New Phytologist

Landscape genomics reveals altered genome wide diversity within revegetated stands of *Eucalyptus microcarpa* (Grey Box)

Rebecca Jordan[1,2], Shannon K. Dillon[3], Suzanne M. Prober[2]* and Ary A. Hoffmann[1]*

[1]Bio21 Institute, School of BioSciences, University of Melbourne, 30 Flemington Rd, Parkville, Vic 3010, Australia; [2]CSIRO Land and Water, 147 Underwood Ave, Floreat, WA 6014, Australia; [3]CSIRO Agriculture, Clunies Ross Street, Black Mountain, ACT 2601, Australia

Author for correspondence:
Rebecca Jordan
Tel: +61 3 8344 2314
Email: rebecca.citroen.jordan@gmail.com

Received: *13 April 2016*
Accepted: *31 May 2016*

New Phytologist (2016) **212:** 992–1006
doi: 10.1111/nph.14084

Key words: Eucalyptus, fragmentation, genetic diversity, genomics, restoration.

Summary

- In order to contribute to evolutionary resilience and adaptive potential in highly modified landscapes, revegetated areas should ideally reflect levels of genetic diversity within and across natural stands. Landscape genomic analyses enable such diversity patterns to be characterized at genome and chromosomal levels.
- Landscape-wide patterns of genomic diversity were assessed in *Eucalyptus microcarpa*, a dominant tree species widely used in revegetation in Southeastern Australia. Trees from small and large patches within large remnants, small isolated remnants and revegetation sites were assessed across the now highly fragmented distribution of this species using the DArTseq genomic approach.
- Genomic diversity was similar within all three types of remnant patches analysed, although often significantly but only slightly lower in revegetation sites compared with natural remnants. Differences in diversity between stand types varied across chromosomes. Genomic differentiation was higher between small, isolated remnants, and among revegetated sites compared with natural stands.
- We conclude that small remnants and revegetated sites of our *E. microcarpa* samples largely but not completely capture patterns in genomic diversity across the landscape. Genomic approaches provide a powerful tool for assessing restoration efforts across the landscape.

Introduction

In highly modified and fragmented landscapes, the conservation and restoration of vegetation does not necessarily equate to the long-term sustainability of plant populations (Broadhurst, 2013). Small size, isolation and environmental stressors can result in low fitness and low genetic variability in populations, reducing their potential to persist and adapt to change (Willi *et al.*, 2006; Savolainen *et al.*, 2007; Hoffmann & Sgrò, 2011). Replanting of vegetation (hereafter 'revegetation') is a common response toward ameliorating the effects of habitat loss and fragmentation in multi-use landscapes, yet its success is often defined by the presence of vegetation without consideration of future reproductive or adaptive abilities (Monks *et al.*, 2012). In a context of ongoing local and global change, including land-use transformations, soil degradation, exotic invasions and climate change, revegetation and landscape management goals need to move beyond mere survival of individuals, towards ensuring that plant populations have high genetic diversity and the capacity to persist in the long term (Sgrò *et al.*, 2011; Weeks *et al.*, 2011).

Understanding how genetic diversity is currently distributed and maintained across highly modified landscapes is an important first step towards managing such diversity (Broadhurst & Young, 2007; Hoffmann *et al.*, 2015). Despite potential negative impacts of reduced population size, small remnants or fragments can remain important sources of genetic diversity and connectivity in modified landscapes (Lander *et al.*, 2010; Breed *et al.*, 2011; Broadhurst, 2013). Because revegetation attempts to enhance landscape genetic diversity and connectivity (Neale, 2012), understanding how diversity is currently distributed across the landscape provides a baseline against which current revegetation efforts can be assessed.

In order to improve future revegetation attempts, past efforts need to be assessed, especially in comparison to natural sites across the landscape (Hufford & Mazer, 2003; Menges, 2008; Neale, 2012; Havens *et al.*, 2015). Despite emerging interest in alternative seed-sourcing strategies (Broadhurst *et al.*, 2008; Breed *et al.*, 2013; Prober *et al.*, 2015), local provenancing is still commonly used in revegetation projects (Hancock & Hughes, 2012). Although it attempts to preserve local variation, the restrictive nature of local seed sourcing can result in trade-offs between maintaining local genotypes and capturing high genetic diversity and thus adaptive potential (Sgrò *et al.*, 2011). These challenges are particularly acute for long-lived species, where population climatic optima may lag behind climate changes, resulting in local genotypes no longer being most fit (Jump &

*These authors contributed equally to this work.

Penuelas, 2005; Savolainen *et al.*, 2007; Aitken *et al.*, 2008; Kuparinen *et al.*, 2010).

Past studies assessing genetic diversity within fragments and revegetation sites have found mixed results (Kramer *et al.*, 2008; Neale, 2012). For revegetation, a loss of diversity, due to genetic bottlenecks from small founding numbers, has been found in some studies (Dolan *et al.*, 2008; Yokogawa *et al.*, 2013) but not in others (Reynolds *et al.*, 2012; Ritchie & Krauss, 2012). Similarly contrasting results have been found in fragments, which may show a loss of diversity (Honnay & Jacquemyn, 2007; Aguilar *et al.*, 2008) but not in all cases (Kramer *et al.*, 2008).

Past studies of fragmented remnants and revegetation sites have relied on marker systems such as allozymes and microsatellites, mostly with undefined locations within the genome. With only a limited number of markers available, inferences regarding genome-wide diversity are based on only small amounts of information. Genome-wide approaches, by contrast, provide many thousands of markers spanning the entire genome, with opportunities for more powerful analysis and reliable estimates of genome wide diversity as well as the identification of genomic areas where there is increased diversity and where adaptive loci are located (Allendorf *et al.*, 2010; Bragg *et al.*, 2015). Landscape genomic approaches can offer a detailed assessment of diversity across modified landscapes as well as revegetation sites within these areas, but, to date, generally have not been used for comparisons involving revegetation sites (Mijangos *et al.*, 2015).

We applied a landscape genomic approach to characterize genetic diversity of revegetation sites of *Eucalyptus microcarpa* (Grey Box), within the context of genetic diversity in natural remnants of this species across a fragmented landscape. *Eucalyptus microcarpa* is a key species used in revegetation in the highly fragmented wheat–sheep belt of south-eastern Australia. Extensive clearing for agriculture has caused severe habitat loss of this species and the ecological community 'Grey Box Grassy Woodlands' is now listed as an Endangered Ecological Community under the Australian EPBC Act (1999). We used a genome-wide sequencing approach to assess (1) population structure and diversity patterns across the geographic range including both large and small habitat remnants; (2) how well this diversity and structure has been captured in current revegetated stands; and (3) how genetic diversity differs at the chromosomal level and whether genomic patterns of divergence are consistent across site types. The results highlight the utility of a genomic approach for assessing genetic diversity of current conservation efforts within the context of the wider modified landscape.

Materials and Methods

Field sites and sampling

Eucalyptus microcarpa (Maiden) Maiden samples were collected from four site types – a revegetation site, a small remnant and a large remnant, sampled in two ways (see later) – within a 30 km radius of each other at seven locations across Victoria and southern New South Wales (NSW; Fig. 1). Revegetation sites varied in age from 4 to 12 yr and included sites established by planting

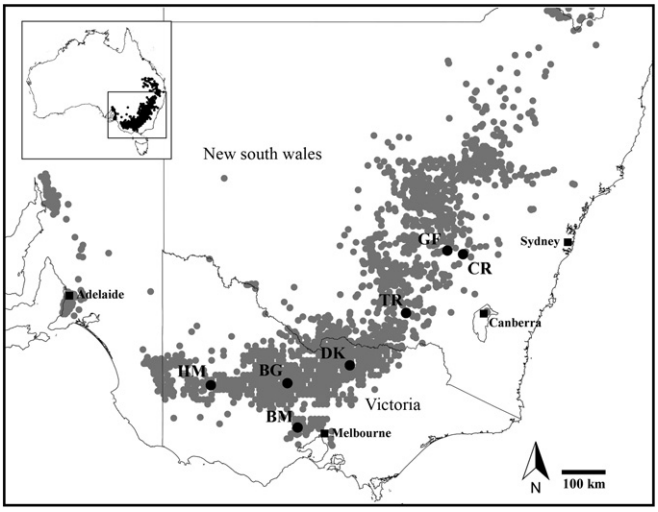

Fig. 1 *Eucalyptus microcarpa* sampling locations across Victoria and New South Wales, Australia. Grey dots indicate recorded occurrences of *E. microcarpa* (data from Atlas of Living Australia; http://www.ala.org.au), providing an indication of the species' distribution. Black dots represent sampled locations. HS, Horsham; BG, Bendigo; BM, Bacchus Marsh; DK, Dookie; TR, The Rock; GF, Grenfell; CR, Crowther. Inset: Map of Australia showing distribution of *E. microcarpa*. Box indicates enlarged region.

seedlings or through direct seeding (Table 1). Unfortunately, there is little documented information regarding provenancing at these sites. Local provenance was known to be used at Dookie (two sites within *c.* 15 km; A. Sands, pers. comm.), whereas at Bendigo multiple seed lots were sourced close to the restoration site (exact distance unknown; J. Crocker, pers. comm.). Material was sourced from local nurseries which were thought to be growing local material at the Bacchus Marsh, The Rock and Grenfell sites (K. Beasley, D. Stein & M. Lewis pers. comm.). There was no provenance information available for Horsham and Crowther.

Small remnants were defined as those sites with *c.* 150 or fewer mature trees, separated from another *E. microcarpa* stand of equal or greater size by at least 500 m. These sites were private property paddock remnants (Bacchus Marsh, Horsham, Bendigo, Dookie and Crowther) or roadside remnants (The Rock and Grenfell). Larger remnants (> 10 ha) had greater numbers of trees and generally connected to wider areas of *E. microcarpa* within the landscape (Table 1). Large remnants comprised State Bushland Reserves (Horsham), State Parks (previously subject to timber harvesting, Bacchus Marsh), roadsides including Travelling Stock Routes and Reserves (Bendigo, Crowther, Grenfell and The Rock) and bushland remnants on private property (Dookie, University of Melbourne Campus). Large trees were sampled at all natural sites to provide trees of comparable ages (Supporting Information Fig. S1).

