5**, 183–193.

Wright S. 1978 *Evolution and the genetics of populations*. University of Chicago Press, Chicago, USA.

Zobel B. and Talbert J. 1991 *Applied forest tree improvement*. Waveland, New York, USA.