Tree Genetics & Genomes (2015) 11: 46
DOI 10.1007/s11295-015-0876-8

ORIGINAL PAPER

Population genetics of the naturally rare tree *Dimorphandra wilsonii* (Caesalpinioideae) of the Brazilian Cerrado

Christina C. Vinson[1] · Thaís C. S. Dal'Sasso[1] · Claudia P. Sudré[2] · Erica Mangaravite[1] · Luiz O. de Oliveira[1]

Received: 6 June 2014 / Revised: 22 March 2015 / Accepted: 16 April 2015 / Published online: 29 April 2015

Abstract Naturally rare species have a higher probability of stochastic extinction due to genetic, demographic, or environmental hazards; human disturbance may intensify these threats. Rare species may therefore be in need of short-term intervention to survive. The ecosystem with the second highest biodiversity in Brazil, the Cerrado, is suffering from fragmentation and threats to its flora. *Dimorphandra wilsonii*, a 30-m tall endemic tree of the Brazilian Cerrado, is listed as critically endangered; only 21 adult trees have been identified. We carried out mating system and pollen flow analyses to understand the current gene flow and limitations in the reproduction of *D. wilsonii*. With seven fluorescently labelled microsatellite primers, we genotyped 20 adult trees and 269 progeny from 13 mother trees. *D. wilsonii* displayed low levels of genetic diversity; bottleneck events are likely to have occurred (H_e=0.60 and 0.29; H_o=0.71 and 0.33, for adults and progeny, respectively). This species is predominantly outcrossing (t_m=0.88), with some selfing (1-t_m=0.12), as well as crossing between related individuals (t_m-t_s=0.11). None of the studied trees was reproductively isolated; a high proportion of pollen (55 %) came from trees yet to be discovered. Two genetic clusters (Northern and Southern) were identified, with high values of genetic divergence among the Southern sites. Planting of seedlings and monitoring of seed dispersion in order to maintain the genetic diversity and genetic structure of *D. wilsonii* are strategies that may ensure the continuation of *D. wilsonii*, but this species does not seem to require reproductive intervention to remain viable.

Keywords Cerrado · Conservation · *Dimorphandra* · Mating system · Microsatellite · Rare species

Communicated by D. Grattapaglia

✉ Luiz O. de Oliveira
luiz.ufv@hotmail.com

[1] Laboratorio de Biologia Molecular e Filogeografia, Instituto de Biotecnologia Aplicada à Agropecuária, Universidade Federal de Viçosa, Viçosa, MG 36570-000, Brazil

[2] Centro de Ciências e Tecnologias Agropecuárias, Universidade Estadual do Norte Fluminense "Darcy Ribeiro", Campos dos Goytacazes, RJ, Brazil

Introduction

Plant species may be naturally rare for several reasons, including specialised habitat requirements, the absence of long-distance pollen and seed dispersal, or a low reproductive rate (Barrett and Kohn 1991; Karron 1991). Rare species can exhibit different population sizes, genetic structure or geographical distributions. They comprise three classes (Karron 1991): (1) sparse species with a large geographical distribution but a low density at the local level, (2) species with specialised habitat requirements, and (3) narrow endemic species with one to five populations. Most naturally rare species, including *Dimorphandra wilsonii* belong to the third class.

Naturally rare species suffer greatly from the evolutionary forces that affect the dynamics of plant populations, which include the loss of genetic variation (Barrett and Kohn 1991; Karron 1991; Hensen and Oberprieler 2005). This reduced level of genetic variation can be due to (1) change in allelic frequencies due to genetic drift or founder events associated with recent speciation or bottlenecks and (2) natural selection towards uniformity, inbreeding and selection against rare alleles (Karron 1991). Meta-analysis studies suggest that the levels of genetic variation are related to population size; overall, small populations contain significantly less genetic variation than large ones (Leimu et al. 2006; Honnay and

Jacquemyn 2007; Aguilar et al. 2008). A meta-analysis of isozyme variation by 57 comparisons at the genus level showed that values for genetic population parameters were significantly lower in rare species than in common ones (Cole 2003). Therefore, it is reasonable to expect rare species to exhibit significantly lower levels of genetic variation than their common congeners (Karron 1991; Cole 2003).

Genetic variability within populations is an important factor to consider as it is associated with concerns over species' persistence during episodes of abiotic and biotic environmental changes (Simberloff 1988; Schaberg et al. 2008). Loss of alleles and erosion of heterozygosity can reduce the diversity of allelic combinations, and therefore potentially lead to the loss of adaptability and environmental fitness (Simberloff 1988; Schaberg et al. 2008). The relationships between population size, genetic variation, and fitness are highly related to plant characteristics, especially life span, mating system, and rarity (Leimu et al. 2006). Small populations are particularly prone to loss of genetic variation and inbreeding; however, this loss is potentially counteracted as long as there is gene flow among populations. Furthermore, the gene flow is highly dependent on specific factors, such as the dispersal efficiency (pollen and seed dispersal, long-distance versus short-distance dispersal), breeding system (hermaphrodite, monoecious or dioecious), pollinator behaviour, outcrossing rate and the degree of geographic isolation (Hensen and Oberprieler 2005).

Habitat fragmentation is one of the greatest threats to biological diversity worldwide; increasing anthropogenic demand for space and the exploitation of the land's resources bring about rapid habitat fragmentation. The Cerrado, the largest savannah in South America and one of Brazil's most threatened ecosystems, lost about 21,000 km^2 of its natural vegetation annually between 2002 and 2008; this deforestation rate was twice that of the Amazon rainforest (Myers et al. 2000). Such high rates of deforestation and fragmentation are major concerns for the survival of many of the Cerrado's endemic species. This is exacerbated by the fact that the Cerrado is a hotspot that harbour more than to 10, 000 species of plants, of which nearly half are endemic. The chief concern is that, when a new species is found, it is already at risk of extinction.