Leaf material was sampled from 20 individuals within each site type at each location and frozen immediately before later freeze-drying. For revegetation sites and small remnants, individuals were sampled as widely as possible across the site. Given the nature of these sites, sampling distances between individuals was often small (< 30 m). For small remnants, this may have resulted

Table 1 *Eucalyptus microcarpa* sampling site information, including location, site type and final number of samples used for genomic analysis (*n*) per site type

		Large remnant[1]		Small remnant		Revegetation			
Location	State	Size (ha)[2]	*n*	Size (ha)	*n*	Size (ha)	*n*	Yr Planted (approx.)	Method
Bacchus Marsh	Vic	> 35	20	12	19	6	11	2006–2007	P
Horsham	Vic	> 60	20	20[3]	20	4	20	2001	P
Bendigo	Vic	> 10	20	7	20	1	20	2005	DS
Dookie	Vic	> 40	20	10	19	2	13	2009	P
The Rock	NSW	> 25	20	1	20	5	20	2007	P
Crowther	NSW	> 30	20	2	20	1	20	2008	P
Grenfell	NSW	> 40	19 (18)	2	19	2	6	2006–2007	P

Vic, Victoria; NSW, New South Wales; P, planted seedlings; DS, direct seeding.
[1]Samples sizes of both 'normal' and 'dense' sample sets within large remnants, expect for Grenfell where 'dense' sample size is given in brackets.
[2]Approximate sampling area within larger habitat area (i.e. values are minimum site area).
[3]Very low density of trees – only *c.* 70 trees across site.

in the sampling of close relatives. To match sampling styles and allow comparable analysis between small and large remnants, two sampling strategies were used at the larger sites. For 'normal' sampling, the first strategy, trees were sampled on average 70 m from a nearest neighbour (across all locations, SD = 43 m) to avoid close relatives and provide a random population sample for population genetic analysis (hereafter large normal-sampled remnant or large$_{nm}$). For the second strategy, 'dense' sampling, trees were sampled at a higher density (28 ± 24 m across all locations), mimicking the sampling style within small remnants (hereafter large dense-sampled remnant or large$_{ds}$).

Throughout this article, 'location' refers to the seven geographic locations where sampling occurred, 'site type' refers to the four different sites sampled at each location (normal-sampling and dense-sampling at large remnants being two different site types), and 'site' refers to the 28 individual sites (four site types × seven locations).

DNA extraction and DArTseq

DNA was extracted from *c.* 20 mg of freeze-dried leaf material. Dried tissue was ground in a mixer mill with a tungsten carbide bead (three rounds of 28 strokes s^{-1} for 1 min) before extraction using a modified CTAB method or Qiagen DNeasy Plant Mini Kit (following the manufacturer's protocol). For CTAB extractions, ground tissue was incubated in 500 µl of lysis buffer (100 mM Tris pH 8.0, 50 mM EDTA pH 8.0, 1.4 M NaCl, 2% PVP, 2% CTAB, 0.2% β-mercaptoethanol plus 56 µg Proteinase K) for 1.5–2 h at 65°C. Lysates were cleaned with two rounds of equal volumes cholorform : isoamyl alcohol (24 : 1 v/v), spun at 10 000 ***g*** for 10 min, with supernatants transferred to new tubes each time. Cleaned lysates were RNase treated with 30 µg RNase A incubated at 60°C for 30–60 min. Finally, DNA was precipitated with 1.5 × volumes of ice-cold isopropanol, incubating at room temperature for 30 min before spinning for 15 min at maximum speed. DNA pellets were washed twice with 500 µl 70% ethanol, spinning for 5 min at maximum speed for each wash before being air-dried overnight and resuspended in 50 µl TE (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0).

Approximately 400 ng of DNA per sample was sent to Diversity Arrays Technology Pty Ltd, Canberra, for individual *Eucalyptus* DArTseq (Sansaloni *et al.*, 2011). Raw, single end, 77-bp reads were split by individual barcode using FASTX Barcodesplitter (v.0.0.14; http://hannonlab.cshl.edu/fastx_toolkit/), allowing for a single mismatch. Barcodes, at the 5′ end, and low quality 3′ end bases were trimmed with TRIMMOMATIC (v.0.32; HEADCROP:8, TRAILING:3, SLIDINGWINDOW:4:15) with a minimum overall read length of 36 bp and minimum average read quality of 20. Trimmed reads were aligned to the *Eucalyptus grandis* genome (v.1.1; Myburg *et al.*, 2014) using BWA-MEM (Li, 2013) with default settings plus a minimum alignment score of 20. Single nucleotide polymorphisms (SNPs) were called from aligned reads with GATK UNIFIEDGENOTYPER (v.3.3; DePristo *et al.*, 2011) using default settings except for a minimum base quality of 20 and minimum variant confidence of 10.

Analysis

An initial principal component analysis (PCA) was performed using the 'ADEGENET' package (Jombart & Ahmed, 2011) in R (R Core Team, 2015) to identify potential outlier samples in the data, for example, those samples with high genetic divergences that might represent species misidentification. VCFTOOLS v.3 (Danecek *et al.*, 2011) was used to filter SNPs to a minimum genotype depth of 10×, minimum genotype quality Phred-score of 30 (99.9% genotype call accuracy, based on given read data), a maximum mean locus depth of 100× (to avoid 'SNPs' from incorrectly aligned paralogs and over-represented organelle reads), a minor allele frequency of > 0.05 and < 20% missing data across all sites. Twenty-nine samples clearly differentiated from the main distribution of samples in the PCA were identified as outlier samples (Fig. S2). These samples represented misidentified juvenile trees in revegetation sites – eight from Bacchus Marsh, seven from Dookie and 14 from Grenfell – and were removed from further analysis.

Filtering for minor alleles across the whole dataset, as described earlier, could result in the exclusion of rare alleles, present at low frequency in only a few sites and thus below the threshold when

assessed across all sites. As these alleles are likely to be informative in assessing population structure and genetic diversity, rather than filtering across all sites, a 'population-level' filtering approach was used for further analyses. SNPs were filtered using the same quality filters as earlier but then filtered to SNPs with < 20% missing data for each site (e.g. present in > 80% of individuals at every site to ensure accurate site-level diversity estimates) and with a minor allele frequency of > 0.05 in at least one site. 'Population-level' filtering was performed excluding the 29 revegetation outlier samples (identified earlier). This filtering resulted in 13 113 polymorphic loci for analysis, distributed across all 11 major Eucalypt scaffolds as well as several minor scaffolds; more than double the 6490 loci attained when filtering across all sites (Table S1). SNP filtering was performed using VCFtools and custom unix bash script (available at https://github.com/rebecca-cj/revegetation).

In order to assess population structure across all site types, and the affinities between natural and revegetated populations, a PCA using population-level filtered SNPs was performed in adegenet. Four additional outlier individuals were identified, based on their clear differentiation from the main distribution of samples, two from small remnants and two from large remnants, and removed from further analyses. Final per site sample sizes are given in Table 1.

Isolation-by-distance (IBD), defined as $F_{ST}/(1-F_{ST})$ vs $\log_{10}(1 + \text{geographic distance})$ (Rousset, 1997), was investigated to compare differentiation across the sampling range between the site types. Pairwise F_{ST} values between the seven locations were calculated for each of the four site types separately in Arlequin v.3.5 (Excoffier *et al.*, 2005). An Arlequin input file was created from a VCFtools PED file output, using PGDspider v.2.0.7.4 (Lischer & Excoffier, 2012). Geographic distance matrices were calculated in GenAlEx v.6.501 (Peakall & Smouse, 2012) and converted to $\log_{10}(1 + \text{geographic distance})$ matrices in R. To test for IBD, and differences in IBD patterns between the four site types, an ANOVA was performed in R using the 'aov' function. The natural log of $F_{ST}/(1-F_{ST})$ was used to meet ANOVA assumptions of normality and homogeneity of variances. Distance, site type and a type-by-distance interaction were used as independent variables. Where significant effects were found, individual pairwise differences were tested using Least Significant Differences with the R package 'agricolae' v.1.2-1 (De Mendiburu, 2014). As site type was significant (see the Results section), separate Mantel tests of $F_{ST}/(1-F_{ST})$ vs $\log_{10}(1 + \text{geographic distance})$ for each of the four site types were performed using the R package 'ade4' v.1.5-2 (Chessel *et al.*, 2004), with 999 permutations.

For within-site genetic diversity comparisons, observed (H_o) and expected heterozygosity (H_e), the percentage of polymorphic loci (%P) and genetic distance (1 – identity-by-state; 1-IBS) between individuals within a site were calculated. Reductions in population size can result initially in a loss of rare alleles (Young *et al.*, 1996). Therefore, heterozygosity, influenced by changes in mid-frequency alleles, provides an indication of changes in allele frequencies, whilst the number of polymorphic loci, influenced by losses of rare alleles, can provide a more sensitive measure to assess early genetic impacts of reduced population size. Genetic distance between individuals can highlight potential losses in diversity, associated with inbreeding in future generations.

Individual genotypes were output, and allele frequencies for the 28 individual sites calculated using VCFtools. Site level H_o was calculated as the proportion of heterozygous loci within an individual averaged across individuals from a site. Site level allele frequencies were used to calculate H_e. As the ancestral allele is not known for *E. microcarpa*, minor allele frequencies for all loci were used to create a folded Site Frequency Spectra (i.e. frequency spectrum between 0 and 0.5 of minor alleles only, instead of all allele frequencies from 0 and 1). To account for variation in the number of samples per site, which may impact %P calculations, a random sampling approach with different sample sizes was used to calculate the percentage of polymorphic loci. Using the PED file created earlier, the percentage of polymorphic loci per site was calculated from 1000 replicates (with replacement between replicates) of n random samples with a custom unix bash script (available at https://github.com/rebecca-cj/revegetation). Significance was tested by ANOVA using individual site data (as described later) for each sample size (n) tested. Average %P per site type per n was calculated for plotting a diversity curve. Genetic distance between individuals was calculated as 1-IBS in Plink v.1.90 (Purcell *et al.*, 2007), with average genetic distance between individuals within a site subsequently calculated in R.

Differences between the four site types for all diversity measures were tested via ANOVA in R, using the 'aov' function, with both site type and location included as independent fixed variables and only main effects were tested. Where significant effects were found, pairwise differences were tested using Least Significant Differences with the R package 'agricolae'.

