



Genome-wide scans detect adaptation to aridity in a widespread forest tree species

Steane, Dorothy A; Potts, Brad M; McLean, Elizabeth; et.al.

<https://research.usc.edu.au/esploro/outputs/journalArticle/Genome-wide-scans-detect-adaptation-to-aridity/99448983202621/filesAndLinks?index=0>

Steane, D. A., Potts, B. M., McLean, E., Prober, S. M., Stock, W. D., Vaillancourt, R. E., & Byrne, M. (2014). Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular Ecology*, 23(10), 2500–2513. <https://doi.org/10.1111/mec.12751>
Document Type: Accepted Version

Link to Published Version: <https://doi.org/10.1111/mec.12751>

UniSC Research Bank: <https://research.usc.edu.au>
research-repository@usc.edu.au

Copyright © 2014 Wiley-Blackwell Publishing Ltd. This is the accepted version of the following article: Steane, D. A., Potts, B. M., McLean, E., Prober, S. M., Stock, W. D., Vaillancourt, R. E. and Byrne, M. (2014), Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular Ecology*, 23: 2500-2513. doi: 10.1111/mec.12751, which has been published in final form at <http://dx.doi.org/10.1111/mec.12751>

Downloaded On 2024/05/02 13:45:16 +1000

Received Date : 23-Dec-2013

Revised Date : 21-Mar-2014

Accepted Date : 30-Apr-2014

Article type : Original Article

Genome-wide scans detect adaptation to aridity in a widespread forest tree species

Dorothy A. Steane^{1,2,6}, Brad M. Potts^{1,6}, Elizabeth McLean³, Suzanne M. Prober⁴,
William D. Stock⁵, René E. Vaillancourt¹ and Margaret Byrne³

¹ School of Biological Sciences and National Centre for Future Forest Industries,
University of Tasmania, Private Bag 55, Hobart, Tasmania 7000, Australia

² Faculty of Science, Health, Education and Engineering and Collaborative Research
Network, University of the Sunshine Coast, Locked Bag 4, Maroochydore,
Queensland, 4558, Australia

³ Science and Conservation Division, Department of Parks and Wildlife, Locked Bag
104, Bentley Delivery Centre, WA 6983 Western Australia

⁴ CSIRO Ecosystem Sciences, Western Australia

⁵ Centre for Ecosystem Management, School of Natural Sciences, Edith Cowan
University, Western Australia

⁶ Joint first author

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.12751

This article is protected by copyright. All rights reserved.

Key Words: *Eucalyptus*, genome-wide scan, outlier analysis, Canonical Analysis of Principal Coordinates, assisted migration, climate-resilience

Corresponding author: Dorothy Steane, School of Plant Science, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001, Australia. Fax: -61-3-62262698. Dorothy.Steane@utas.edu.au

Running title: Genomic adaptation in *Eucalyptus*

Abstract

Patterns of adaptive variation within plant species are best studied through common garden experiments, but these are costly and time-consuming, especially for trees that have long generation times. We explored whether genome-wide scanning technology combined with outlier marker detection could be used to detect adaptation to climate and provide an alternative to common garden experiments. As a case study, we sampled nine provenances of the widespread forest tree species, *Eucalyptus tricarpa*, across an aridity gradient in southeastern Australia. Using a Bayesian analysis we identified a suite of 94 putatively adaptive (outlying) sequence-tagged markers across the genome. Population-level allele frequencies of these outlier markers were strongly correlated with temperature and moisture availability at the site of origin, and with population differences in functional traits measured in two common gardens. Using the output from a canonical analysis of principal coordinates we devised a metric that provides a holistic measure of genomic adaptation to aridity that could be used to guide assisted migration or genetic augmentation.

Introduction

Recent climate change projections suggest that 2°C global average warming by the end of this century is inevitable, and an increase of at least 4°C is possible (Peters *et al.* 2013). Changes in precipitation and more frequent extreme climatic events, such as droughts and heatwaves, are also predicted (Ganguly *et al.* 2009). Syntheses clearly demonstrate that global warming is already impacting on ecosystems and biodiversity around the world (Chen *et al.* 2011; Parmesan 2006). Thus, new approaches are needed to predict the climate-resilience of species that are vital to ecosystem stability and to enhance ecological management and restoration.

Widespread tree species are important targets when investigating resilience to future climate change because of their foundational role in maintaining ecological function (Lunt *et al.* 2013) and because they may contain genetic material of unrealised adaptive potential.

A ubiquitous practice towards managing and restoring terrestrial ecosystems is to re-establish indigenous or other desirable plant species (e.g. Breed *et al.* 2013; Vallejo *et al.* 2012). This can include extensive plantings to revegetate degraded landscapes and restore biodiversity, sometimes in conjunction with production uses, such as perennial cropping systems, agro-forestry or carbon sequestration (Smith *et al.* 2013). However, investments in revegetation activities currently take little account of climate change (Vallejo *et al.* 2012; Breed *et al.* 2013), and typically focus on maintaining local patterns of genetic variation, assuming that local populations are maximally adapted to local conditions and aiming to avoid any risk of outbreeding depression (Broadhurst *et al.* 2008; Byrne *et al.* 2011; Hereford 2009). Under changing environmental conditions local germplasm may not be optimally

Accepted Article

adapted to future conditions and a static view of 'local is best' may no longer be relevant (Breed *et al.* 2013). Consequently, a new genetic framework is needed to promote climate-resilient revegetation, based on a predictive understanding of patterns of adaptive variation in plant species and future climates (Booth 2013).

Historically, patterns of adaptive variation within plant species have been studied through physiological, morphometric and fitness comparisons in common gardens across climatic gradients (Alberto *et al.* 2013; Kremer *et al.* 2013; Savolainen *et al.* 2013). However, this approach can be expensive and may take years to yield results, especially with organisms, such as trees, that have long generation times (Neale & Kremer 2011). Recent developments in genomic technologies allow greater insights into patterns of adaptive variability (Neale & Kremer 2011; Savolainen *et al.* 2013; Strasburg *et al.* 2012) and have the potential to form the basis of tools determining patterns of adaptive diversity in plant species (Alberto *et al.* 2013; Allendorf *et al.* 2010; Neale & Kremer 2011). Indeed, the development of predictive models of adaptation based on genome-wide scans - that do not need individual loci to be identified - have been envisaged, analogous to genomic selection methods used in breeding (Alberto *et al.* 2013).