The genus *Dimorphandra* Schott (Fabaceae, Caesalpinioideae) contains 26 species and four subspecies; the geographic range of *Dimorphandra* comprises Central and South America. Although the genus is widespread, the distribution of most of its species is very restricted, leading to the possibility that there are additional naturally rare species of *Dimorphandra* yet to be described (Silva 1986). Within this genus, *D. wilsonii Rizzini*, known locally as "faveiro de Wilson", is a geographically narrow endemic species of the Brazilian Cerrado. It was first described in 1969 when 11 trees were identified (Rizzini 1969; Silva 1986). Recently, the Belo Horizonte Zoo-Botanic Foundation carried out intensive searches for *D. wilsonii*; only 21 trees are known to exist, near Belo Horizonte and neighbouring cities, in the central region of the State of Minas Gerais, Brazil (Fernandes et al. 2007; Viana e Souza and Lovato 2010). *Dimorphandra wilsonii* is considered to be seriously threatened with extinction; the species has been listed since 2006 as critically endangered (IUCN 2013, IUCN Red List of Threatened Species, version 2013.1, http://www.iucnredlist.org). Belo Horizonte and its neighbouring cities have grown extensively over the last 200 years, which has increased the threat of habitat destruction and fragmentation. Cerrado-disturbing activities, such as clearing and logging for charcoal production, conversion to agriculture and pasture, and *Eucalyptus* cultivation, also contribute to reduction of the natural habitat of *D. wilsonii* (Fernandes et al. 2007). *Dimorphandra wilsonii* has an annual flowering cycle, flowering from December to February and producing mature fruit from July to October (Fernandes et al. 2007). Although *D. wilsonii* pollinators and dispersers have not been described, the common congener species *D. mollis* is pollinated by small insects and its seeds are dispersed by tapirs (Bizerril et al. 2005). In addition, inter simple sequence repeat markers have shown low genetic diversity among the adult trees of *D. wilsonii* and high population differentiation compared with that of *D. mollis* (Viana e Souza and Lovato 2010); however, mating system studies have not been carried out.

In this study, we analysed the population genetics and mating system of the naturally rare tree species *D. wilsonii*. We address the following questions: (a) What levels of genetic diversity and genetic structure does *D. wilsonii* possess? (b) What are the estimates for outcrossing and mating between related individuals? Are geographically isolated trees of *D. wilsonii* reproductively isolated? (c) What are the implications of these findings for the conservation of the remaining individuals of *D. wilsonii*? Our results should shed light on the population genetics of a rare narrow endemic tree of the Cerrado and contribute to the implementation of management strategies for in situ and ex situ conservation of *D. wilsonii*.

Methodology

Study site and sample collection

The Brazilian Institute of Environment and Renewable Natural Resources (IBAMA) provided us with data on the location of the trees (Fig. 1, Table 1). A total of 20 adult trees, 20–40 fruit from 13 trees, and 10 regenerants of *D. wilsonii* were sampled (Table 1). Amongst the regenerants, one had a natural origin; the other nine had been planted by the Belo Horizonte Zoo-Botanic Foundation as part of a conservation program (Fernandes et al. 2007). For mating system studies,

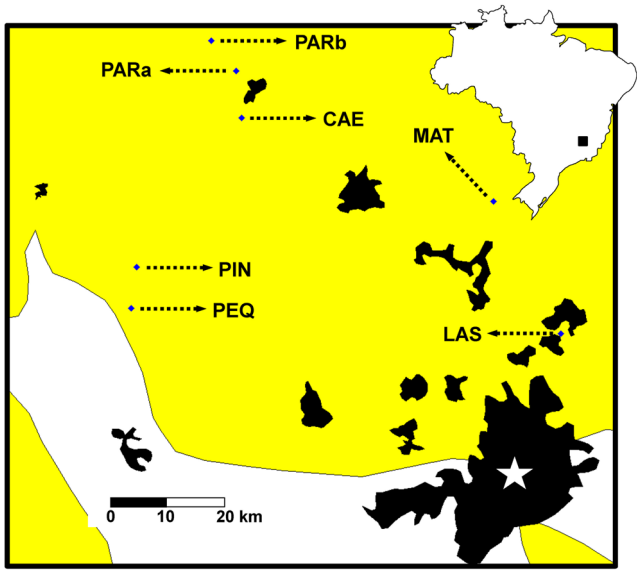

Fig. 1 Geographic distribution and sampling locations of adult trees and progeny of *Dimorphandra wilsonii*. *Yellow*, Cerrado; *white*, seasonal forests; *star*, Belo Horizonte, the capital city of the State of Minas Gerais; contours of neighbouring cities are given in *gray* for reference purposes. Please refer to Table 1 for codes

fruit were sampled from apparently reproductively isolated trees at three sites: Caetanópolis (CAE), Pindaiba (PIN) and Paraopeba (PARb). Although PARb had two trees, one of them was clearly dying and had only a few leaves left; therefore, we assumed that the remaining living tree at PARb was reproductively isolated.

The sampling methods for adults, regenerants and progeny varied as follows: adults and regenerants—all trees were sampled by removing a small amount of leaf tissue that was later dried in silica gel; progeny—a total of 20–40 fruit per 'mother tree' were collected directly from all trees that displayed fruiting in that season (Table 1); fruit were collected from different branches around the tree in order to achieve representative sampling. Seeds were planted in a nursery; when seedlings developed at least two leaflets, leaf tissue was taken and DNA was extracted; for genotyping, we used only one seedling per fruit.

Laboratory experiments

DNA was extracted using a 2 % cetyl trimethylammonium bromide (CTAB) protocol (Riahi et al. 2010), adapted for tropical forest species. Forward primers of seven loci were fluorescently labelled with 6-FAM, HEX (MWG-Biotech) and NED (Applied Biosystems). The polymerase chain reaction (PCR) protocols were performed in accordance with Vinson et al. (2013). The sizes of the fragments were determined on a 96 capillary sequencer, ABI PRISM 3130xl DNA Analyser (Applied Biosystems), using Applied Biosystems' GeneScan 500 Rox size standard. The fragments were scored using GeneMapper v. 4.0 software (Applied Biosystems).