In order to further explore this large dataset and highlight the potential that detailed genomic approaches provide, diversity, H_e and %P, were assessed at the chromosome level. Only SNPs on the first 11 'scaffolds', representing the 11 eucalypt chromosomes (Myburg *et al.*, 2014), were used in these calculations. These represented the vast majority of SNPs used in this study, with only 166 of the 13 113 SNPs (1.2%) not located within these scaffolds. Custom Python scripts were used to calculate, per chromosome, H_e (as earlier) and %P from site level allele frequency data (scripts available at https://github.com/rebecca-cj/revegetation). Deviation from the genome-wide average (Average $\text{diversity}_{[\text{chromosome}]}$ – Average $\text{diversity}_{[\text{genome}]}$) was calculated, per chromosome, per site, to assess relative variation in diversity between the chromosomes.

In order to assess whether chromosomes with higher relative diversity may be more susceptible to diversity loss in small remnants and revegetation sites, chromosome-level changes in diversity at small remnants and revegetation sites, compared with large remnants, were calculated (e.g. per chromosome: Average chromosome $\text{diversity}_{[\text{small remnant}]}$ – Average chromosome $\text{diversity}_{[\text{normal-sampled large remnant}]}$). Linear models were used to assess whether chromosome-level change in diversity at small remnant and revegetation sites were related to the relative diversity for that chromosome, with chromosome-level variation from the genome-wide average in normal-sampled large remnants used as the measure of relative diversity. Linear models, with relative

chromosome-level diversity at large remnants and location as independent variables, were performed in R using the 'lm' function. Where location was significant, individual regression lines for each location were plotted.

The potential impact of linkage-disequilibrium was assessed by repeating site and chromosome-level H_e tests on a dataset with linked loci pruned out using the 'INDEP-PAIRWISE' function in PLINK (SNP window of 200, with 5 SNP steps and an r^2 cut-off of 0.2). Results, including significance, were qualitatively similar and therefore all loci were retained for analysis.

Data accessibility

Both 'global' and 'population-level' filtered SNP data in VCF format and associated sample information are available from Dryad: doi: 10.5061/dryad.9gp8g.

Results

Landscape genomic patterns across the fragmented distribution of *E. microcarpa* were investigated at 28 individual sites from seven geographic locations. At each location, four sites were sampled, representing a revegetation site (actively replanted), a small natural remnant (*c.* 150 or fewer trees, > 500 m from an equal or greater size stand) and a large natural remnant (> 150 trees, > 10 ha), the latter sampled using both a normal-sampling scheme (widely sampled trees) and dense-sampling scheme (closely sampled trees for comparison to small remnant sampling; see the Materials and Methods section).

Landscape patterns

A strong geographical pattern for *E. microcarpa* was seen in the PCA (Fig. 2). The first axis shows a strong latitudinal cline from Bacchus Marsh in the south, northward to Crowther and Grenfell in central NSW. The second axis separates Horsham, in western Victoria, from the more easterly, central Victorian sites, and potentially indicates an east–west cline. The low levels of variation explained by these two axes (1.75% in total) and a lack of strong population differentiation apart from a clinal pattern suggest limited IBD, likely due to gene flow across the range of *E. microcarpa*.

Small remnants appeared to be representative of their local geographic area, clustering closely to their corresponding large remnant sites. Revegetation sites, however, did not cluster as strongly, being more similar to broader geographic regions than the finer scale local area. The revegetation site at Horsham (western Victoria) appeared to include material from central Victoria. In NSW, revegetation at The Rock and Crowther appeared to have been sourced from more northern provenances.

Consistent with PCA results, significant IBD was found (Fig. 3). Both distance ($F_{[1,76]} = 11.02$, $P = 0.001$) and site type ($F_{[3,76]} = 28.40$, $P < 0.001$) had a significant effect on genetic differentiation between sites ($F_{ST}/(1-F_{ST})$). There was no significant interaction between type and distance ($F_{[3,76]} = 1.26$, $P = 0.294$) so this term was omitted from the ANOVA for pairwise comparisons. Patterns of IBD across the sampling locations were significantly

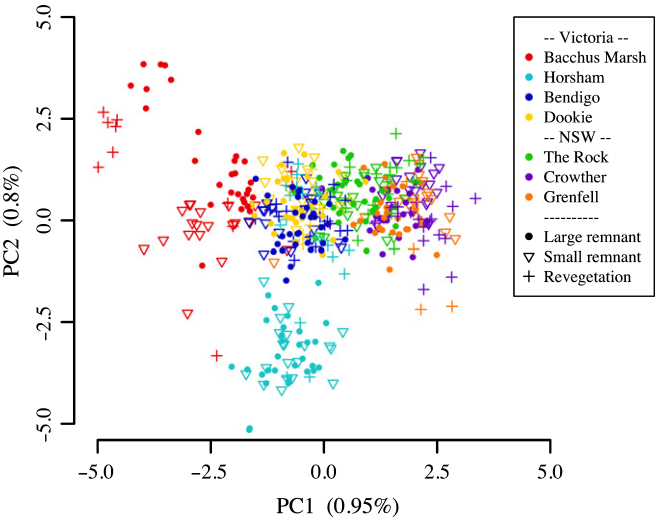

Fig. 2 Principal component analysis (PCA) on 13 113 polymorphic loci for large and small remnants and revegetation sites of *Eucalyptus microcarpa*, showing the first two main axes. NSW, New South Wales.

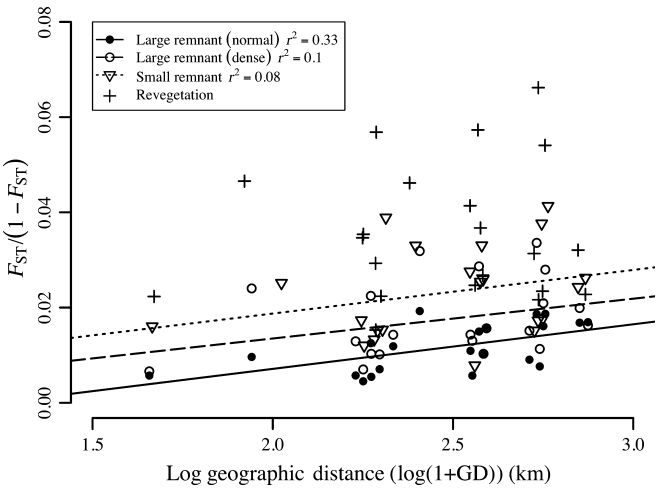

Fig. 3 Isolation-by-distance for each of the four site types across the seven sampled locations of *Eucalyptus microcarpa*. Regression lines included for site types where the Mantel test was significant.

different between all four site types. Individual Mantel tests for each site type found the strongest IBD relationship between large normal-sampled remnants across the sampling distribution ($r^2 = 0.33$, $P = 0.002$; Fig. 3). Weaker IBD was found in large dense-sampled remnants ($r^2 = 0.10$, $P = 0.047$) and small remnants ($r^2 = 0.08$, $P = 0.037$). Both dense-sampling of large remnants and small remnants showed greater differentiation for any given geographic distance, compared with large normal-sampled remnants. Genetic differentiation was highest between revegetation sites and not related to geographic distance.

Within-site diversity

Genetic diversity within individual sites (H_e and %P) varied significantly between site types and locations (Table 2). Mean H_e ranged from 0.074 to 0.083 (Fig. 4a) with revegetation sites

Table 2 ANOVA results for tests of differences in within-site genetic diversity between site types and locations of *Eucalyptus microcarpa*

Diversity measure	Effect	*F*	df	*P*
Expected heterozygosity (H_e; Fig. 4a)	Site Type	9.023	3,18	< 0.001
	Location	17.190	6,18	< 0.001
Observed heterozygosity (H_o; Fig. 4b)	Site Type	0.337	3,18	0.799
	Location	5.021	6,18	0.003
Genetic distance (1-IBS; Fig. 6) [1]	Site Type	6.972	3,17	0.003
	Location	20.038	6,17	< 0.001

[1]Excluding Grenfell revegetation site.

significantly lower than the other three site types (revegetation = 0.0765 vs $large_{nm}$ remnants = 0.0784, $large_{ds}$ remnants = 0.0783 and small remnants = 0.0785). Despite lower H_e on average, revegetation sites at The Rock and Crowther had H_e levels similar to those of large remnants (Fig. 4a). There was no significant difference in H_e between small and large remnants, or between large remnants sampled in different ways. For comparison of diversity levels to other studies that use 'global filtering', average site-level H_e for large and small remnants at all locations was also calculated using the 6490 globally filtered SNPs and found to range from 0.262 to 0.271.

The %P estimate was significantly lower at revegetation sites than large remnants at all sample sizes except $n = 3$ (Fig. 5; Table S2). Small remnants had significantly lower %P than large normal-sampled remnants, but not large dense-sampled remnants, for sample sizes of $n = 9$–19. There was no significant difference between either sampling strategy in large remnants at any sample size, nor between revegetation and small remnants at $n = 16$–19. Lower %P, especially in revegetation sites, was reflected in a reduced incidence of low frequency alleles in contrast to alleles present at an intermediate frequency that were similar between all four site types (Fig. S3).

The mean H_o of individuals showed similar patterns to H_e across sites and locations, ranging from 0.077 to 0.082 (Fig. 4b). Unlike H_e, there was no significant difference between the four site types but there was a significant difference between locations (Table 2), with H_o for Horsham (location mean = 0.0812) being higher than for all other locations (location means, excluding Horsham, ranging from 0.0781 to 0.0794).

The average genetic distance (1-IBS) between individuals within a site ranged from 0.064 to 0.073 (Fig. 6). No significant difference was found between the four site types. However, when the revegetation site at Grenfell (with a small sample size and high estimate variance) was removed from the analysis, average genetic distance between individuals was significantly different between both site types and location (Table 2), with revegetation sites significantly lower than the other three site types (revegetation = 0.0663 vs $large_{nm}$ remnants = 0.0683, $large_{ds}$ remnants = 0.0681 and small remnants = 0.0676).