To date, most genomic studies of local adaptation in trees have used either a candidate gene approach (Eckert *et al.* 2009; Strasburg *et al.* 2012) or a large number of anonymous, mainly neutral markers, e.g., AFLPs (Jump *et al.* 2006; Strasburg *et al.* 2012) and SSRs (Strasburg *et al.* 2012). In the candidate gene approach, discovery is usually limited by the finite number of genes being screened (except in some model organisms, e.g., *Arabidopsis*; Hancock *et al.* 2011). At the other end of the spectrum, large numbers of anonymous markers have been used to

Accepted Article

produce evidence of local adaptation (see Strasburg *et al.* 2012) but, by their very nature, these lack the potential to identify the underlying genomic regions that are under selection. With the increasing accessibility of next-generation sequencing, new methods (e.g., RAD sequencing (Davey & Blaxter 2010), DArTseq (Sansaloni *et al.* 2011)) have been developed that provide non-model organisms with genome-wide sets of sequence-tagged markers that can be linked to a genomic database, thereby providing scope to identify regions of a genome that may be linked to local adaptation (Stapley *et al.* 2010).

Eucalypts are dominant tree species in many ecosystems and are planted worldwide. Thus, an understanding of the genetic basis for adaptation to climate is critical for climate-resilient forest management and reforestation. *Eucalyptus tricarpa* (red ironbark) is an ecologically important widespread species that grows across a rainfall gradient in agricultural and forest landscapes of southeastern Australia. The previous establishment (Stackpole & Harwood 2001) of a range of provenances of *E. tricarpa* in two common gardens located near each end of the gradient provides an excellent study system to investigate local adaptation and phenotypic plasticity in a widespread eucalypt. A recent study McLean *et al.* (2014) found evidence of local adaptation in functional traits in *E. tricarpa*, leading to a hypothesis of genetic-based climate adaptation. Our aim in the present study was to test this hypothesis using an indirect approach, where we use genomic scans and outlier marker detection to detect signals of adaptation to aridity and cross-reference our results with direct evidence from common garden field trials. We propose that it might be feasible to use such an approach to bypass the need for common garden field trials for assessments of local adaptation in widespread non-model species.

This is particularly relevant when time and resources are limited, and we here explore a framework to implement such an approach.

Materials and methods

Study species, field sites and common gardens

Eucalyptus tricarpa (red ironbark) is an important component of Victoria's threatened box-ironbark forests (Kelly & Mercer 2005) and is used for revegetation. The species varies in form across a moisture gradient, growing as dry woodlands in the northwest (mean annual precipitation [MAP] of 450 mm) and as tall, dense, wet sclerophyll forests in the southeast (MAP of 1200 mm). We selected three provenances from the 'dry' end of the aridity gradient (1 Tarnagulla, 2 Mt Bealiba, 3 Craigie); three provenances from the 'wet' end of the gradient (7 Mt Nowa Nowa, 8 Tuckerbox and 9 Martins Creek); and three provenances between the two extremes (4 Heathcote, 5 Heyfield, 6 Christmas Hills). Provenances were numbered from 1 (driest) to 9 (wettest) according to the MAP at the site of origin (Figure 1, Table 1).

Two common garden field trials of *E. tricarpa* were established in 2000 using individual tree, open-pollinated seedlots (families) collected from multiple native provenances from throughout the natural range in southeastern Australia (Stackpole & Harwood 2001). The trials were planted towards each end of the rainfall gradient (Figure 1) and included the nine provenances that were sampled in the wild. The trials have been thinned to 60% of the original planting density of each family, on the basis of tree size (but not form) (see McLean *et al.* 2014).

Analysis of functional traits

A study of physiological and morphometric traits (hereafter ‘functional traits’) across the aridity gradient, based on measurements taken from eight to 10 trees in each of the two field trials and 10 trees from each wild population, was conducted in parallel with the genomic study. Traits are listed in Table 2. The results of the functional trait study are reported elsewhere (McLean *et al.* 2014).

Analysis of Variance (ANOVA), using the PROC GLM procedure of SAS (SAS Institute; Version 9.1) with a one-factor fixed effects model, was used to test whether there was significant variation in functional traits among provenances in the wild and/or in the common gardens. To account for multiple testing, for each class of response variable (i.e., within ‘wild provenances’, and each trial), probabilities were corrected for a 5% False Discovery Rate (FDR) using a conservative ‘dependent’ FDR method that allows for correlation between tests (DFDR; Benjamini & Yekateuli 2001), using PROC MULTTEST of SAS with the DFDR option. Analysis of covariance, fitting site, a covariate (CAP1, see below) and site x CAP1 interaction was also undertaken using PROC GLM.

Genomic methods

Foliage samples were collected from the nine wild provenances across the rainfall gradient (Figure 1, Table 1). Thirty adult trees were sampled at each locality, except in provenance 2 Mount Bealiba and provenance 6 Christmas Hills where 35 and 29 adult trees, respectively, were sampled. To avoid sampling close relatives within a provenance, samples were taken from trees that were separated by a minimum distance of 100 m (at least two tree heights) (Skabo *et al.* 1998). The location of

each tree was recorded using a Garmin GPSMAP 60CSx (accuracy ranging from 4 to 10 m).

DNA was extracted from leaf samples by the Australian Genome Research Facility (Adelaide, Australia), using an in-house CTAB protocol. DNA from each tree was standardised to approximately 50 ng μl^{-1} and sent to Diversity Arrays Technology Pty. Ltd. (Canberra, Australia) for genotyping using DArTseq technology (Sansaloni *et al.* 2011). DArTseq markers are expected to have many of the same qualities as DArT markers: dominant in nature, random dispersal across a genome and a large proportion coming from coding regions (Petroli *et al.* 2012). To ensure that all DArTseq markers were highly reproducible, only those with a 'Q' value > 2.5 ('signal-to-noise' measure comprising the ratio of the average number of counts for a sequence among the samples and their standard deviation) and a Call Rate $\geq 90\%$ (percentage of valid scores made across all samples) were included in the final data set (about 50% of the markers were discarded).

Defining an adaptively enriched genetic space

Outlier loci representative of diversifying selection (i.e., loci whose allele frequencies differ more among the nine provenances than would be expected through drift alone) were identified using Bayescan v. 2.1 (Foll & Gaggiotti 2008). Although the island model of allele frequency correlation is an underlying assumption of this software, various simulation studies have indicated that Bayescan is one of the most robust differentiation methods for outlier detection (Narum & Hess 2011; Perez-Figueroa *et al.* 2010; Savolainen *et al.* 2013; Vilas *et al.* 2012), even when this assumption is violated, for example, in a population with hierarchical structure (De Mita *et al.* 2013). Markers with a low '1-Ratio' (i.e. 'allele' or 'band' frequency < 0.10 or > 0.90) were

excluded from the Bayescan analysis, as recommended by the software manual, leaving 3,684 markers. Default parameters were used for the Bayescan searches, prior odds for the neutral model were set to 100 and the F_{IS} prior was set to 'uniform between 0.0 and 0.3' in accordance with typical values of this inbreeding statistic for eucalypts (Byrne 2008). A 'q value' maximum threshold of 0.05 was used to identify outlier loci (this yielded a FDR of approximately 5%), which were used to define an 'adaptively enriched genetic space'.