Genetic data analyses

Genetic diversity We defined the following arrays: trees (all trees assessed, trees from each site), progeny (all progeny assessed, progeny from each mother as a group: total of 13 families) and regenerants (all regenerants assessed). Genetic diversity analysis was performed using the software GDA (Lewis and Zaykin 2001); total number of alleles (A_t), average number of alleles (A), number of private alleles (A_p), expected heterozygosity (H_e), observed heterozygosity (H_o) and fixation index (F) were calculated. The unpaired t test ($P<0.05$) was used to compare the number of alleles, H_e and H_o between progeny and adult samples. For F_{ST} analysis, we defined the following arrays: trees (all trees assessed), progeny (all progeny assessed) and regenerants (all regenerants assessed). For F_{ST} analysis between progeny, we defined the following arrays: between families (progeny from each mother as a group: total of 13 families), between sites (progeny from each site as a group: seven sites) and two clusters found in STRUCTURE analysis (Northern and Southern; see Results). We estimated F_{ST} and Hedrick's G_{ST} (Hedrick 2005) using FSTAT 2.9.3.2 (Goudet 1995). The frequency of null alleles and scoring errors were estimated using the software MICRO-CHEKER 2.2.3 (Oosterhout et al. 2004).

Table 1 Sampled individuals of *Dimorphandra wilsonii* with abbreviations, sample size and geographic location

Site	Number of trees	Number of mother trees	Code of progeny	Number of regenerants	Geographic coordinates
Paraopeba	7	7	PARa	9	S19 25.6593, W44 43.2738
Paraopeba	2	1	PARb	0	S19 12.456, W44 28.413
Caetanópolis	1	1	CAE	1	S19 19.846, W44 25.449
Pindaiba	1	1	PIN	0	S19 34.174, W44 35.644
Pequi	2	1	PEQ	0	S19 38.539, W44 36.207
Matozinhos	3	1	MAT	0	S19 46.4357, W44 01.6324
Lagoa Santa	4	1	LAS	0	S19 40.559, W43 54.434

Genetic structure For the analyses, we defined the following array: trees (all trees assessed), progeny (progeny from each mother as a group: total of 13 families) and regenerants (all regenerants assessed). To assess genetic population structure, a Bayesian Markov chain Monte Carlo (MCMC) approach was performed with all 269 progeny, adult trees and regenerants using STRUCTURE v. 2.2 (Pritchard et al. 2000). We used an admixture model and correlated allele frequencies among populations according to Falush et al. (2003). To obtain the best K, the criteria used were both ln[Pr (X | K)] and ΔK (Evanno et al. 2005) using Structure Harvester (Earl and VonHoldt 2011). We tested the number of populations (K) in the range of 1 to 18. For the MCMC, we set the number of runs to 750,000, after a burn in period of 250,000, and 20 independent runs for each value of K.

Mating system analysis and paternity analysis The mating system of *D. wilsonii* was analysed using data from 269 progeny from 13 trees. Analyses used both mixed mating (Ritland and Jain 1981) and correlated mating (Ritland 1989) models in the program MLTR version 3.2 (Ritland 2002). Parameters were estimated using the maximum expectation method. Standard errors of the parameters were calculated from 500 bootstraps, using families as the resample unit. The effective number of pollen donors was calculated as $N_{ep}=1/r_{p(m)}$, where $r_{p(m)}$ is the paternity multilocus correlation value. The genetic structure of the progeny was estimated using the average coefficient of coancestry (θ_{xy}) between individuals within families according to Carneiro et al. (2011). The variance effective population size ($Ne_{(v)}$) and the number of seed-trees to be sampled to obtain a reference effective population size (m) of 150 trees were calculated in accordance with Carneiro et al. (2011). Paternity analysis was performed by maximum likelihood assignment using the program Cervus 3.0 (Marshall et al. 1998; Kalinowski et al. 2007) using the following inputs: 10,000 simulations, mistyped rate of 0.01, sampled candidate parent proportion of 0.5–0.7, confidence values of 80 and 95 %, and a minimum of four typed loci. All trees from the same county (maximum distance between trees were 20 km) were considered probable candidate parents in the paternity analysis, and selfing was also considered. Therefore, four inputs of progeny, known mothers and candidate fathers of each county were run together: (1) PARa, PARb and CAE; (2) PIN and PEQ; (3) MAT; and (4) LAS) (see Tables 1 and 2).

Results

Genetic diversity and genetic structure

The highly polymorphic nature of microsatellite markers makes them a very effective tool to uncover details about population genetic structure, gene flow and parentage. The number of alleles varied greatly among loci; it ranged from three (Dw33) to eight (Dw105) alleles per locus for the adults, and from four (Dw78) to nine (Dw105) for the progeny. After Bonferroni correction, the test of pairwise linkage showed values higher than the significant threshold value of 0.005, which indicates that these loci are not linked; they can therefore be used as independent markers. Null alleles occurred only in locus Dw17 (for adults) and in locus Dw52 (for PARa1 progeny), which means that our data can be used for mating system analysis. Overall, microsatellite markers showed levels of polymorphism that allowed for the distinction among sites and among most trees; PARa was the only exception, as we identified only three genotypes among the seven adult trees.