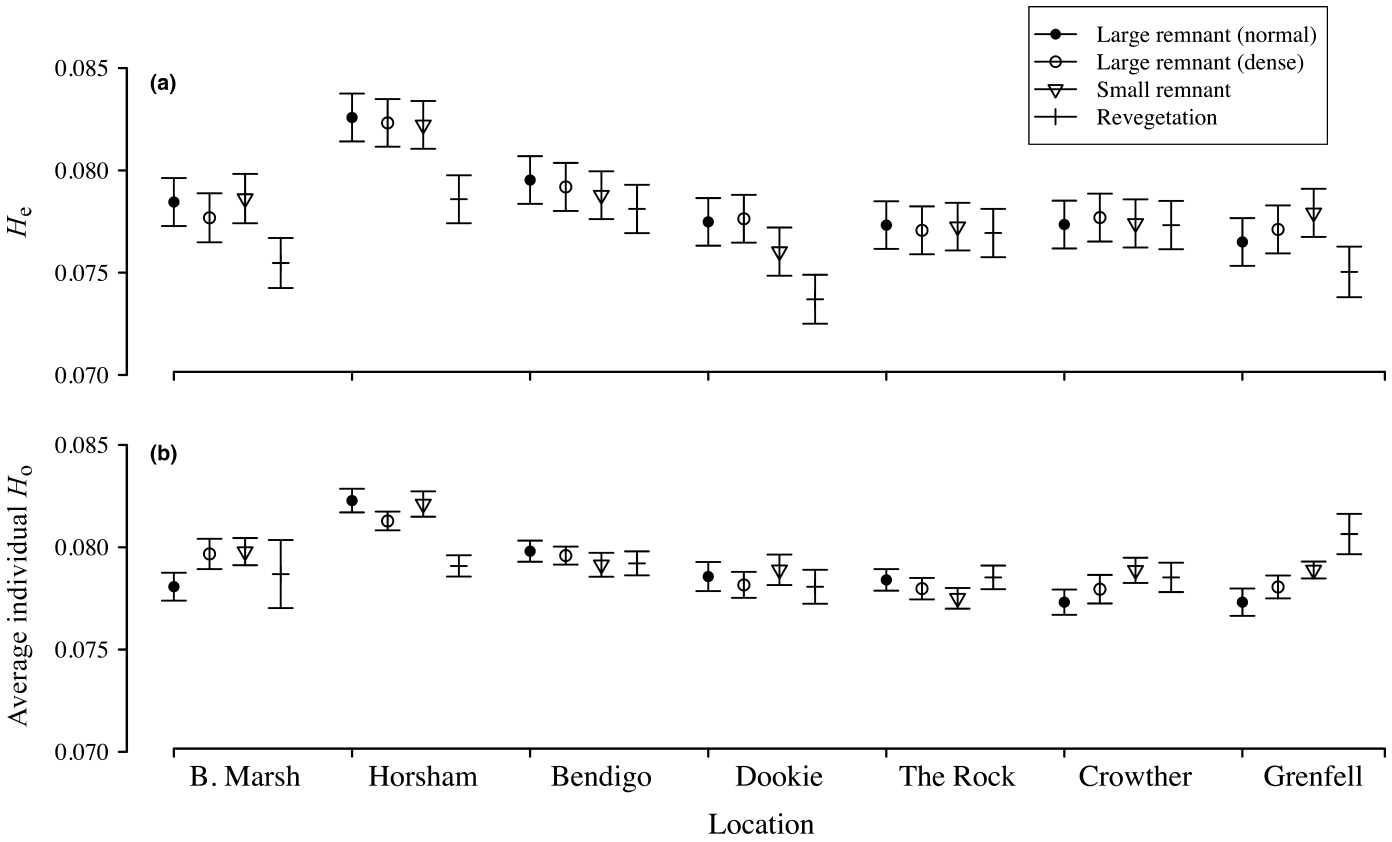

Fig. 4 (a) Expected (H_e) and (b) observed heterozygosity (H_o) for each of the four site types at the seven sampling locations of *Eucalyptus microcarpa*. Error bars ± 1 SE.

Within-site genetic diversity was not distributed evenly across chromosomes, with both within-site H_e and %P varying between chromosomes (Figs 7a, 8a, S4, S5). Across all 28 individual sites, common trends towards higher or lower diversity on particular chromosomes, compared with the genome-wide average, were found: for example, higher relative H_e on chromosomes 6 and 7, in contrast to substantially lower or average H_e on chromosomes 8 and 9, respectively (Fig. 7a).

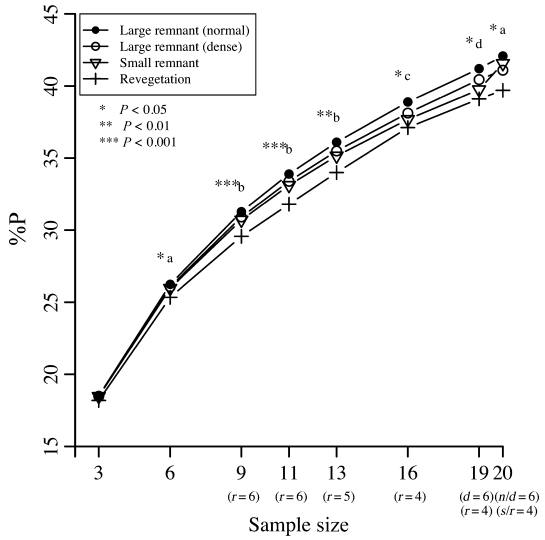

Fig. 5 Average percentage polymorphic loci per site type vs sample size (*n*) in *Eucalyptus microcarpa*. Percentage polymorphic loci (%P) was calculated from 1000 random resamples for each sample size at each of the 28 individual sites. The average %P by site type, across the seven locations, is plotted. Significance calculated by ANOVA using individual site data (see the Materials and Methods section); (a) revegetation lower than all other site types; (b) revegetation lower than all other sites, and small remnants lower than large normal-sampled remnants ($large_{nm}$) but not large densely-sampled remnants ($large_{ds}$); (c) revegetation and small remnants lower than $large_{nm}$ but not $large_{ds}$; (d) revegetation and small remnant lower than $large_{nm}$ though only revegetation lower than $large_{ds}$; numbers in brackets on the *x*-axis indicate the number of sites used to calculate site type average, where this differs from seven; *n*, $large_{nm}$; *d*, $large_{ds}$; *s*, small remnant; *r*, revegetation.

Whilst general trends of higher and lower relative diversity were similar across individual sites, the actual level of diversity differed between large normal-sampled remnants and the other site types among chromosomes (Figs 7b, 8b). Despite little difference in the average genetic diversity of either small sampling area site types – large dense-sampled remnants and small remnants – compared with large normal-sampled remnants, differences in H_e and %P varied somewhat between chromosomes, with some chromosomes suggesting higher diversity compared with large normal-sampled remnants whilst others suggest lower diversity or no difference. The decreased average diversity in revegetation sites compared with large normal-sampled remnants was reflected in more consistent trends towards decreased diversity at the individual chromosome level. However, some chromosomes showed a greater decrease than others (Figs 7b, 8b), and for H_e, some chromosomes even had higher diversity compared with large normal-sampled remnants, despite lower overall genome-wide H_e.

Finally, decreases in chromosome-level diversity in small remnants and revegetation sites, compared with large normal-sampled remnants were greater on chromosomes with higher relative diversity (Fig. 9). Significant negative associations were found between relative chromosome-level diversity, based on large normal-sampled remnants, and changes in both H_e and %P at small remnants and revegetation sites compared with the large remnants (Table 3). With the exception of variation in H_e in small remnants, location also had a significant effect on changes in diversity, with some locations showing greater losses of diversity than others. Lower values of %P for three of the revegetation sites (Fig. 9d) is likely due to low samples size, which whilst impacting the absolute values, should not impact the trends seen. These patterns highlight that regions of higher genomic variability contributed most to differences among site types.

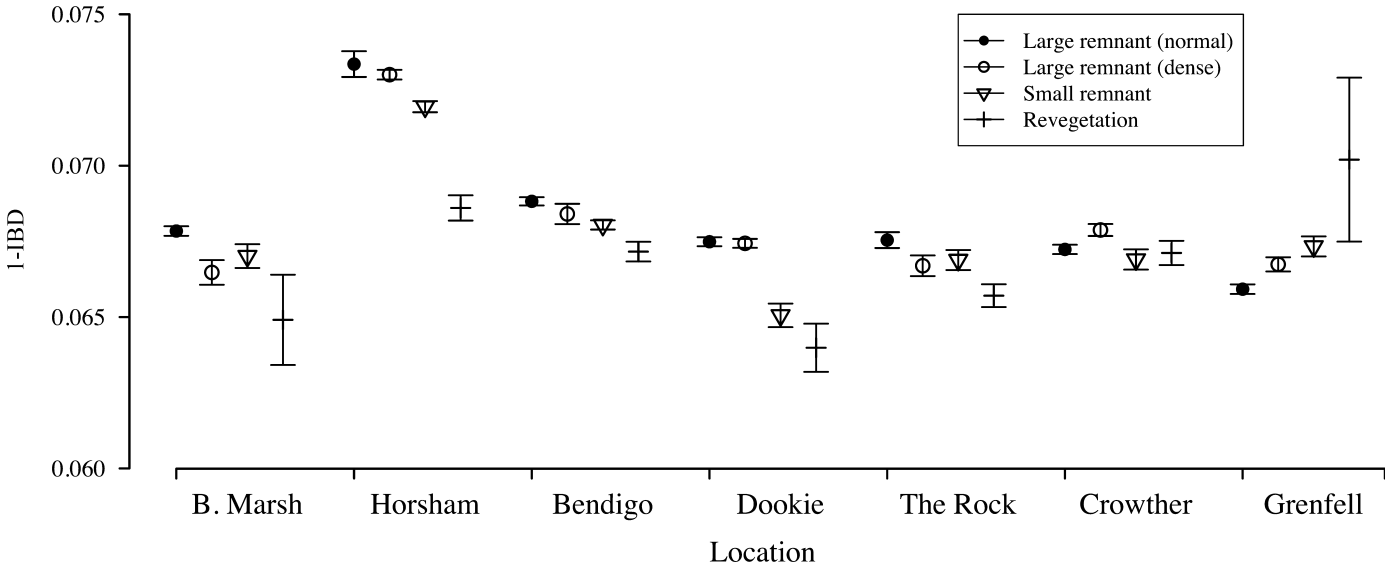

Fig. 6 Average genetic distance (1 – identity-by-state; 1-IBS) between individuals within a site, for each of the four site types at the seven sampling locations of *Eucalyptus microcarpa*. Error bars ± 1 SE.