The full DArTseq data set and the data set comprising 94 outlier loci were each analysed in a population genetics framework, using GenAlEx 6.41 (Peakall & Smouse 2006) to generate matrices of pairwise genetic distances among individuals (Huff *et al.* 1993) and provenances (the widely-used Nei (1972) genetic distance; Hedrick 2000; Nei 1972). The matrix of genetic distance calculated using outlier markers and all individuals is called the 'adaptive genetic distance matrix'. GenAlEx 6.41 was also used to perform Analyses of Molecular Variances (AMOVAs) and Mantel tests for correlation between provenance-level genetic distance (Nei's genetic distance; Hedrick 2000; Nei 1972) and geographic distance.

Thirty-five climatic variables (Supporting Information 1) for each tree position (based on GPS coordinates) were derived from climatic surfaces in the ANUCLIM 6.1 software package (Xu & Hutchinson 2011). ANUCLIM has a resolution of about 1 km and incorporates altitude data into its climate estimates (such that estimates within a 1 km grid can vary depending on altitude). In addition, a 'moisture index' (ratio of precipitation to potential evaporation) from the Atlas of Living Australia (ALA) (<http://www.ala.org.au/>) was also used. Climatic variables were normalised using Primer-E (Clarke & Gorley 2006).

Accepted Article

For each marker in both 'outlier' and 'neutral' data sets, linear regression was used to identify significant associations between the population-level allele frequencies and environmental variables as well as 14 functional traits (Table 2). A chi-square goodness-of-fit test was used to determine whether there were significant differences between the rate of association discovery in adaptive markers *versus* neutral markers for classes of climatic variables (rainfall, radiation, temperature, moisture indices), soil variables or functional traits.

Aligning the adaptively enriched genetic space with environment

The matrices of pairwise genetic distances among individuals that were calculated in GenAlEx 6.41, were analysed using Permanova software (Anderson *et al.* 2008). Primer-E (Clarke & Gorley 2006) was used to identify sets of highly correlated climatic variables (Supporting Information 2). Of 35 climatic variables from ANUCLIM, a subset of the 15 least intercorrelated, representing four classes, was used for canonical analyses of principal coordinates (CAP, see below): (i) temperature (mean annual temperature, mean diurnal range, maximum temperature of the warmest period, minimum temperature of the coldest period, mean temperature of the driest quarter), (ii) rainfall (mean annual precipitation, precipitation of the wettest period, precipitation of the coldest quarter), (iii) solar radiation (mean annual radiation, radiation seasonality, radiation of the driest quarter, radiation of the coldest quarter), and (iv) moisture indices (annual mean moisture index, highest period moisture index, mean moisture index of highest quarter).

Principal coordinates analyses (PCoAs) and canonical analyses of principal coordinates (CAP; Anderson & Robinson 2003; Anderson & Willis 2003) were undertaken with Permanova (Anderson *et al.* 2008). CAP is a classical canonical

Accepted Article

correlation analysis between a subset of the principal coordinates that summarises a resemblance matrix and a set of environmental variables associated with each object. The purpose of the CAP was to find axes through the multivariate cloud of points (corresponding to the adaptively enriched genetic space of *E. tricarpa*) that had the strongest correlation with another set of continuous variables, in this case, climatic variables (above) and soil variables (describing texture, chemistry - NH_4 , NO_3 , P, K, S, organic C, conductivity, pH in CaCl_2 , pH in water - and soil particle size - clay, coarse sand, sand, fine sand, silt) (McLean *et al.* 2014). The first axis of each of these analyses, CAP1, represented the direction of molecular change most closely associated with change in environmental variables. This strategy for alignment of a molecular and environmental space is conceptually similar to the approach undertaken by Parisod & Christin (2008). CAP1 provided what we hereafter term an 'adaptive genetic index' that was used as a quantitative predictor to model genetic adaptation to either aridity or soil. Since the soils showed no correlation with the adaptively enriched genetic variation (see Results), we henceforth refer to the adaptive genetic index (CAP1) with regard to climate.

As the similarity in environmental variables among trees from the same provenance (arising from the 1 km precision of the ANUCLIM climate surface) may result in pseudoreplication, the strength of the CAP1 association with environmental variables was confirmed by linear regression analyses at the provenance level ($n=9$). We tested all climatic variables at this stage to 'determine whether there were variables other than those used in the CAP that were also closely associated with CAP1. In each test, the dependent variable was CAP1 and the independent variable was the environmental variable. The environmental variables were divided into two classes: (a) 35 climatic variables (Supporting Information 1) plus a 'moisture index' from ALA;

and (b) 15 soil variables (Table 2; Supporting Information 3). Linear regression was done at the provenance level (n=9), using the PROC REG procedure of SAS. DFDR corrections for multiple testing were made within each class of environmental variables, as previously detailed. Similarly, we used linear regression to test whether the adaptive genetic index (CAP1) was associated with any functional traits, using provenance means for functional trait measurements from three data sets (wild provenances, Lake Tyers Trial and Huntly Trial), and controlling for multiple testing using a 5% DFDR within each of the three data sets.

The genetic architecture of adaptation

Each DArTseq marker is tagged by approximately 60 bp of DNA sequence data. The DArTseq analytical service included a BLAST analysis of each marker against the *E. grandis* complete genome sequence (JGI ver. 1.1), providing information about the position of each marker on a linkage group (chromosome). Most of the markers gave a single alignment, and some could not be located at all on the *E. grandis* genome. When more than one alignment was obtained, the alignment with the best score was used to assign a position to the marker. To test whether outlier markers were clustered or distributed randomly across the genome, we:

- (a) tested for differences in the relative frequency of outlier markers between chromosomes using PROC GLIMMIX of SAS to fit a single factor binary model with a logit link function;
- (b) tested whether there were more sequential runs (clusters) of outlier markers along each chromosome (with >3 outlier markers) than would be expected by chance alone using the Wald-Wolfowitz Test for Randomness (Mendenhall *et al.* 1986); and

(c) identified 3 Mb fragments of the genome that had a higher density of outlier markers than would be expected by chance alone, with 10,000 random allocations to chromosomes giving a frequency distribution for the random expectation.

Matching provenances to sites to which they are best adapted

To determine whether the results of the CAP could be used for predictive purposes, we used the weightings (canonical eigen vectors) derived from CAP of the 15 climatic variables to calculate a genetic-based Aridity Index (AI) for each site of interest, using the following formula (adapted from equation 5.14 in Permanova manual):

$$AI = \sum a_i b_i$$

where a = normalised climatic variable x, and b = the weighting of the climate variable on the canonical eigenvector aligned with CAP.