Table 2 Distance (km) between sites of sampled individuals of *Dimorphandra wilsonii*

	PARa	PARb	CAE	PIN	PEQ	MAT	LAS
PARa	0						
PARb	9.04	0					
CAE	7.02	14.9	0				
PIN	39	34.31	44.21	0			
PEQ	46.41	40.31	50.21	20.2	0		
MAT	49.06	45.83	55.44	62.41	64.66	0	
LAS	73.9	67.77	81.5	72.4	73.67	27.26	0

Adults showed a low level of allelic diversity (a total of 32 alleles for seven loci, average of 2.04 alleles per locus; Table 3). The observed heterozygosity was higher than the expected heterozygosity (H_e=0.56, H_o=0.68; Table 3), which resulted in a negative fixation index. There was a significant difference in the number of alleles and the expected and observed heterozygosity between progeny and adults ($P<0.05$). The total numbers of alleles were 32 and 43, and the average numbers of alleles were 4.57 and 6.14 for adults and progeny, respectively (Table 3). This indicates an influx of 11 distinct alleles through pollen flow from trees not recorded in our study (Table 4). Both adults and progeny had private alleles in most sites (Table 3). On average, F was negative for progeny and adults, which indicated a proportion of heterozygotes higher than expected under Hardy-Weinberg equilibrium (HWE). Progeny obtained from two sites—LAS and MAT—displayed fewer alleles than progeny from other sites with low values for observed heterozygosity and F close to zero (0.04 and 0.00, respectively), which suggested HWE (Table 3).

The analysis of adults, regenerants and progeny with STRUCTURE (Fig. 2) revealed two Bayesian groups (K= 2); these groups were congruent with the geographic location. The Northern group (depicted in green in Fig. 2) contained the seven progeny from PARa, along with PARb and CAE. The

Table 3 Genetic diversity of *Dimorphandra wilsonii* estimated from seven microsatellite loci

	n	A_t	*A*	A_p	H_e	H_o	*F*
Adults							
PARa	7	15	2.14	6	0.46	0.78	−0.79
PARb	1.8	12	1.71		0.48	0.57	−0.50
CAE	1	11	1.57	1	0.57	0.57	0.00
PIN	1	12	1.71	1	0.71	0.71	0.00
PEQ	2	17	2.43		0.69	0.86	−0.41
MAT	3	15	2.57	1	0.58	0.57	0.02
LAS	4	18	2.14	2	0.46	0.71	−0.71
Average		14.3	2.04		0.56	0.68	−0.46
Total		32	4.57				
Progeny							
PARa1	20.43	18	2.57		0.47	0.58	−0.23
PARa2	23.43	22	3.14	1	0.50	0.65	−0.31
PARa3	32.86	25	3.57		0.45	0.60	−0.33
PARa4	19.28	22	3.14	1	0.47	0.62	−0.32
PARa5	17.00	19	2.71		0.44	0.59	−0.36
PARa6	17.00	21	3.00		0.47	0.56	−0.20
PARa7	20.14	22	3.14	3	0.47	0.53	−0.12
CAE	21.29	20	2.86	1	0.44	0.62	−0.42
PARb	19.29	21	3.00		0.48	0.59	−0.23
PIN	11.00	20	2.86	1	0.47	0.58	−0.24
PEQ	17.71	24	3.43		0.52	0.64	−0.24
MAT	14.86	17	2.43	2	0.42	0.40	0.04
LAS	11.86	17	2.43	1	0.40	0.41	0.00
Average		21.5	2.94		0.46	0.57	−0.23
Total		43	6.14				
Regenerants	9.28	20	3.14	0	0.53	0.65	−0.24

Please refer to Table 1 for codes

n number of individuals with locus successfully amplified, A_t total number of alleles, *A* average number of alleles, A_p number of private alleles, H_e expected heterozygosity, H_o observed heterozygosity, *F* fixation index

Southern group (depicted in red in Fig. 2) consisted of LAS and MAT, and included PEQ and PIN; the latter two progeny came from mother trees located at sites with an intermediate location (Fig. 1). There was some admixture between the two groups. In the Northern group, some seedlings contained a genetic component from the Southern group, especially those from PARa2, PARa4 and PARb. The genetic profile of the mother tree of PARb was mostly of Southern origin, which explained the admixture in some of the PARb seedlings. However, the presence of admixture in seedlings of PARa more likely arose due to gene flow through pollen donors from outside the site. In the Southern group, both MAT and its mother tree exhibited some degree of admixture, while PIN, PEQ and LAS and their respective mother trees showed little evidence of interbreeding. The only natural regenerant that we found was located at CAE; the regenerant and the adult tree exhibited similar genetic profiles. The nine regenerants that the Belo Horizonte Zoo-Botanic Foundation had planted at PARa, as part of a conservation program (Fernandes et al. 2007), exhibited varying degrees of admixture; these nine regenerants may have been obtained from the same site as the mother tree.

The values of genetic divergence (Θ_p (F_{ST}) among all adults, all progeny and all regenerants were close to zero, which indicated no difference among them; therefore, progeny and regenerants represent the adults well. There was significant genetic divergence (Θ_p (F_{ST}) and Hedrick's G_{ST}) for most progeny comparisons (Table 5). Among progeny, there was strong differentiation between families, between sites and between the two groups identified by the STRUCTURE analysis (Table 5). The values of genetic divergence were smaller between families than between sites, which indicated that the difference was not between mother trees of the same site; most likely, these differences were between sites. Comparisons of F_{ST} and Hedrick's G_{ST} between the two groups revealed by STRUCTURE showed a significant difference, which indicated that pollen clouds had distinct sources.

Mating systems and pollen flow

Dimorphandra wilsonii is an outcrossed species, with a mean outcrossing rate (t_m) of 0.88 (see Goodwillie et al. (2005); an outcrossed species has t_m higher than 0.8) (Table 4). Therefore, 12 % of all progeny resulted from selfing. The level of mating among relatives (biparental inbreeding), measured by the difference between single and multilocus outcrossing rates (t_m-t_s), indicated that 12 % of the crossings occurred between genetically related individuals (Table 4). The multilocus correlation of outcrossed paternity ($r_{p(m)}$) was 0.38, which suggested that progeny within the family had the same parent, meaning that they could be half-sibs (same father or mother) or self-sibs and self-half-sibs (selfing, same mother and father). In addition, only a few trees took part in the fertilisation process, with an average number of effective pollen donors (N_{ep}) of 2.61 father trees per mother tree. The father trees were not genetically related as there was no significant difference between $r_{p(m)}$ and $r_{p(s)}$ (Table 4). The mean estimated coancestry coefficient among individuals within progeny (θ_{xy}) was low at 0.062, which is below the half-sib value (0.125). The mean variance effective population size ($Ne_{(v)}$=1.061) was very low, so the number of seed-trees (m) for seed collection of 150 reference trees was high, with a mean value of 141.