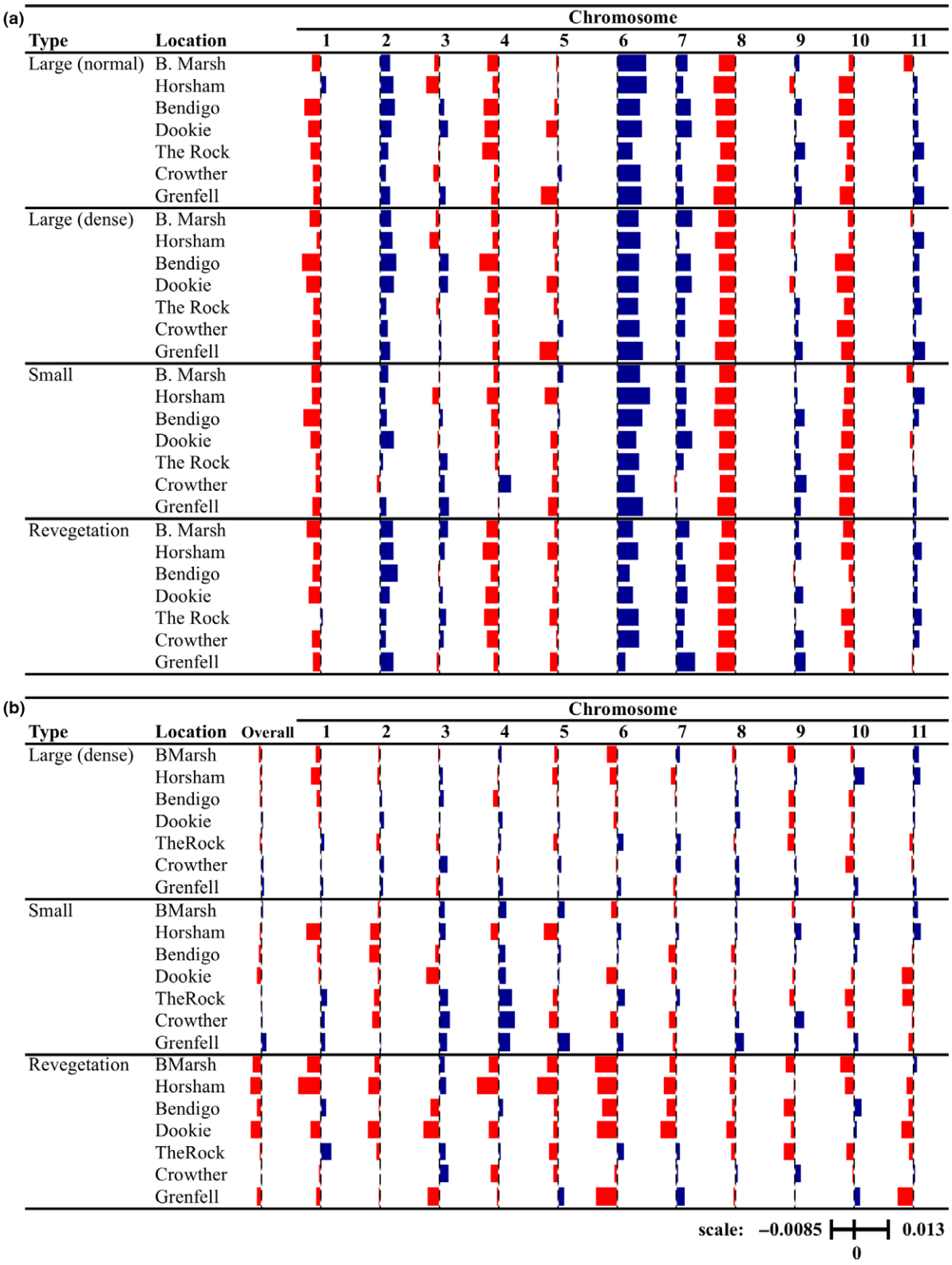

Fig. 7 Per chromosome variation in expected heterozygosity (H_e), from the site genome-wide average (a) and from the chromosome-level large normal-sampled remnant average (b) in *Eucalyptus microcarpa*. For comparison, 'Overall' in (b) indicates the difference between the site and the large normal-sampled remnant genome-wide averages. Blue bars, increased diversity (positive values); red bars, decreased diversity (negative values).

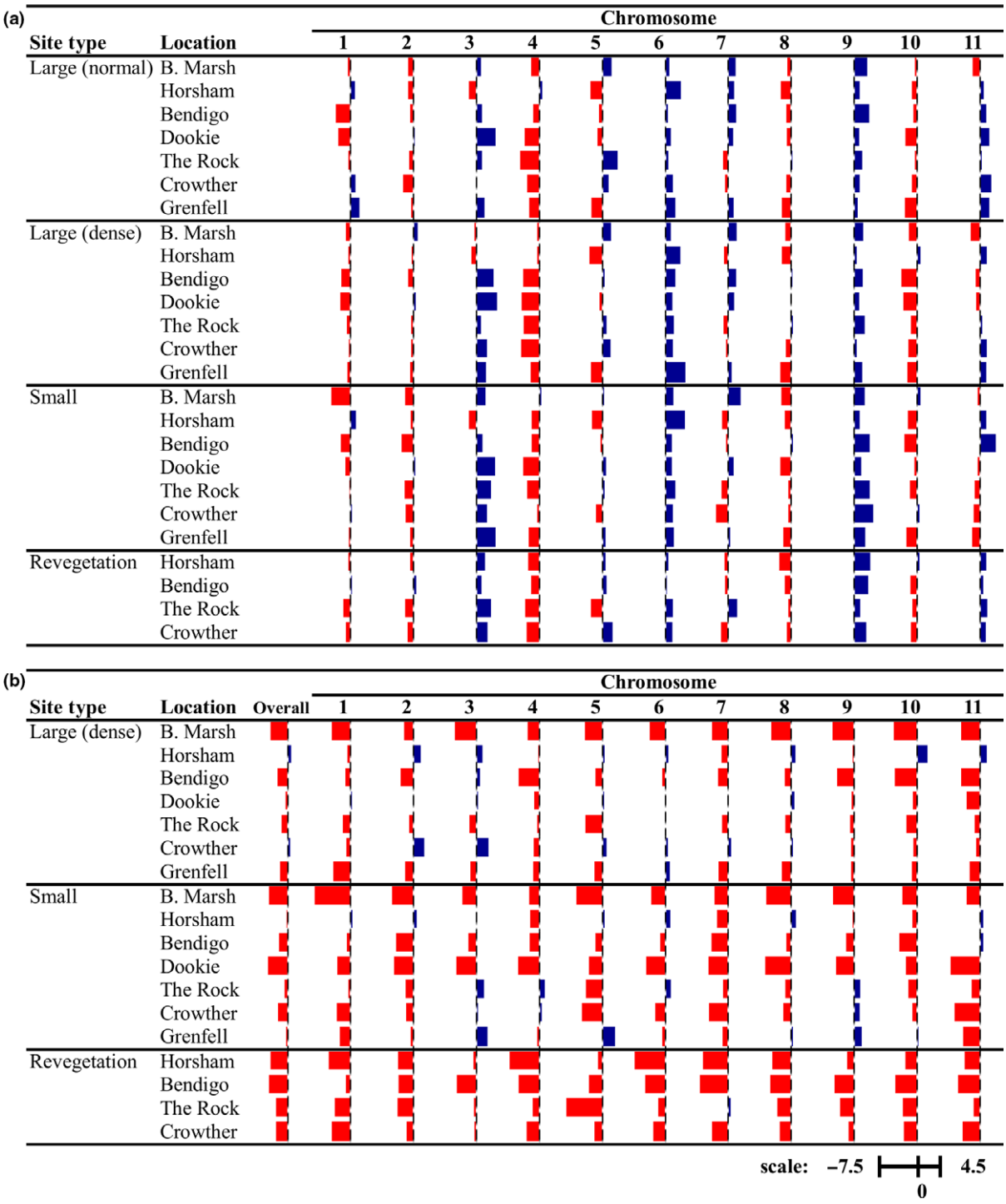

Fig. 8 Per chromosome variation in percentage polymorphic loci, from the site genome-wide value (a) and from the chromosome-level large normal-sampled remnant value (b) in *Eucalyptus microcarpa*. For comparison, 'Overall' in (b) indicates the difference between the site and the large normal-sampled remnant genome-wide values. Blue bars, increased diversity (positive values); red bars, decreased diversity (negative values). Three revegetation sites with low sample sizes, and thus potentially biased estimates, were excluded from the graph (see Supporting Information Fig. S5 for data including these sites).

Discussion

Our genomic assessment of *Eucalyptus microcarpa* showed similar patterns of landscape-wide genetic diversity in both large and small habitat remnants across a fragmented landscape. Small remnants retained high levels of genetic diversity comparable to larger sites, although the strength of landscape-level patterns was reduced. At revegetation sites, genetic diversity was often somewhat lower, and landscape-level patterns were not fully maintained. Differences in diversity at revegetation sites compared