The AI for each of the collection sites and the two common gardens was calculated, to predict (i) provenances that would perform well in the common gardens and (ii) patterns of adaptation to aridity across the geographic range. The AIs of the common garden sites were used – through site matching – to predict the provenances that should perform best in each garden, based on the age 13 growth data for the trees from which functional traits were measured (McLean *et al.* 2014).

Spatial analysis

Geographic locations of all records of *E. tricarpa* were downloaded from ALA (accessed 18 September 2013), and outlying samples likely to be errors were

removed. The distribution area was defined by drawing a polygon encompassing a 15 km buffer around each retained record (see Figure 6). Ordinary kriging (Schabenberger and Gotway, 2004) was used in ArcGIS 10.1 (ESRI, 2012) to interpolate the AIs calculated for each ALA record and each provenance study site, and create a continuous surface of AI for the distribution area. A spherical semivariogram, with variable search radius and considering the values of 12 neighbouring records, was used for the spatial calculations (Schabenberger & Gotway 2004).

Results

Provenance differentiation

AMOVA of the full *E. tricarpa* DArTseq data set (274 individual trees by 6,544 markers; 4.6% missing data) revealed significant molecular differentiation between provenances ($P < 0.001$) accounting for 7% of the total molecular variation. The genetic distances among provenances were strongly positively correlated with geographic distances ($r^2 = 0.72$; $P = 0.001$), with distinct spatial structuring of genetic variation from east to west, along the prevailing aridity gradient (Figure 2A).

Defining an adaptively enriched genetic space

Bayescan 2.1 identified 94 loci (2.6% of the Bayescan input) that had outlier F_{st} values indicative of disruptive selection (none of the outlier loci had negative alpha values indicative of stabilizing selection; see Supporting Information 4 for F_{st} distribution data). A greater proportion of the variation in the PCoA was accounted for by the first dimension (49%) when using the outlier data set (Figure 2B)

compared to the PCoA when using the full data set (7.3%; Figure 2A). AMOVA of the outlier data set showed that 36% of the variation between trees could be accounted for by differences among provenances, a five-fold increase over that observed in the full data set. There was still a strong correlation between provenance-level geographic and genetic distances ($r^2 = 0.81$; $P < 0.001$; $n=9$), but this was expected from the northwest-southeast alignment of the aridity gradient.

Regression analyses using provenance-level data showed a significant ($X^2_1 = 3214$; $P < 0.001$) eight-fold increase in the proportion of outlier loci exhibiting a significant association with provenance variation in the climatic variables compared to the remaining putatively 'neutral' markers (Table 2). Testing the subset of 15 climatic variables used for the CAP (below), the enrichment was similar, at seven-fold ($X^2_1 = 1392$; $P < 0.001$). The enrichment of climate-marker associations in the outlier data set was highest for ANUCLIM moisture indices, followed by radiation variables. No significant enrichment was observed for variables describing the physical and chemical soil properties (Table 2), indicating that the main signal of selection on these outlier markers is not associated with soil properties. Further evidence that the outlier markers show signals of climate adaptation was provided by their association with genetic-based provenance variation in functional traits (Table 2). Most of the functional traits assessed from trees growing in the common-environment field trials exhibited significant ($P < 0.05$) differences among provenances (Supporting Information 5), demonstrating significant quantitative genetic differentiation across the species' range. While only 1.5% of provenance-level regressions between the outlier markers and trial traits were significant ($P < 0.001$), this was a significant ($X^2_1 = 68.2$; $P < 0.001$) three-fold increase over that observed with the putatively neutral loci (Table 2).

A climate-aligned adaptive genetic index

The climate-based adaptive genetic variation amongst trees was summarised by the CAP, undertaken using the adaptive genetic distance matrix and the set of 15 representative climatic variables for the 274 individual trees. Eighty-three percent of the variance in the adaptive genetic distance matrix for individuals was explained by the optimum number of PCo axes for this analysis ($m=13$; see Anderson *et al.* (2008)). The integrated molecular change defined by the main CAP1 axis was strongly correlated with the climatic data set (squared canonical correlation, $\delta^2 = 0.98$; $P=0.001$), and accounted for 50% of the variation in the adaptive genetic distance matrix. This compares with only an additional 7% being explained by CAP2 ($\delta^2 = 0.79$; $P=0.001$). Molecular variation along CAP1 reflected the aridity ranking of the provenances, with trees from provenances from wetter regions having more negative CAP1 values and those from the drier regions having more positive values (Figure 2C). Thus, CAP1 represents a 'climate-aligned adaptive genetic index', with increasing values along CAP1 aligning closely with vectors for climatic variables that represent high temperature and low rainfall (Figure 2C).

At the provenance level, linear regressions (Supporting Information 6) indicated 23 (66%) of the original 35 climatic variables were significantly associated with change in CAP1 (after DFDR adjustment; $P<0.05$). For 13 (37%) of the variables, the associations were significant ($P<0.001$). Increases in CAP1 at the provenance level were most closely associated with an increase in the maximum temperature in the warmest month (TMXWM, $R^2 = 0.98$) and decreases in the mean rainfall in the driest periods (RDRYQ, $R^2 = 0.96$; RDRYM, $R^2 = 0.96$), the lowest period moisture index

(MIL, $R^2 = 0.97$) and the mean moisture index of the lowest quarter (MIMLQ, $R^2 = 0.97$).

While variation along CAP1 was not associated with significant phenotypic differences among provenances in traits measured from the wild (Supporting Information 7), it was associated with genetically-based differences in functional traits measured in the common gardens (Supporting Information 8). At the dry common garden (Huntly), the only significant associations found were greater leaf thickness (differences among provenances $F_{8,80} = 8.5$, $P < 0.001$; association with CAP1, $R^2 = 0.82$, $P(\text{DFDR}) < 0.05$) and lower SLA ($F_{8,80} = 7.5$, $P < 0.001$; $R^2 = 0.79$, $P(\text{DFDR}) < 0.05$) with increasing values of CAP1 (i.e., increasing molecular adaptation to aridity). At the wet common garden (Lake Tyers), the only significant association found was between CAP1 and leaf size, where provenances with higher CAP1 values tended to have genetically smaller leaves ($F_{8,76} = 6.1$, $P < 0.001$; $R^2 = 0.81$, $P(\text{DFDR}) < 0.05$) (Figure 3). The trial-specific nature of these associations was confirmed by analysis of covariance (Table 3) that showed differences in the slopes of the provenance-level relationships between CAP1 and these functional traits in the two common gardens. Thus, the provenance-level changes in the outlier loci, as described by CAP1, were strongly associated with climate variation and quantitative genetic-based changes in functional traits, arguing that CAP1 describes molecular change associated with adaptation of the provenances to increasing aridity.