In general, mating system parameters varied amongst mother trees. Most mother trees presented high values for the outcrossing rate; nevertheless, two mother trees (located in LAS or MAT; with t_m=0.50 and t_m=0.67, respectively, Table 4) presented noticeably lower outcrossing rates than all of the others, which had an impact on the observed heterozygosity and fixation index compared with those of the other

Table 4 Estimates of mating system and paternity analysis of 269 progeny from 13 trees of *Dimorphandra wilsonii*

Parameters	PARa1	PARa2	PARa3	PARa4	PARa5	PARa6	PARa7	PARb	CAE	PIN	PEQ	MAT	LAS	All
Progeny size	21	24	34	24	18	18	22	21	23	12	19	19	14	269
Mating system analysis														
Multilocus crossing rate: t_m	0.902 (0.011)	0.941 (0.016)	0.901 (0.021)	0.997 (0.002)	0.991 (0.004)	0.827 (0.045)	0.942 (0.004)	0.999 (0.002)	0.885 (0.003)	0.916 (0.000)	0.917 (0.008)	0.655 (0.016)	0.463 (0.014)	0.883 (0.046)
Single locus crossing rate: t_s	0.896 (0.015)	0.914 (0.012)	0.907 (0.023)	0.939 (0.016)	0.914 (0.014)	0.917 (0.020)	0.894 (0.011)	0.927 (0.017)	0.887 (0.018)	0.815 (0.035)	0.823 (0.028)	0.674 (0.117)	0.508 (0.123)	0.768 (0.087)
Mating among relatives rate: t_m-t_s	0.006 (0.010)	0.027 (0.012)	−0.006 (0.010)	0.058 (0.015)	0.078 (0.012)	−0.090 (0.035)	0.048 (0.009)	0.072 (0.016)	−0.002 (0.017)	0.100 (0.036)	0.094 (0.026)	−0.018 (0.111)	−0.045 (0.109)	0.116 (0.046)
Multilocus correlation of paternity $r_{p(m)}$	0.133 (0.169)	0.249 (0.221)	0.040 (0.116)	0.176 (0.190)	0.158 (0.183)	0.094 (0.143)	0.108 (0.151)	0.359 (0.272)	0.479 (0.333)	0.189 (0.209)	0.484 (0.339)	0.213 (0.220)	0.228 (0.263)	0.382 (0.091)
Single locus correlation of paternity $r_{p(s)}$	0.106 (0.153)	0.115 (0.158)	0.082 (0.135)	0.114 (0.158)	0.111 (0.160)	0.100 (0.147)	0.102 (0.147)	0.144 (0.182)	0.193 (0.229)	0.112 (0.169)	0.171 (0.205)	0.113 (0.153)	0.110 (0.150)	0.290 (0.071)
Crossing between pollination related trees $r_{p(s)}$-$r_{p(m)}$	0.027 (0.022)	0.134 (0.064)	−0.041 (0.020)	0.062 (0.033)	0.047 (0.026)	−0.006 (0.006)	0.006 (0.009)	0.215 (0.104)	0.286 (0.141)	0.077 (0.050)	0.313 (0.150)	0.100 (0.083)	0.119 (0.151)	0.092 (0.037)
Effective number of pollen donors: N_{ep}= $1/r_{p(m)}$	7.52	4.02	25.00	5.68	6.33	10.64	9.26	2.79	2.09	5.29	2.07	4.69	4.39	2.61
Genetic divergence of crossed pollen among mother trees: Φ_{fi}= $r_{p(s)}/2$	0.053	0.0575	0.041	0.057	0.0555	0.05	0.051	0.072	0.0965	0.056	0.0855	0.0565	0.055	0.123
Full-sibs proportion: P_{IC}= $rp(s)tm$	0.096	0.108	0.074	0.114	0.110	0.083	0.096	0.144	0.171	0.103	0.157	0.074	0.051	0.244
Half-sibs proportion: P_{MI}= $(1\text{-}t_m)$ $r_{p(s)}$	0.010	0.007	0.008	0.000	0.001	0.017	0.006	0.000	0.022	0.009	0.014	0.039	0.059	0.003
Coancestry coefficient within families: θ_{xy}	0.052	0.075	0.038	0.098	0.089	0.007	0.084	0.129	0.053	0.062	0.085	−0.102	−0.215	0.062
Variance effective size: $N_{e(v)}$	1.044	1.066	1.032	1.087	1.074	1.001	1.077	1.117	1.041	1.046	1.074	0.903	0.798	1.061
Number of seed-trees for seed collection: m	144	141	145	138	140	150	139	134	144	143	140	166	188	141
Paternity analysis														
Selfing	Na	Na	Na	Na	Na	Na	0.19	0.0	0.13	0.08	0.16	0.47	0.5	
Known father	0.57	0.67	0.53	0.5	0.61	0.5	0.29	Na	Na	Na	0.26	0.37	0.0	
Unknown father	0.43	0.33	0.47	0.5	0.39	0.5	0.52	1.00	0.87	0.92	0.58	0.16	0.5	0.55
Different alleles from mother tree	5	9	12	10	7	9	9	9	9	8	9	5	5	
Different alleles from all sampled trees	1	2	1	2	3	3	3	0	2	1	0	2	1	12