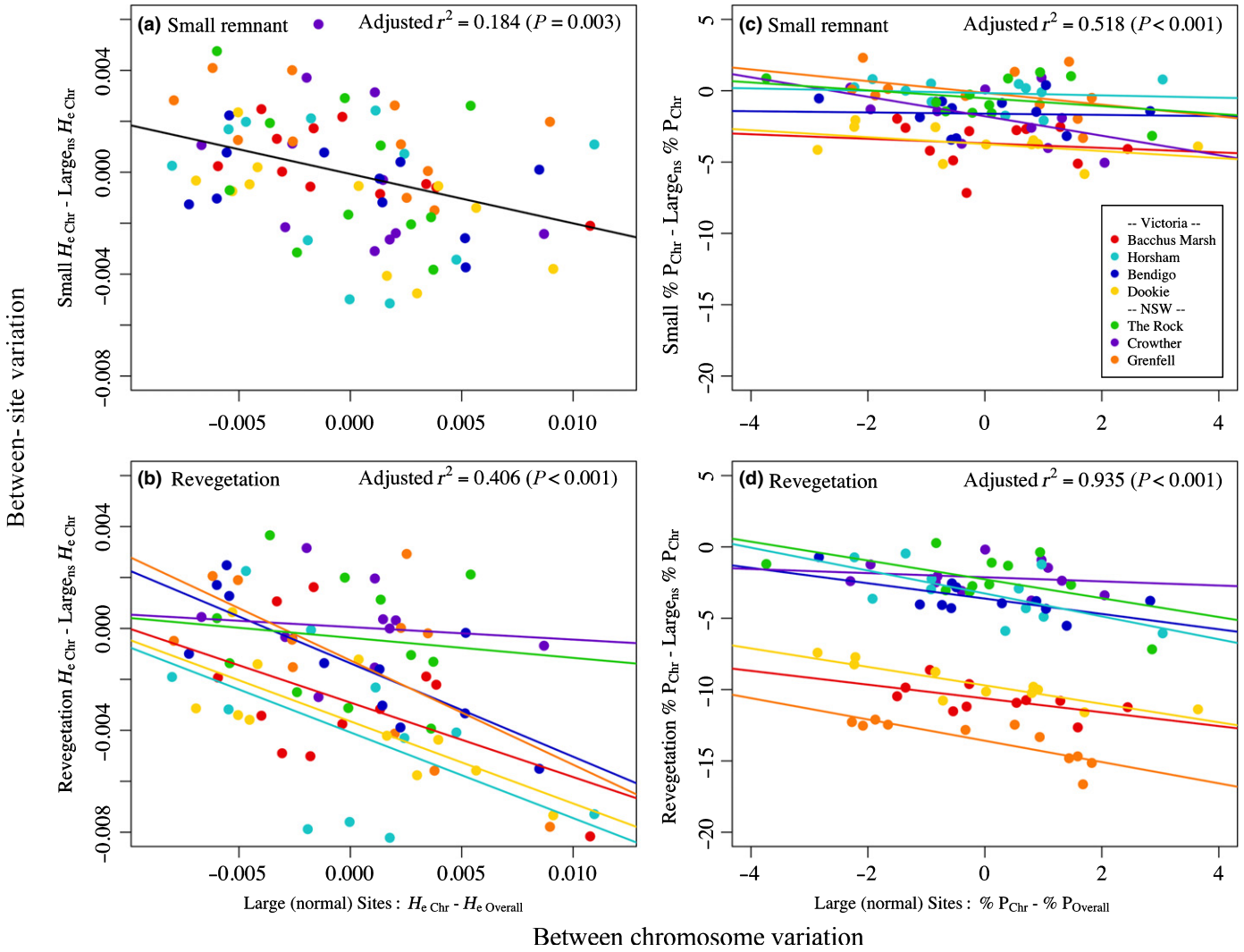

Fig. 9 Relationship between chromosome-level variation (expressed as deviation from genome-wide average in large normal-sampled remnants) and chromosome level change in variation for (a, c) small remnants or (b, d) revegetation sites compared with large normal-sampled remnants of *Eucalyptus microcarpa*. (a, b) Expected heterozygosity (H_e) and (c, d) percentage polymorphic loci (%P). Regression lines are plotted for each location where the location effect was significant; otherwise, an overall regression line was plotted. Chr, chromosome level estimate. Overall, genome-wide estimate.

Table 3 Results of linear model testing the effect of relative chromosome level genetic diversity on diversity changes in small remnants and revegetation compared with natural sites of *Eucalyptus microcarpa*

		H_e			% polymorphic loci		
	Effect	F	df	P	F	df	P
Small remnant	(Relative) Genetic diversity	13.24	1,69	< 0.001	6.52	1,69	0.013
	Location	1.83	6,69	0.105	13.57	6,69	< 0.001
Revegetation	(Relative) Genetic diversity	26.76	1,69	< 0.001	45.46	1,69	< 0.001
	Location	5.45	6,69	< 0.001	175.37	6,69	< 0.001

with natural sites were nevertheless small and varied between localities, suggesting that revegetation has the potential to contribute to landscape-wide diversity. Our results also demonstrate the power of genomics to assess and detect small changes in genetic diversity, as well as the ability to investigate genomic differences at the chromosome level.

Landscape genomic patterns in natural remnants

Evidence of isolation-by-distance (IBD), as found in *E. microcarpa*, has been found in other widespread tree species (e.g. Geraldes *et al.*, 2014; Holliday *et al.*, 2015), including other eucalypts (Prober & Brown, 1994; Bradbury *et al.*, 2013) where

long-generation times, large effective population sizes and long distance gene flow (Alberto *et al.*, 2013) may minimize differentiation across species ranges. In *E. microcarpa*, the position of Horsham and Bacchus Marsh samples at the edge of the main PCA cluster may reflect their geographic location at the western and southern edges of its Victorian distribution. In particular, Bacchus Marsh occurs in the only section of *E. microcarpa*'s distribution that is south of the Great Dividing Range, a biogeographical barrier along the Australian east coast. Alternatively, the distinction at these two sites may be the result of hybridization or introgression, known to occur in many eucalypt species (Grattapaglia *et al.*, 2012) potentially including *E. microcarpa* (Bean, 2009).

Within-site levels of genetic diversity were also similar to other tree species. Using genome-wide SNP data, mean expected heterozygosity (H_e) in eucalypts and other trees has been found to range from 0.18 to 0.25 (Prunier *et al.*, 2011; Mosca *et al.*, 2012; De Kort *et al.*, 2014; Dillon *et al.*, 2014). When calculated using globally filtered single nucleotide polymorphisms (SNPs), comparable to these other studies, average site-level H_e in large and small remnants was similar to the higher end of this range.

Similarities to nearby larger sites, as well as similar levels of heterozygosity and percentage polymorphic loci (%P), suggest that small habitat areas, both within large remnants and separate small, fragmented remnants, can represent wider landscape-scale genetic diversity. Expected decreases in genetic diversity due to reductions in population size (Leimu *et al.*, 2010) are not always seen in fragmented tree populations, including eucalypts (Lowe *et al.*, 2005; Kramer *et al.*, 2008; Mimura *et al.*, 2009; Lander *et al.*, 2010; Ottewell *et al.*, 2010; Breed *et al.*, 2015; Dillon *et al.*, 2015). For long-lived species with long generation times, relatively recent habitat fragmentation may mean that insufficient generations have passed for altered mating dynamics to have resulted in decreased genetic diversity (Aguilar *et al.*, 2008; Kramer *et al.*, 2008; Vranckx *et al.*, 2012; James & Jordan, 2013). The small remnants sampled here likely represent the genetic variation of larger historic populations (Kramer *et al.*, 2008), given that genetic diversity of larger remnants was reflected within small areas and that small remnants may contain pre-European settlement trees (Spooner *et al.*, 2010) due to the relatively recent clearing in the early to mid-1800s (Pryor, 1976). Small habitat areas therefore remain important stores of genetic diversity in these modified landscapes.

High current levels of standing variation, however, do not necessarily ensure the health of future generations. Effects of fragmentation are more commonly seen in seed or post-fragmentation progeny (Vranckx *et al.*, 2012), with progeny showing decreased diversity (Prober & Brown, 1994; Broadhurst, 2011) and increased inbreeding (Aguilar *et al.*, 2008; Lander *et al.*, 2010). In addition to these direct genetic effects, small population sizes also can eventually impact adaptive potential (Leimu & Fischer, 2008; Pickup *et al.*, 2012).

Indeed, some changes in genetic diversity were found in small habitat areas, with increased differentiation, reduced IBD and lower %P at some sample sizes. Greater differentiation between smaller remnants likely results from a loss of variation and changes in allele frequencies due to both loss of individuals and altered mating dynamics (where next-generation trees are present; Young *et al.*, 1996; Leimu *et al.*, 2010), with similar losses of landscape-level genetic structure in small remnants seen in *E. albens* (Prober & Brown, 1994). Differences between small area within large remnants (dense sampling) and small, fragmented remnants suggested that these changes were due to reduced population size in small remnants, not simply altered sampling strategy.

Revegetation site diversity within the wider landscape context

The similarity to nearby large remnants but greater differentiation in revegetated sites of *E. microcarpa* suggests that 'local' provenancing strategies were most likely used and were partly successful in maintaining landscape diversity patterns in revegetation sites. The incomplete capture of genetic variation from natural stands, for example due to poor sampling, may have contributed to differences between natural and revegetation sites. As a consequence of sampling and sourcing strategies, revegetation sites can become genetically distinct from known source populations (Campanella *et al.*, 2012; Yokogawa *et al.*, 2013), even having higher genetic identity with natural nonsource sites than their original source populations (Liu *et al.*, 2008).

Reduced diversity within revegetation sites has been found previously (Dolan *et al.*, 2008; Aavik *et al.*, 2012; Yokogawa *et al.*, 2013; Mijangos *et al.*, 2015). In restored sites of *E. melliodora*, microsatellite analysis indicated that H_e and allelic richness were 4–35% and 3–32% lower, respectively, than natural trees in the surrounding landscape (Broadhurst, 2013). Decreases in H_e (11–21%), allelic richness (8–16%) and number of alleles (13–37%) also have been found in restored sites of perennial herb species compared with source populations (Dolan *et al.*, 2008; allozymes; Yokogawa *et al.*, 2013; microsatellites). In comparison to these findings, the somewhat lower genomic diversity in *E. microcarpa* revegetation sites appears to be minor, with decreases in H_e from 0% to 4.8%, and changes in %P (akin to allelic richness and number of alleles), adjusted to equal sample sizes ($n = 6$), ranging from an increase of 3% at one site to decreases of 1–7% at the remaining sites. Whilst diversity is only slightly lower, the lower genetic distance between individuals in revegetation may increase the susceptibility to future inbreeding depression and loss of diversity (Ellstrand & Elam, 1993; Young *et al.*, 1996).

The relatively small difference between revegetation and natural sites highlights the capacity for landscape diversity to be maintained during revegetation, consistent with findings from other studies (Liu *et al.*, 2008; Lloyd *et al.*, 2012; Reynolds *et al.*, 2012; Ritchie & Krauss, 2012). Careful selection of widely sourced seed, including mixing multiple source populations, can even lead to greater diversity at restored sites than source populations (Smulders *et al.*, 2000; Fant *et al.*, 2013).

Genome-wide diversity patterns in fragmented and revegetated sites

Genetic diversity has long been held as a measure of population health (Ellstrand & Elam, 1993; Lowe *et al.*, 2005; Honnay &

Jacquemyn, 2007; Weeks *et al.*, 2011) with increased average diversity associated with increased fitness and adaptability (Reed & Frankham, 2003). Genomics, however, has the potential to characterize not just overall diversity, but patterns of increased or decreased diversity and differentiation at different scales across the genome (e.g. Hudson *et al.*, 2015) as well as regions under selection (e.g. Zhou *et al.*, 2014).