Further evidence that provenance variation in CAP1 is adaptive comes from the relative growth of the provenances at the two common garden sites. Analysis of covariance (Table 3) indicated that the slope of the provenance-level relationships differed significantly between the wet (Lake Tyers) and dry (Huntly) sites for CAP1

and growth, as measured by either tree cross sectional area (trial by CAP1 interaction $F_{1,14} = 10.2$, $P < 0.01$) or tree height ($F_{1,14} = 9.8$, $P < 0.01$). This result indicates significant genotype-by-environment interactions for growth traits, and provides independent evidence that variation in CAP1 has fitness consequences. At the Lake Tyers (wet) common garden, there were significant differences in performance among provenances (height, $F_{8,76} = 4.6$, $P < 0.001$; total cross sectional area, $F_{8,76} = 7.2$, $P < 0.001$), with those originating from wetter sites showing better performance than those from drier sites. For example, provenance-level regression of total cross sectional area ($R^2 = 0.48$, $P = 0.04$) and tree height ($R^2 = 0.59$, $P = 0.02$) against CAP1 values yielded significant negative correlations, indicating that the wet-adapted provenances (negative values of CAP1) grew taller and had a greater total cross sectional area than dry-adapted provenances when grown at the wet trial site. Similarly, at the dry site (Huntly) there were significant, or near significant, differences among provenances in performance (height, $F_{8,80} = 2.8$, $P < 0.01$; cross-sectional area, $F_{8,80} = 1.8$, $P < 0.1$). For example, the two best performing provenances were those originating from the two most arid sites (1 Tarnagulla and 2 Bealiba), and CAP1 values were positively linearly associated with cross-sectional area ($R^2 = 0.44$, $P < 0.05$). Thus, a provenance's CAP1 value – its climate-aligned adaptive genetic index – appears to predict provenance growth and performance to some degree, with more arid-adapted provenances tending to perform relatively better on more arid sites and more wet-adapted provenances performing better on wetter sites.

At the provenance level, all 94 outlier loci were correlated strongly with either climate or functional traits or both (Supporting Information 9). Twenty-nine percent of the outlier markers showed strong, significant correlations with CAP1 ($P < 0.001$), in contrast to $< 1\%$ of neutral DArTseq markers (i.e., greater than 1000-fold enrichment in the outlier loci; Table 2). Strong associations were found between 82% of the outlier markers and genetically-based differences amongst provenance functional traits (78% in the dry trial, 77% in the wet trial). Seventy markers (75%) were associated with both climate and genetic traits (Supporting Information 9).

Of the 6,544 markers used in this study, 4,134 could be located on the *E. grandis* genome sequence and 3,808 of these mapped to the 11 main eucalypt chromosome scaffolds (Figure 5; Supporting Information 9). Of the 94 outlier loci, 48 mapped to these 11 chromosomes, four mapped to other contigs (those that have yet to be assigned to one of the 11 chromosomes) and 42 could not be located on the *E. grandis* genome sequence (possibly, in part, due to differences in sequence between *E. tricarpa* and *E. grandis*). The neutral markers provided reasonably uniform coverage of the genome (Figure 5). Outlier markers associated with adaptation to aridity occurred on all chromosomes and, while the proportion per chromosome varied 10-fold (from 0.27% on chromosome 2 to 2.7% on chromosome 4), these differences were not statistically significant (GLIMMIX $F_{10, 3797} = 1.28$, $P = 0.2376$). Nevertheless, several lines of evidence indicate that the outlier markers were not distributed randomly throughout the genome. Firstly, tests for sequential runs of outliers along each chromosome revealed significant clumping on chromosome 8 (Wald-Wolfowitz Z value = -6.74, $P < 0.001$ after accounting for multiple testing

through a Bonferroni adjustment). Secondly, 10,000 random allocations of outliers to the DArT marker positions that could be mapped onto 3 Mb fragments of the chromosomes shown in Figure 5, indicated that an observed 3 Mb fragment with five outlier markers (see below) was highly unlikely through chance alone. This conclusion was reached regardless of whether randomisation was undertaken across ($P=0.0013$) or within ($P=0.0028$) chromosomes. In contrast, the frequency of 3 Mb fragments observed with three outlier markers was not significantly greater than expected by chance ($P>0.05$). Evidence of non-random distribution of the outlier markers along the genome is further support for selection impacting regions of the genome in which these outlying markers occur.

The statistically significant cluster of adaptive markers was at the end of chromosome 8 and encompassed a block of 6 markers (Figure 5). Four of these outlier markers lay within 403 kb of each other and another two markers were within 2.9 Mb of these four (Supporting Information 9). Comparison with the *E. grandis* annotated genome sequence showed that these markers lie in the vicinity of a Histidine Kinase, a putative microRNA (potentially involved in leaf longevity and leaf morphogenesis) and a Cytochrome P450 gene, all of which could have adaptive significance across an aridity gradient. Nevertheless, while there may be localised hotspots for adaptation, the widespread distribution of the outlier markers across the eleven chromosomes argues that adaptation to aridity is a genome-wide phenomenon and is likely to involve multiple traits and genes.

Matching provenances to sites to which they are best adapted

In order to extrapolate the identified adaptive change associated with CAP1 to the whole geographic range of *E. tricarpa* we calculated the AI for all known locations of

E. tricarpa, using the 15 normalised climatic variables used in the original CAP.

Values of AI were initially calculated for the sites of origin of each provenance and for the common gardens. As expected, the trend in growth patterns described for CAP1 is reflected by the AI; in general, the better performing provenances in each trial were from sites with AI most similar to those of the trial sites (Figure 4).

Calculating the AI for other known locations of the species allows visualisation of predicted climate adaptation across the *E. tricarpa* distribution; Figure 6 shows a clear trend from highly positive AIs (i.e., dry-adapted provenances) in the northwest of the range through to highly negative values (i.e., wet-adapted) in the east of the range. There are patches of higher/lower rainfall in the middle of the range that do not fit with the general west-to-east trend.

The AI exhibited an 11-fold enrichment in marker associations with outlier markers compared to neutral markers (Table 2). While the AI calculated here was derived from 15 selected climatic variables, single climatic variables, such as the lowest period moisture index (MIL) or maximum temperature of the warmest period (TMXWM) could provide good surrogates for practical application in this case. For example, MIL was strongly correlated ($P<0.001$) at the provenance level with CAP1 ($R^2=0.97$) and with 30% of the outlier markers. TMXWM was also strongly correlated with CAP1 ($R^2=0.96$) and 21% of the outlier markers.