Refer to Table 1 for codes; standard errors in parentheses

Na not applicable

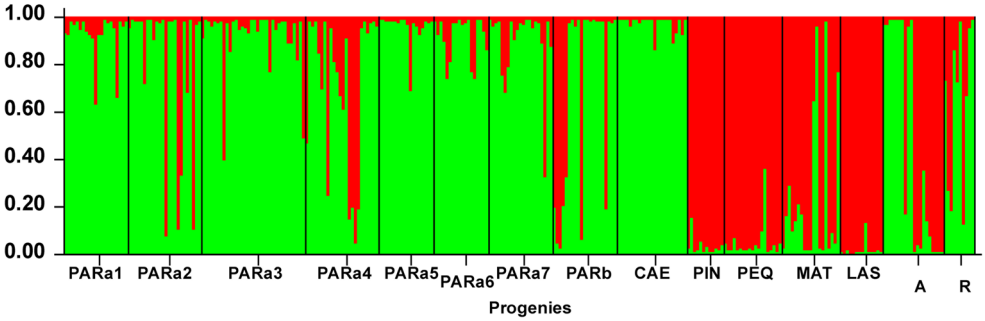

Fig. 2 Results of STRUCTURE analysis for *Dimorphandra wilsonii* using seven microsatellite loci. In each plot, colour blocks represent the proportion of membership for 13 progeny, 20 adult trees (A) and 10 regenerants (R) in the inferred Bayesian groups (K=2). Please refer to Table 1 for codes

progeny (Table 3). The levels of mating among relatives (biparental inbreeding) ranged from 0 to 10 % (Table 4). The multilocus correlation of outcrossed paternity ($r_{p(m)}$) varied greatly. Mother trees from TABa exhibited the lowest values of $r_{p(m)}$, which suggested that the progeny from these trees had different fathers. Whereas, two geographic isolated trees (at CAE and PARb) and PEQ had higher values of $r_{p(m)}$, which indicated that the progeny from these isolated trees had few fathers (Table 4). Most mother trees presented very low values for $Ne_{(v)}$ variance and consequently a high number of seed-trees for seed collection (Table 4), with the exception of LAS and MAT, which presented even lower values of $Ne_{(v)}$ and consequently a higher number of seed-trees for seed collection.

Paternity analyses were performed using a total of 269 progeny. Estimates of selfing for each mother tree agreed with the results from mating system analysis as all trees showed a rate of selfing that ranged from 0 to 20 %; the exception was MAT, for which the rate was higher than 47 %. Six adult trees from PARa (PARa1 to PARa6) displayed only two distinct genotypes; therefore, it was not possible to estimate selfing for them. One of the trees that we found at PARa (PARa7) exhibited a unique genotype, so it was possible to estimate the rate of selfing from this tree (selfing=0.19; Table 4). Geographic isolated trees from three sites (CAE, PARb and PIN) were not reproductively isolated, as selfing estimates ranged from 0 to 13 %. The two sites PARa and PEQ exhibited up to 20 % selfing, with 30 % of pollen from inside the site and 50 % of pollen from unknown trees. Trees at MAT and LAS had up to 50 % selfing, with 16 and 50 % unknown pollen donors, respectively. In general, there was a high proportion of pollen from unknown trees (55 %). Most of these unknown trees had similar alleles to the adults that we sampled, but there were 12 alleles that had to have originated from individuals not included in our study. These alleles could be from unknown *D. wilsonii* trees or from the common congener *D. mollis* as a result of rare interspecific hybridisation events that allowed gene introgression from the widespread congener *D. mollis* into *D. wilsonii*.

Table 5 *F*-statistic estimation of F_{ST} (theta) and Hedrick's G_{ST} for *Dimorphandra wilsonii*

	$\Theta_p(F_{ST})$	G_{ST} (Hedrick)
Adults × progeny	0.009^{NS}	0.019
Adults × regenerants	0.050*	0.096
Progeny × regenerants	0.038*	0.070
Progeny: between sites	0.231*	0.729
Progeny: between families	0.174*	0.534
Groups: Northern × Southern	0.141*	0.252

NS non-significant

*Significant ($P<0.05$)

Discussion

Genetic diversity, genetic structure and likely bottleneck events

Rare plants are expected to show low levels of genetic variation; however, there is variation among rare species, with higher variation for species that originally had a wider distribution and have declined more recently, and lower values for species with a recent origin that never occupied a large range (Karron 1991). *Dimorphandra wilsonii* presented low levels of genetic variation, as would be expected if this species is a rare plant with a recent origin that has never occupied a large range. A molecular phylogeny of *Dimorphandra* suggests a recent origin for *D. wilsonii*, which indicates that this species is not an evolutionary relic (Vinson et al., unpublished data). In addition, all known trees of *D. wilsonii* are restricted to a small area near the city of Belo Horizonte (Rizzini 1969; Silva 1986). Our results agree with the microsatellite marker study of Viana e Souza and Lovato (2010) where, overall, low levels of genetic diversity were found for the rare *D. wilsonii*, but higher levels were found for the congener species *D. mollis*. A recent bottleneck event is a factor likely to account for the negative values for the fixation index (*F*) and the associated heterozygote excess compared with HWE expectations for *D. wilsonii*. This is supported by the observation that allelic diversity (especially rare alleles) is lost more rapidly than heterozygosity during bottlenecks. Deforestation, forest fragmentation, indiscriminate exploitation and other activities that

degrade natural environments are actions that lead to the reduction in effective population size and affect fitness-related characteristics (as observed for *Dictamnus albus*, Hensen and Oberprieler 2005). As *D. wilsonii* is a naturally rare species with a geographical distribution around the capital city of a Brazilian state, fragmentation has been occurring over the last 200 years, reducing the number of trees of its small populations. In addition, the dispersion of seeds is very limited; we observed only one natural regenerant of *D. wilsonii* in the field. In its common congener species, *D. mollis*, most seeds germinate under the tree and some are found in tapir faeces. The original dispersers of these species were likely part of the South American megafauna (ground sloths, gomphotheres (mastodon-like proboscidians), glyptodonts), which became extinct around 10,000 years ago (Bizerril et al. 2005).