In the current study, comparisons of H_e between small habitat areas and large remnants highlight how the distribution of genetic diversity across chromosomes can differ between individual sites despite similar overall diversity. Such patterns could help explain finer-scale differences found between the four site types. Similarly, decreases in overall genome-wide diversity in revegetation sites compared with large remnants were not distributed evenly across all chromosomes. The negative correlation between relative chromosome-level diversity and diversity change in small remnants and revegetation sites indicates that variable regions of the genome are more prone to changes in diversity, presumably due to loss of rare alleles. The loss of such diversity could affect adaptive potential based on standing variation, given that low frequency alleles can be important in this process (Barrett & Schluter, 2008; Alberto *et al.*, 2013).

Variation in genetic diversity across the genome, compared with overall averages, has been noted in other organisms (Hohenlohe *et al.*, 2010; Zhou *et al.*, 2014; Hudson *et al.*, 2015). Moreover, the consistency in patterns of relative diversity across chromosomes across sites has been noted in populations of other species including black cottonwood (*Populus trichocarpa*) (Zhou *et al.*, 2014) and three-spined sticklebacks (*Gasterosteus aculeatus*) (Hohenlohe *et al.*, 2010). This pattern may even extend across species; heterozygosity varies across the *E. grandis* genome (Myburg *et al.*, 2014), with peaks in a number of regions including areas on chromosomes 6 and 7, which were found to have higher relative heterozygosity in *E. microcarpa*.

An understanding of how genetic diversity and regions of adaptive differentiation are distributed across the genome eventually may lead to targeted screening of genetic variation important for conservation and restoration. Genomics studies are beginning to identify islands of divergence, which represent genomic regions of particular importance in adaptation (Zhou *et al.*, 2014; Holliday *et al.*, 2015), as well as genomic regions that are highly differentiated across populations and species (Steane *et al.*, 2011; Hudson *et al.*, 2015). In these regions there are likely to be different patterns of population structure between adaptive and neutral loci (Moura *et al.*, 2014; Steane *et al.*, 2014). Identifying such patterns will be useful when selecting restoration material that increases adaptive potential for future climate change and other stressors (Hoffmann *et al.*, 2015).

Maintaining genetic diversity in highly modified landscapes

In order to enhance evolutionary potential, there is a need to maintain high genetic diversity (Jump *et al.*, 2009; Sgrò *et al.*, 2011). Such diversity currently appears to be present in both large and small remnants of *E. microcarpa* across the landscape. As effects of habitat loss may not be seen until future generations, there is a need to maintain landscape-level diversity and connectivity. Creating genetically diverse revegetated stands will therefore be important not only for ensuring the longevity of revegetation itself, but also for supporting genetic diversity at the wider landscape scale.

Although current revegetation in *E. microcarpa* does not entirely capture the diversity seen in the wider landscape, differences are small. As thinking around seed sourcing is shifting away from local provenancing towards capturing high genetic diversity and 'adaptive potential' (Broadhurst *et al.*, 2008; Breed *et al.*, 2013; Maschinski *et al.*, 2013; Havens *et al.*, 2015; Prober *et al.*, 2015), a loss of landscape-wide diversity patterns in revegetation may be of less importance than a decrease in genetic diversity. The use of multiple populations and consideration of landscape-wide species genetic variation can help to improve restoration success (Godefroid *et al.*, 2011; Mijangos *et al.*, 2015).

In summary, this study demonstrates the utility of a genomics approach to uncover subtle but significant patterns of genetic diversity across habitat components of a highly modified landscape. Genomics affords new opportunities for assessing the genetic health of populations, providing not only traditional genome-wide population estimates, but also moving beyond these averages to in-depth assessment of the distribution of genetic diversity across the genome (Hoffmann *et al.*, 2015). In addition, the ability of genomics to detect putatively adaptive regions will help in assessing genetic adaptive potential more directly.

Acknowledgements

Thanks to Greening Australia, especially G. Fifield and J. Gardner, for their assistance in locating revegetation sites, as well as L. Evans and A. Sands (Goulburn Broken Indigenous Seedbank), J. Crocker (Mid-Loddon Landcare Network), M. Lewis (Lachlan CMA) and D. Stein (Murrumbidgee CMA). Thanks to K. Ussher, W. Gebert, S. Panter, S. Culver, K. Bell, K. Roberts, L. Bedel and C. O'Meara for their assistance with fieldwork and to all the landholders for permission to access their property. Thanks to P. Griffin, D. Bush, H. L. Yeap and three anonymous reviewers for their constructive feedback that helped improve this manuscript. All material was collected under permits from Department of Environment and Primary Industry, Victoria (10006855), Parks and Wildlife Service, NSW (SL101225), Forestry Corporation of NSW (SPPR0034) and with support and permission from LHPA to sample in NSW Travelling Stock Reserves. This study was supported by a CSIRO Julius Award (S.M.P.), Science and Industry Endowment Fund (A.A.H.), Australian Postgraduate Award (R.J.), and Holsworth Wildlife Research Endowment (R.J.).

Author contributions

All authors contributed to the development of the research question and design; R.J. performed the fieldwork and analysis and prepared the manuscript; S.K.D., S.M.P. and A.A.H. assisted with analysis and editing of the manuscript.

References

Aavik T, Edwards PJ, Holderegger R, Graf R, Billeter R. 2012. Genetic consequences of using seed mixtures in restoration: a case study of a wetland plant *Lychnis flos-cuculi*. *Biological Conservation* **145**: 195–204.

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. *Molecular Ecology* **17**: 5177–5188.

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**: 95–111.

Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, Lefèvre F, Lenormand T, Yeaman S, Whetten R *et al.* 2013. Potential for evolutionary responses to climate change – evidence from tree populations. *Global Change Biology* **19**: 1645–1661.

Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**: 697–709.

Barrett RDH, Schluter D. 2008. Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**: 38–44.

Bean AR. 2009. Taxonomic and nomenclatural notes on the Eastern grey boxes (Eucalyptus ser. Moluccanae Chippendale, Myrtaceae) and the reinstatement of *Eucalyptus woollsiana* R.T.Baker. *Austrobaileya* **8**: 25–34.

Bradbury D, Smithson A, Krauss SL. 2013. Signatures of diversifying selection at EST-SSR loci and association with climate in natural *Eucalyptus* populations. *Molecular Ecology* **22**: 5112–5129.

Bragg JG, Supple MA, Andrew RL, Borevitz JO. 2015. Genomic variation across landscapes: insights and applications. *New Phytologist* **207**: 953–967.

Breed MF, Ottewell KM, Gardner MG, Lowe AJ. 2011. Clarifying climate change adaptation responses for scattered trees in modified landscapes. *Journal of Applied Ecology* **48**: 637–641.

Breed MF, Ottewell KM, Gardner MG, Marklund MHK, Stead MG, Harris JBC, Lowe AJ. 2015. Mating system and early viability resistance to habitat fragmentation in a bird-pollinated eucalypt. *Heredity* **115**: 100–107.

Breed MF, Stead MG, Ottewell KM, Gardner MG, Lowe AJ. 2013. Which provenance and where? Seed sourcing strategies for revegetation in a changing environment. *Conservation Genetics* **14**: 1–10.

Broadhurst L, Young A. 2007. Seeing the wood and the trees – predicting the future for fragmented plant populations in Australian landscapes. *Australian Journal of Botany* **55**: 250–260.

Broadhurst LM. 2011. Genetic diversity and population genetic structure in fragmented *Allocasuarina verticillata* (Allocasuarinaceae) – implications for restoration. *Australian Journal of Botany* **59**: 770–780.

Broadhurst LM. 2013. A genetic analysis of scattered Yellow Box trees (*Eucalyptus melliodora* A. Cunn. ex Schauer, Myrtaceae) and their restored cohorts. *Biological Conservation* **161**: 48–57.

Broadhurst LM, Lowe A, Coates DJ, Cunningham SA, McDonald M, Vesk PA, Yates C. 2008. Seed supply for broadscale restoration: maximizing evolutionary potential. *Evolutionary Applications* **1**: 587–597.

Campanella JJ, Bologna PAX, Smalley JV, Avila DN, Lee KN, Areche EC, Slavin LJ. 2012. An analysis of the population genetics of restored *Zostera marina* plantings in Barnegat Bay, New Jersey. *Population Ecology* **55**: 121–133.

Chessel D, Dufour AB, Thioulouse J. 2004. The ade4 package - I : one-table methods. *R News* **4**: 5–10.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

De Kort H, Vandepitte K, Bruun HH, Closset-Kopp D, Honnay O, Mergeay J. 2014. Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Molecular Ecology* **23**: 4709–4721.

De Mendiburu F. 2014. *agricolae: Statistical procedures for agricultural research. R package. version 1.2-1.* [WWW document] URL http://CRAN.R-project.org/package=agricolae [accessed 29 July 2015].

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M *et al.* 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**: 491–498.

Dillon S, McEvoy R, Baldwin DS, Rees GN, Parsons Y, Southerton S. 2014. Characterisation of adaptive genetic diversity in environmentally contrasted populations of *Eucalyptus camaldulensis* Dehnh. (river red gum). *PLoS ONE* **9**: e103515.

Dillon S, McEvoy R, Baldwin DS, Southerton S, Campbell C, Parsons Y, Rees GN. 2015. Genetic diversity of *Eucalyptus camaldulensis* Dehnh following population decline in response to drought and altered hydrological regime. *Austral Ecology* **40**: 558–572.

Dolan RW, Marr DL, Schnabel A. 2008. Capturing genetic variation during ecological restorations: an example from Kankakee Sands in Indiana. *Restoration Ecology* **16**: 386–396.

Ellstrand NC, Elam DR. 1993. Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics* **24**: 217–242.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1**: 47–50.

Fant JB, Kramer A, Sirkin E, Havens K. 2013. Genetics of reintroduced populations of the narrowly endemic thistle, *Cirsium pitcheri* (Asteraceae). *Botany-Botanique* **91**: 301–308.

Geraldes A, Farzaneh N, Grassa CJ, McKown AD, Guy RD, Mansfield SD, Douglas CJ, Cronk QCB. 2014. Landscape genomics of *Populus trichocarpa*: the role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution* **68**: 3260–3280.

Godefroid S, Piazza C, Rossi G, Buord S, Stevens A-D, Aguraiuja R, Cowell C, Weekley CW, Vogg G, Iriondo JM *et al.* 2011. How successful are plant species reintroductions? *Biological Conservation* **144**: 672–682.