Discussion

Many authors have signalled the potential of genomic techniques to inform management decisions regarding conservation, restoration or assisted migration (Allendorf *et al.* 2010; Funk *et al.* 2012). Maximizing survival in plantings for

ecological restoration or climate change mitigation relies in part on choosing the most suitable germplasm. Species and populations that are more plastic are likely to adjust to climatic changes *in situ*, while in those showing strong local adaptation, assisted migration is likely to be beneficial (Aitken & Whitlock 2013; Hof *et al.* 2011; Lunt *et al.* 2013; Richter *et al.* 2012). Knowledge of the patterns of local adaptation and plasticity in a species is, therefore, critical for making sound seed sourcing decisions for revegetation and restoration. Information underpinning such decisions should come from multiple sources, including ecological similarity, phenotypic similarity (including plasticity), genome-wide similarity at neutral markers and genetic similarity at adaptive loci (Allendorf *et al.* 2010).

The best method of detecting local genetic adaptation within a species is to establish multiple common garden experiments across a range of environments (Aitken & Whitlock 2013; Kremer *et al.* 2013; Matyas 1996), but this requires significant resources for establishment and analysis, particularly when involving long-lived organisms such as trees (Neale & Kremer 2011). One motivation for this research was to determine whether genome-wide scanning technology, combined with outlier marker detection, could be used to bypass time-consuming and expensive common garden experiments. The study demonstrated the feasibility of a relatively simple approach in which plants are sampled from across the full range of ecological variants, genotyped using a high-throughput genome-wide approach, and assessed for population structure and outlier loci. If outlying markers are detected, linear regression can be used to screen them for associations with environmental variables, such as climate or soil. CAP can be used to identify the vector describing the change in the environmental variables most strongly correlated with the adaptive

molecular change variation (e.g., the AI) to allow the predictions of contemporary and future adaptive surfaces for species.

Our genome-wide scanning approach provides a means of modelling adaptation and has the potential to match overall adaptation of provenances to sites without identifying particular loci (Alberto *et al.* 2013), nor necessarily knowing the underlying functional traits (e.g. frost or drought resistance, heat tolerance) involved in adaptation. Although we have shown that there are strong population-level associations between outlier loci in *E. tricarpa* and various climatic variables and some functional traits, there are numerous ecophysiological traits that we did not assess, but for which we may, unknowingly, have detected adaptive loci. The genome-wide scanning approach to adaptation detection allows for the identification of a wide range of adaptive loci, without the *a priori* constraints imposed by a targeted candidate-gene survey (Franks & Hoffmann 2012).

Further validation of the methodology will involve confirmation that allele frequencies of putatively adaptive loci are correlated to climate and/or adaptive traits in provenances of *E. tricarpa* other than those that have been tested here.

Provenances of particular interest include those that grow in zones where the rainfall does not align with the general west-east trend of the rainfall gradient. Our approach will also be useful for identifying changes in population-level adaptive allele frequencies over time, both within and between generations (e.g., see (Jump *et al.* 2006), as they – or the Aridity Indices derived from them - could provide a means of monitoring adaptation to environmental change, as well as hybridization (e.g., where differentiated germplasm has been planted) and genetic drift.

Many studies of adaptation focus on individual mutations in selected candidate genes. However, focusing on particular genes could result in other important sources of adaptive variation being overlooked (Allendorf *et al.* 2010). The genome-wide distribution of adaptive markers found in *E. tricarpha* is concordant with theoretical predictions (Le Corre & Kremer 2003) and other experimental results (e.g., Eckert *et al.* 2010) that have suggested that adaptive genetic variation is a genome-wide phenomenon (Neale & Kremer 2011). Adaptation is generally thought to involve multiple loci of small effect that are spread across the genome, combinations of which may produce an 'adaptive phenotype' (Berg & Coop 2013; Pritchard & Di Rienzo 2010). 'Hot spots' of adaptive differentiation, such as the cluster of six adaptive markers identified on chromosome eight, may represent co-adapted gene complexes that are under divergent selection (Nosil *et al.* 2009). Alternatively, they may be a consequence of genetic hitch-hiking (Barton 2000; Schlotterer 2003), where selectively neutral markers are linked to a single gene that is under positive selection.

In summary, we have shown that a genome-wide scan and outlier analysis can detect evidence of adaptation and its environmental drivers in wild populations of non-model species. This approach will be useful for determining climatic adaptation in species identified for environmental plantings so that selection of seed sources accounting for climate adaptation ('climate-adjusted provenancing') confers greater resilience to future climates.

Acknowledgements

This work was funded by a grant from the National Climate Change Adaptation Research Facility (TB11 03). We thank: David Bush (CSIRO), Martyn Lavery (Arianda Seed), Shane Lavell and Paul Turnbull (DSE Victoria), Craig Macfarlane (CSIRO), Gay Blake and Didier Alanoix (DPaW/CSIRO) for assistance with field work; Paul Tilyard (UTAS) for technical assistance; the project's end-user advisory group, David Freudenberger, Gary Howling, Neil Riches and Richard Mazanec; Cristina Ramalho (DPAW) for production of Figure 6; Sue Baker and Jules Freeman (UTAS) for useful discussions; Tom Baker (UMelb) and Des Stackpole (Australian Low Rainfall Tree Improvement Group) for establishing field trials.

References

- Aitken SN, Whitlock M (2013) Assisted gene flow to facilitate local adaptation to climate change. *Annual Review of Ecology, Evolution and Systematics* **44**, 13.11-13.22.
- Alberto FJ, Aitken SN, Alia 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.
- Anderson MJ, Gorley RN, Clarke KR (2008) *Permanova+ for Primer: guide to software and statistical methods*. PRIMER-E, Plymouth.
- Anderson MJ, Robinson J (2003) Generalized discriminant analysis based on distances. *Australian and New Zealand Journal of Statistics* **45**, 301-318.
- Anderson MJ, Willis TJ (2003) Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* **84**, 511-525.
- Barton NH (2000) Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **355**, 1553-1562.
- Benjamini Y, Yekateuli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29**, 1165-1188.
- Berg JB, Coop G (2013) The population genetic signature of polygenic local adaptation. University of California, Davis.
- Booth TH (2013) Eucalypt plantations and climate change. *Forest Ecology and Management* **301**, 28-34.