Geographic isolated, but not reproductively isolated, trees and outcrossed mating system

Endangered species could be a result of environmental catastrophes or genetic deterioration (Aguilar et al. 2008). Inbreeding, through selfing or mating between related individuals, is directly involved in genetic deterioration because it reduces reproduction and survival rates (inbreeding depression). On the other hand, selfing may be favourable in a rare species in cases where there are few breeding partners and inefficient pollinators. Although there have been studies on the genetic diversity of naturally rare plants, there have been few about their reproductive biology (Aguilar et al. 2008; Honnay and Jacquemyn 2007). Our results indicate that the naturally rare species *D. wilsonii* has an outcrossed mating system.

D. wilsonii was shown to be outcrossed and to exhibit low levels of inbreeding; however, two sites presented high levels of selfing (MAT and LAS). The low levels of inbreeding in *D. wilsonii* allow a reproductively isolated tree to produce seeds. Currently, most *D. wilsonii* trees are having success as outcrossers, with an availability of pollen donors and efficient pollinators. Therefore, the resulting progeny exhibit considerable levels of genetic diversity, which prevents the detrimental effects of the reduced levels of heterozygosity and the expression of deleterious, recessive alleles, as well as making genetic rescue unnecessary (e.g., *Medusagyne oppositifolia*, Finger et al. 2011).

Our study did not identify any reproductively isolated trees, and most assessed trees showed high levels of outcrossing and progeny had levels of genetic diversity and heterozygosity which were significantly higher than those of adults. The average number of trees that contributed to the pollination in each population was relatively low. Furthermore, the multilocus correlation of outcrossed paternity ($r_{p(m)}$) suggested that the isolated mother trees that we sampled received pollen from fewer pollen donors (yet to be located) than mothers surrounded by other trees. In addition, paternity correlations revealed correlated mating, which indicated that offspring were not composed exclusively of half-sibs, but a mixture of half-sibs, full-sibs and self-half-sibs. This led to low individual and average variance effective population sizes within families, with even lower values for LAS and MAT (Table 4); the values were very low compared with those expected in panmictic populations (N_e=4). On the basis of this result, we calculated the minimum number of trees from which seed should be collected for ex situ conservation. With the aim of maintaining long-term genetic diversity, our analyses indicated that seed should be collected from at least 141 trees (range: from 138 to 188), for a reference of 150 trees. Given that only 21 trees are known, all trees should be included for seed sampling, and the recommendation of sampling 141 trees should be considered with caution as it cannot be implemented with the current known population size.

Similarly to *D. wilsonii*, the rare *Dyckia ibiramensis* showed a mixed mating system and a low number of pollen donors, due to isolated populations and the fact that the pollinator deposited 50 % of pollen in the first visited flower (Hmeljevski et al. 2011). Fragmentation of the Cerrado has resulted in few trees of *D. wilsonii* per site, and these sites are very far from each other, from 8 to 100 km apart, which may have contributed to the low number of pollen donors.

Implications for conservation

Human intervention to aid the reproduction of *D. wilsonii* seems unnecessary at present; fruit bearing viable seeds were produced under natural conditions. However, to increase the average number of effective pollen donors (N_{ep}=2.61), trees may be reintroduced deliberately at sites with an intermediate location, with the specific aim of facilitating the movement of pollinators and, consequently, increasing the flow of viable pollen over longer distances. It would be worthwhile to attempt to identify additional trees whose existence was predicted by our paternity tests. Most importantly, these trees merit immediate protection given that their distinct genetic composition adds new alleles to the depauperate genetic diversity of *D. wilsonii*. If these vulnerable trees are left unprotected, the risk of genetic erosion is high owing to several factors that threaten the species persistence over time. During the elaboration of management strategies, conservation agencies (such as IBAMA and the Belo Horizonte Zoo-Botanic Foundation) should consider that the species comprises two Bayesian groups (Fig. 2). The genetic compositions of these groups contrast; therefore, both groups deserve long-term protection to conserve in situ the extant genetic diversity of *D. wilsonii*. We recommend that recovery programs should exchange seedlings among sites. Although we did not quantify seedling recruitment, our observations suggest that the rate of establishment of seedlings is very low in the conditions in the field,

which is worrying; the lack of replacements for the long-lived perennial trees will soon lead to species extinction. There is thus an urgent need to determine the causes of the low seedling recruitment in *D. wilsonii* (e.g., ineffective seed dispersion due to megafauna extinction). Meanwhile, we suggest that *D. wilsonii* be considered amongst the candidate species for restoration of degraded habitats within the region; cultivation of *D. wilsonii* ex situ may provide juvenile trees for reintroduction programs.

Acknowledgments The authors thank all owners of the visited sites, and Mr. Joseph Hein (Fazenda Cauaia) and Cimentos Liz, for allowing the collection of material; in addition, IBAMA granted a special licence for sampling *D. wilsonii* (Licence No. 24385-1). We also thank Eveline Caixeta and other members of the Laboratório de Biotecnologia do Cafeeiro for the use of the DNA Analyser, and members of our laboratory for support. CCV received a post-doc fellowship from Fundação de Amparo à Pesquisa do estado de Minas Gerais (BPD-00037-10) and a grant from the International Foundation for Science (D/5064-1). TCSS received a grant from CNPq (507951/2010-7). LOO received a fellowship from Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (PQ 304153/2012-5) and a grant from Fundação de Amparo à Pesquisa do estado de Minas Gerais (PPM-00291-13).

Conflict of interest The authors declare that they have no commercial or financial relationships that could be construed as a potential conflict of interest.

Data archiving statement Data has been submitted to Dryad (https://datadryad.org/) as three files: (1) ID of samples, X and Y coordinates, (2) microsatellite genotypes of all adult trees and (3) microsatellite genotypes of all progeny.