Grattapaglia D, Vaillancourt RE, Shepherd M, Thumma BR, Foley W, Külheim C, Potts BM, Myburg AA. 2012. Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. *Tree Genetics & Genomes* **8**: 463–508.

Hancock N, Hughes L. 2012. How far is it to your local? A survey on local provenance use in New South Wales. *Ecological Management & Restoration* **13**: 259–266.

Havens K, Vitt P, Still S, Kramer AT, Fant JB, Schatz K. 2015. Seed sourcing for restoration in an era of climate change seed sourcing for restoration in an era of climate change. *Natural Areas Journal* **35**: 122–133.

Hoffmann A, Griffin P, Dillon S, Catullo R, Rane R, Byrne M, Jordan R, Oakeshott J, Weeks A, Joseph L *et al.* 2015. A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses* **2**: 1.

Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. *Nature* **470**: 479–485.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**: e1000862.

Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW. 2015. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. *New Phytologist* **209**: 1240–1251.

Honnay O, Jacquemyn H. 2007. Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. *Conservation Biology* **21**: 823–831.

Hudson CJ, Freeman JS, Myburg AA, Potts BM, Vaillancourt RE. 2015. Genomic patterns of species diversity and divergence in *Eucalyptus*. *New Phytologist* **206**: 1378–1390.

Hufford KM, Mazer SJ. 2003. Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends in Ecology & Evolution* **18**: 147–155.

James EA, Jordan R. 2013. Limited structure and widespread diversity suggest potential buffers to genetic erosion in a threatened grassland shrub *Pimelea spinescens* (Thymelaeaceae). *Conservation Genetics* **15**: 305–317.

Jombart T, Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**: 3070–3071.

Jump AS, Marchant R, Peñuelas J. 2009. Environmental change and the option value of genetic diversity. *Trends in Plant Science* **14**: 51–58.

Jump AS, Penuelas J. 2005. Running to stand still: adaptation and the response of plants to rapid climate change. *Ecology Letters* **8**: 1010–1020.

Kramer AT, Ison JL, Ashley MV, Howe HF. 2008. The paradox of forest fragmentation genetics. *Conservation Biology* **22**: 878–885.

Kuparinen A, Savolainen O, Schurr FM. 2010. Increased mortality can promote evolutionary adaptation of forest trees to climate change. *Forest Ecology and Management* **259**: 1003–1008.

Lander TA, Boshier DH, Harris SA. 2010. Fragmented but not isolated: contribution of single trees, small patches and long-distance pollen flow to genetic connectivity for *Gomortega keule*, an endangered Chilean tree. *Biological Conservation* **143**: 2583–2590.

Leimu R, Fischer M. 2008. A meta-analysis of local adaptation in plants. *PLoS ONE* **3**: e4010.

Leimu R, Vergeer P, Angeloni F, Ouborg NJ. 2010. Habitat fragmentation, climate change, and inbreeding in plants. *Annals of the New York Academy of Sciences* **1195**: 84–98.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint*: arXiv:1303.3997.

Lischer HEL, Excoffier L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**: 298–299.

Liu M-H, Chen X-Y, Zhang X, Shen D-W. 2008. A population genetic evaluation of ecological restoration with the case study on *Cyclobalanopsis myrsinaefolia* (Fagaceae). *Plant Ecology* **197**: 31–41.

Lloyd MW, Burnett RK, Engelhardt KAM, Neel MC. 2012. Does genetic diversity of restored sites differ from natural sites? A comparison of *Vallisneria americana* (Hydrocharitaceae) populations within the Chesapeake Bay. *Conservation Genetics* **13**: 753–765.

Lowe AJ, Boshier D, Ward M, Bacles CFE, Navarro C. 2005. Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity* **95**: 255–273.

Maschinski J, Wright SJ, Koptur S, Pinto-Torres EC. 2013. When is local the best paradigm? Breeding history influences conservation reintroduction survival and population trajectories in times of extreme climate events. *Biological Conservation* **159**: 277–284.

Menges ES. 2008. Restoration demography and genetics of plants: when is a translocation successful? *Australian Journal of Botany* **56**: 187–196.

Mijangos JL, Pacioni C, Spencer PBS, Craig MD. 2015. Contribution of genetics to ecological restoration. *Molecular Ecology* **24**: 22–37.

Mimura M, Barbour RC, Potts BM, Vaillancourt RE, Watanabe KN. 2009. Comparison of contemporary mating patterns in continuous and fragmented *Eucalyptus globulus* native forests. *Molecular Ecology* **18**: 4180–4192.

Monks L, Coates D, Bell T, Bowles M. 2012. Determining success criteria for reintroductions of threatened long-lived plants. In: Maschinski J, Haskins KE, eds. *Plant reintroduction in a changing climate: promises and perils*. Washington, DC, USA: Island Press, 189–208.

Mosca E, Eckert AJ, Di Pierro EA, Rocchini D, La Porta N, Belletti P, Neale DB. 2012. The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. *Molecular Ecology* **21**: 5530–5545.

Moura AE, Kenny JG, Chaudhuri R, Hughes MA, Welch A, Reisinger RR, De Bruyn PJN, Dahlheim ME, Hall N, Rus Hoelzel A. 2014. Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Molecular Ecology* **23**: 5179–5192.

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D *et al.* 2014. The genome of *Eucalyptus grandis*. *Nature* **510**: 356–362.

Neale JR. 2012. Genetic considerations in rare plant reintroduction: practical applications (or How are we doing?). In: Maschinski J, Haskins KE, eds. *Plant reintroduction in a changing climate: promises and perils*. Washington, DC, USA: Island Press, 71–88.

Ottewell KM, Donnellan SC, Paton DC. 2010. Evaluating the demographic, reproductive, and genetic value of Eucalypt paddock trees for woodland restoration in agricultural landscapes. *Restoration Ecology* **18**: 263–272.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics* **28**: 2537–2539.

Pickup M, Field DL, Rowell DM, Young AG. 2012. Predicting local adaptation in fragmented plant populations : implications for restoration genetics. *Evolutionary Applications* **5**: 913–924.

Prober SM, Brown AHD. 1994. Conservation of the grassy White Box woodlands: population genetics and fragmentation of *Eucalyptus albens*. *Conservation Biology* **8**: 1003–1013.

Prober SM, Byrne M, McLean EH, Steane DA, Potts BM, Vaillancourt RE, Stock WD. 2015. Climate-adjusted provenancing: a strategy for climate-resilient ecological restoration. *Frontiers in Ecology and Evolution* **3**: 65.

Prunier J, Laroche J, Beaulieu J, Bousquet J. 2011. Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular Ecology* **20**: 1702–1716.

Pryor LD. 1976. *The biology of Eucalypts*. London, UK: Edward Arnold Ltd.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ *et al.* 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**: 559–575.

R Core Team. 2015. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [WWW document] URL https://www.r-project.org/ [accessed 29 July 2015].

Reed DH, Frankham R. 2003. Correlation between fitness and genetic diversity. *Conservation Biology* **17**: 230–237.

Reynolds L, Waycott M, McGlathery K, Orth R, Zieman J. 2012. Eelgrass restoration by seed maintains genetic diversity: case study from a coastal bay system. *Marine Ecology Progress Series* **448**: 223–233.

Ritchie AL, Krauss SL. 2012. A genetic assessment of ecological restoration success in *Banksia attenuata*. *Restoration Ecology* **20**: 441–449.

Rousset F. 1997. Genetic differentiation. *Genetics* **145**: 1219–1228.

Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A. 2011. Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proceedings* **5**: 54.

Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics* **38**: 595–619.

Sgrò CM, Lowe AJ, Hoffmann AA. 2011. Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications* **4**: 326–337.

Smulders MJM, Schoot J, Geerts RHEM, Antonisse-de Jong AG, Korevaar H, Werf A, Vosman B. 2000. Genetic diversity and the reintroduction of meadow species. *Plant Biology* **2**: 447–454.

Spooner PG, Lunt ID, Smallbone L. 2010. Using growth-form attributes to identify pre-settlement woodland trees in central NSW, Australia. *Australian Journal of Botany* **58**: 188–197.

Steane DA, Nicolle D, Sansaloni CP, Petroli CD, Carling J, Kilian A, Myburg AA, Grattapaglia D, Vaillancourt RE. 2011. Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping. *Molecular Phylogenetics and Evolution* **59**: 206–224.

Steane DA, Potts BM, McLean E, Prober SM, Stock WD, Vaillancourt RE, Byrne M. 2014. Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular Ecology* **23**: 2500–2513.

Vranckx G, Jacquemyn H, Muys B, Honnay O. 2012. Meta-analysis of susceptibility of woody plants to loss of genetic diversity through habitat fragmentation. *Conservation Biology* **26**: 228–237.

Weeks AR, Sgro CM, Young AG, Frankham R, Mitchell NJ, Miller KA, Byrne M, Coates DJ, Eldridge MDB, Sunnucks P *et al.* 2011. Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications* **4**: 709–725.

Willi Y, Van Buskirk J, Hoffmann AA. 2006. Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution, and Systematics* **37**: 433–458.

Yokogawa M, Kaneko S, Takahashi Y, Isagi Y. 2013. Genetic consequences of rapid population decline and restoration of the critically endangered herb *Polemonium kiushianum*. *Biological Conservation* **157**: 401–408.

Young A, Boyle T, Brown T. 1996. The population genetic consequences of habitat fragmentation for plants. *Tree* **11**: 413–418.

Zhou L, Bawa R, Holliday JA. 2014. Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cottonwood (*Populus trichocarpa*). *Molecular Ecology* **23**: 2486–2499.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Boxplot of diameter at breast height (DBH) for all *Eucalyptus microcarpa* trees sampled at natural sites.

Fig. S2 Initial PCA for *Eucalyptus microcarpa* to identify outlier samples using 6490 'globally filtered' SNPs.

Fig. S3 Folded site frequency spectrum per location per site type for *Eucalyptus microcarpa*.

Fig. S4 Expected heterozygosity by chromosome for *Eucalyptus microcarpa*.

Fig. S5 Percentage polymorphic loci by chromosome for *Eucalyptus microcarpa*.

Table S1 Size (kbp) and number of variants identified in *Eucalyptus microcarpa* for the 11 major Eucalytpus scaffolds (chromosomes) and minor scaffolds

Table S2 Percentage polymorphic loci per site in *Eucalyptus microcarpa*, unadjusted and adjusted for smallest sample of six

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.