- 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 LM, Lowe A, Coates DJ, *et al.* (2008) Seed supply for broadscale restoration: maximizing evolutionary potential. *Evolutionary Applications* **1**, 587-597.
- Byrne M (2008) Phylogeny, diversity and evolution of eucalypts. In: *Plant Genome Biodiversity and Evolution. Part E. Phanerogams - Angiosperms* (eds. Sharma A, Sharma A), pp. pp. 303-346. Science Publishers, Berlin.
- Byrne M, Stone L, Millar MA (2011) Assessing genetic risk in revegetation. *Journal of Applied Ecology* **48**, 1365-1373.
- Chen IC, Hill JK, Ohlemuller R, Roy DB, Thomas CD (2011) Rapid range shifts of species associated with high levels of climate warming. *Science* **333**, 1024-1026.
- Clarke KR, Gorley RN (2006) *Primer v6: User Manual/Tutorial* PRIMER-E, Plymouth.
- Davey JW, Blaxter ML (2010) RADSeq: next generation population genetics. *Brief. Funct. Genomics* **9**, 416-423.
- De Mita S, Thuillet AC, Gay L, *et al.* (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular ecology* **22**, 1383-1399.
- Eckert AJ, Bower AD, Gonzalez-Martinez SC, *et al.* (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular ecology* **19**, 3789-3805.
- Eckert AJ, Wegrzyn JL, Pande B, *et al.* (2009) Multilocus patterns of nucleotide diversity and divergence reveal positive selection at candidate genes related to cold hardiness in coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*). *Genetics* **183**, 289-298.
- Foll M, Gaggiotti O (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977-993.
- Franks SJ, Hoffmann AA (2012) Genetics of climate change adaptation. *Annual Review of Genetics* **46**, 185-208.
- Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in ecology & evolution* **27**, 489-496.
- Ganguly AR, Steinhäuser K, Erickson DJ, *et al.* (2009) Higher trends but larger uncertainty and geographic variability in 21st century temperature and heat waves. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 15555-15559.
- Hancock AM, Brachi B, Faure N, *et al.* (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**, 83-86.
- Hedrick PW (2000) *Genetics of Populations*, 2nd edn. Jones and Bartlett, Boston.
- Hereford J (2009) A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist* **173**, 579-588.
- Hof C, Levinsky I, Araújo MB, Rahbek C (2011) Rethinking species' ability to cope with rapid climate change. *Global Change Biology* **17**, 2987-2990.
- Huff DR, Peakall R, Smouse PE (1993) RAPD variation within and among natural populations of outcrossing buffalo grass, *Buchloe dactyloides* (Nutt.) Engelm. . *Theoretical and Applied Genetics* **86**, 927-934.

- Jump AS, Hunt JM, Martinez-Izquierdo JA, Penuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Molecular ecology* **15**, 3469-3480.
- Kelly M, Mercer D (2005) Australia's box-ironbark forests and woodlands: saving the fragments of a threatened ecosystem. *Australian Geographer* **36**, 19-37.
- Kremer A, Potts BM, Delzon S (2013) Genetic divergence in forest trees: understanding the consequences of climate change. *Functional Ecology*.
- Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics* **164**, 1205-1219.
- Lunt ID, Byrne M, Hellmann JJ, *et al.* (2013) Using assisted colonisation to conserve biodiversity and restore ecosystem function under climate change. *Biological Conservation* **157**, 172-177.
- Matyas C (1996) Climatic adaptation of trees: Rediscovering provenance tests. *Euphytica* **92**, 45-54.
- McLean EH, Steane DA, Prober SM, *et al.* (2014) Plasticity of functional traits varies clinally along a rainfall gradient in *Eucalyptus tricarpa*. *Plant, Cell and Environment* **accepted for publication 30/11/13**.
- Mendenhall W, Scheaffer RL, Wackerley DD (1986) *Mathematical Statistics with Applications*, 3rd edn. Duxbury Press, California.
- Narum SR, Hess JE (2011) Comparison of F-ST outlier tests for SNP loci under selection. *Molecular ecology resources* **11**, 184-194.
- Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* **12**, 111-122.
- Nei M (1972) Genetic distance between populations. *American Naturalist* **106**, 283-292.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular ecology* **18**, 375-402.
- Parisod C, Christin PA (2008) Genome-wide association to fine-scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytologist* **178**, 436-447.
- Parmesan C (2006) Ecological and evolutionary responses to recent climate change. In: *Annual Review of Ecology Evolution and Systematics*, pp. 637-669.
- Peakall R, Smouse P (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.
- Perez-Figueroa A, Garcia-Pereira MJ, Saura M, Rolan-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology* **23**, 2267-2276.
- Peters GP, Andrew RM, Boden T, *et al.* (2013) COMMENTARY: The challenge to keep global warming below 2 degrees C. *Nature Climate Change* **3**, 4-6.
- Petroli CD, Sansaloni CP, Carling J, *et al.* (2012) Genomic characterization of DArT markers based on high-density linkage analysis and physical mapping to the *Eucalyptus* genome. *PloS one* **7**, e44684. doi:44610.41371.
- Pritchard JK, Di Rienzo A (2010) Adaptation - not by sweeps alone. *Nature Reviews Genetics* **11**, 665-667.
- Richter S, Kipfer T, Wohlgemuth T, *et al.* (2012) Phenotypic plasticity facilitates resistance to climate change in a highly variable environment. *Oecologia* **169**, 269-279.

- Sansaloni CP, Petrolis CD, Jaccoud D, *et al.* (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**, P54.
- Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**, 807-820.
- Schabenberger O, Gotway CA (2004) *Statistical methods for spatial data analysis*. Chapman and Hall/CRC, London, U.K.
- Schlotterer C (2003) Hitchhiking mapping - functional genomics from the population genetics perspective. *Trends in Genetics* **19**, 32-38.
- Skabo S, Vaillancourt RE, Potts BM (1998) Fine-scale genetic structure of *Eucalyptus globulus* ssp. *globulus* forest revealed by RAPDs. *Australian Journal of Botany* **46**, 583-594.
- Smith FP, Prober SM, House APN, McIntyre S (2013) Maximizing retention of native biodiversity in Australian agricultural landscapes-The 10:20:40:30 guidelines. *Agriculture Ecosystems & Environment* **166**, 35-45.
- Stackpole DJ, Harwood C (2001) *Eucalyptus tricarpa*: genetic improvement strategy in southern Australia. In: *Rural Industries Research and development Corporation Publication* pp. 94-113, Canberra.
- Stapley J, Reger J, Feulner PGD, *et al.* (2010) Adaptation genomics: the next generation. *Trends in Ecology and Evolution* **25**, 705-712.
- Strasburg JL, Sherman NA, Wright KM, *et al.* (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the Royal Society B-Biological Sciences* **367**, 364-373.
- Valladares F, Sanchez-Gomez D, Zavala MA (2006) Quantitative estimation of phenotypic plasticity: bridging the gap between the evolutionary concept and its ecological applications. *Journal of Ecology* **94**, 1103-1116.
- Vallejo VR, Smanis A, Chirino E, *et al.* (2012) Perspectives in dryland restoration: approaches for climate change adaptation. *New Forests* **43**, 561-579.
- Vilas A, Perez-Figueroa A, Caballero A (2012) A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers. *Journal of Evolutionary Biology* **25**, 1364-1376.
- Xu T, Hutchinson M (2011) *ANUCLIM version 6.1 user guide*. The Australian National University, Fenner School of Environment and Society, Canberra.