References

Aguilar R, Quesada M, Ashworth L, Herrerias-Diego Y, Lobo J (2008) Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. Mol Ecol 17:5177–5188. doi:10.1111/j.1365-294X.2008.03971.x

Barrett SCH, Kohn JR (1991) Genetic and evolutionary consequences of small population size in plants: implications for conservation. In: Falk DD, Holsinger KE (eds) Genetics and conservation of rare plants. Oxford University Press, New York, pp 3–30

Bizerril MXA, Rodrigues FHG, Hass A (2005) Fruit consumption and seed dispersal of *Dimorphandra mollis* Benth. (Leguminosae) by the lowland tapir in the Cerrado of Central Brazil. Braz J Biol 65: 407–413. doi:10.1590/S1519-69842005000300005

Carneiro FS, Lacerda AEB, Leme MR, Gribel R, Kanashiro M, Wadt LHO, Sebbenn AM (2011) Effects of selective logging on the mating system and pollen dispersal of *Hymenaea courbaril* L. (leguminosae) in the Eastern Amazon as revealed by microsatellite analysis. For Ecol Manage 262:1758–1765

Cole CT (2003) Genetic variation in rare and common plants. Annu Rev Ecol Syst Evol 34:213–237. doi:10.1146/annurev.ecolsys.34.030102.151717

Earl DA, VonHoldt BM (2011) STRUCTURE HARVESTER; A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. doi:10.1007/s12686-011-9548-7

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genet 164:1567–1587

Fernandes FM, Fonseca AG, Kaechele K, Goulart MF, Marinho W, Souza HAV, Queiroz AR, Giorni VT, Oliveira G, Rodrigues MJ, Bacelar M, Lovato MB (2007) Tentando evitar mais uma extinção: o caso do "Faveiro de Wilson" (*Dimorphandra wilsonii* Rizzini). In: Pereira TS, Costa MLMN, Jackson PW (Eds.) Recuperando o verde para as cidades: a experiência dos jardins botânicos brasileiros. Rede Brasileira de Jardins Botânicos, Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rio de Janeiro. pp 87–98

Finger A, Kettle CJ, Kaiser-Bunbury CN, Valentin T, Doudee D, Matatiken D, Ghazoul J (2011) Back from the brink: potential for genetic rescue in a critically endangered tree. Mol Ecol 20:3773–3784. doi:10.1111/j.1365-294X.2011.05228.x

Goodwillie C, Kalisz S, Eckert CG (2005) The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. Annu Rev Ecol Evol Syst 36:47–79

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486

Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59:1633–1638

Hensen I, Oberprieler C (2005) Effects of population size on genetic diversity and seed production in the rare *Dictamnus albus* (Rutaceae) in central Germany. Conserv Genet 6:63–73. doi:10.1007/s10592-004-7745-6

Hmeljevski K, Reis A, Montagna T, Reis MC (2011) Genetic diversity, genetic drift and mixed mating system in small subpopulations of *Dyckia ibiramensis*, a rare endemic bromeliad from Southern Brazil. Conserv Genet 12:761–769. doi:10.1007/s10592-011-0183-3

Honnay O, Jacquemyn H (2007) Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. Conserv Biol 21:823–831. doi:10.1111/j.1523-1739.2006.00646.x

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106. doi:10.1111/j.1365-294X.2007.03089.x

Karron JD (1991) Patterns of genetic variation and breeding systems in rare plant species. In: Falk DD, Holsinger KE (eds) Genetics and conservation of rare plants. Oxford University Press, New York, pp 87–98

Leimu R, Mutikainen P, Koricheva J, Fischer M (2006) How general are positive relationships between plant population size, fitness and genetic variation? J Ecol 94:942–952. doi:10.1111/j.1365-2745.2006.01150.x

Lewis, PO, Zaykin D (2001) GDA – Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewishome/software.html

Marshall TC, Slate J, Kruuk LEB et al (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655. doi:10.1046/j.1365-294x.1998.00374.x

Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858. doi:10.1038/35002501

Oosterhout CV, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Not 4:535–538. doi:10.1111/j.1471-8286.2004.00684.x

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genet 155:945–959

Riahi M, Zarre S, Maassoumi AA, Attar F, Kazempour Osaloo S (2010) An inexpensive and rapid method for extracting papilionoid

genomic DNA. Genet Mol Res 9:1334–1342. doi:10.4238/vol9-3gmr839

Ritland K (1989) Correlated matings in the partial selfer *Mimulus guttatus*. Evolution 43:848–859. doi:10.2307/2409312

Ritland K (2002) Extensions of models for the estimation of mating systems using n independent loci. Heredity 88:221–228. doi:10.1038/sj.hdy.6800029

Ritland K, Jain S (1981) A model for the estimation of outcrossing rate and gene-frequencies using n independent loci. Heredity 47:35–52. doi:10.1038/hdy.1981.57

Rizzini CT (1969) Espécies novas de árvores do Planalto Central Brasileiro. An Acad Bras de Ciênc 41:239–244

Schaberg PG, DeHayes DH, Hawley GJ, Nijensohn SE (2008) Anthropogenic alterations of genetic diversity within tree populations: Implications for forest ecosystem resilience. Fores Ecol and Manag 256:855–862. doi:10.1016/j.foreco.2008.06.038

Silva MF (1986) Flora Neotrópica*: Dimorphandra* (Caesalpiniaceae). The New York Botanical Garden Press, New York 44:1–128

Simberloff D (1988) The contribution of population and community biology to conservation science. Annu Rev Ecol Syst 19:473–511. doi:10.1146/annurev.es.19.110188.002353

Viana e Souza HA, Lovato MB (2010) Genetic diversity and structure of the critically endangered tree *Dimorphandra wilsonii* and of the widespread in the Brazilian Cerrado *Dimorphandra mollis*: implications for conservation. Biochem Syst Ecol 38:49–56. doi:10.1016/j.bse.2009.12.038

Vinson CC, Azevedo VCR, Mendonça MAC, Ciampi AY, Oliveira LO (2013) Microsatellite markers for the rare tree *Dimorphandra wilsonii* (Caesalpinioideae, Fabaceae) and transferability to *Dimorphandra* species. Mol Ecol Resour 13(2):341–343. doi:10.1111/1755-0998.12061