Data Accessibility

DARTseq data, climate data and functional trait data may be accessed via Dryad:

doi:10.5061/dryad.qq20s

Author Contributions

All authors contributed to the design of the project. DS and BP analysed molecular data and wrote manuscript with assistance from all other authors. EM conducted field work and produced functional trait data. All authors read and approved final manuscript.

Table 1. Location and mean annual climatic conditions (1925-1995; Atlas of Living Australia) for the nine provenances and the two common garden sites of *Eucalyptus tricarpa*. Provenance numbering corresponds to Figure 1 and is in order of mean annual precipitation (MAP). The provenances are numbered in order of increasing MAP.

Location (code)	Latitude (°N)	Longitude (°E)	MAP (mm)	Daily max. temp. (°C)	Daily min. temp. (°C)	Annual evaporation (mm)	Annual solar radiation (MJ m ⁻²)	CAP1-based Aridity Index
Provenances								
1. Tarnagulla	-36.76	143.85	460	20.4	7.4	1335	199	3.56
2. Mt Bealiba	-36.81	143.65	511	19.9	7.2	1306	197	2.04
3. Craigie	-37.08	143.77	543	19.7	7.1	1249	195	2.08
4. Heathcote	-36.98	144.75	621	19.3	7.3	1297	193	1.07
5. Heyfield	-37.94	146.73	683	19.1	7.6	1177	184	-1.29
6. Christmas Hills	-37.69	145.31	787	18.6	8.1	1194	177	-1.79
7. Mt Nowa Nowa	-37.7	148.11	860	19.2	7.5	1241	184	-2.38
8. Tuckerbox	-37.63	148.24	879	18.7	7	1217	183	-2.72
9. Martins Creek	-37.47	148.58	1020	18.4	6.4	1241	180	-3.70
Common gardens								

Huntly	-36.63	144.31	472	20.6	7.6	1395	199	3.94
Lake Tyers	-37.82	148.10	840	19.4	8.4	1261	185	-0.82

Table 2. Proportions of significant associations ($P < 0.001$) found in marker-trait association tests comparing results from 94 ‘outlier’ markers to 3590 ‘neutral’ markers in *Eucalyptus tricarpa*. Provenance-level allele frequencies of DArTseq markers were regressed against climatic variables and soil traits at each site, as well as functional traits measured in the wild and in two field trials. The number of marker-trait associations reflects the number of tests done. Chi-squared values in bold are those for which expected values of significant associations in the outlier data were adjusted upwards to 5 (from values < 5). This adjustment made the tests more conservative and therefore did not bias the test in favour of high significance.

TRAITS/VARIABLES (number)	Marker Status	No. Marker-trait Associations		Percentage $P < 0.001$	Enrichment $P < 0.001$	P (ChiSq)
		TOTAL	$P < 0.001$			
CAP1 (1)	Neutral	3590	1	0.03		
	Outlier	94	27	28.7	1031.2	< 0.001
Aridity Index (1)	Neutral	3590	131	3.6		
	Outlier	94	39	41.5	11.4	< 0.001
Climatic variables (35) ^A	Neutral	125650	2894	2.3		
	Outlier	3290	625	19.0	8.2	< 0.001
Rainfall (8) ^A	Neutral	28720	773	2.7		
	Outlier	752	151	20.1	7.46	< 0.001
Temperature (11) ^A	Neutral	39490	713	1.8		
	Outlier	1034	146	14.1	7.82	< 0.001
Moisture indices (8) ^A	Neutral	28720	787	2.7		
	Outlier	752	190	25.3	9.22	< 0.001

Radiation (8) ^A	Neutra					
	I	28720	621	2.2		
	Outlier	752	138	18.4	8.49	<0.001
Soil variables (15) ^B	Neutra					
	I	53850	54	0.1		
	Outlier	1410	1	0.1	0.7	0.0736
Wild traits (14) ^C	Neutra					
	I	50260	137	0.3		
	Outlier	1316	5	0.4	1.4	0.9689
Trial traits (28)	Neutra					
	I	100520	432	0.4		
	Outlier	2632	39	1.5	3.4	<0.001
Huntly (dry) (14) ^D	Neutra					
	I	50260	151	0.3		
	Outlier	1316	15	1.1	3.79	<0.001
Lake Tyers (wet) (14) ^D	Neutra					
	I	50260	281	0.6		
	Outlier	1316	24	1.8	3.26	<0.001
Plasticity traits (7) ^E	Neutra					
	I	25130	181	0.7		
	Outlier	658	20	3.0	4.2	<0.001

^A See Supplementary Material 1

^B Soil variables: texture, soil particle size (clay, coarse sand, sand, fine sand, silt), chemistry (NH₄, NO₃, P, K, S, organic C, conductivity, pH in CaCl₂, pH in water) (McLean *et al.*, 2014).

^C Functional traits measured in wild populations: leaf size, leaf thickness, leaf density, specific leaf area, circumference of main stem, cellulose ¹³C, leaf ¹³C, leaf ¹⁵N, C mass, C:N ratio, Leaf N area, Leaf N mass, number of stems, total stem cross sectional area (at 1.3 m).

^D Functional traits measured in common garden trials: leaf size, leaf thickness, leaf density, specific leaf area, circumference of main stem, cellulose ¹³C, leaf ¹³C, leaf ¹⁵N, C mass, C:N ratio, Leaf N area, Leaf N mass, total cross sectional area, height of main stem.

^E Plasticity (Relative Trait Range; (Valladares *et al.* 2006) of: leaf size, leaf thickness, cellulose ¹³C, , leaf density, height of main stem, total stem cross sectional area (at 1.3 m), Specific Leaf Area (SLA).

Table 3. Analysis of Covariance (ANCOVA) fitting *Eucalyptus tricarpa* provenance level traits to site (i.e. common garden), CAP1 (i.e. adaptive molecular index) and CAP1 X site interaction. Sig – significance level - * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, ns - $P > 0.05$ (not significant). SLA – specific leaf area; CSA – cross-sectional area. Leaf N_{mass} – leaf nitrogen content per unit mass; Leaf N_{area} – leaf nitrogen content per unit area.

Trait	Site		CAP1		CAP1 X Site	
	$F_{1,14}$	Sig	$F_{1,14}$	Sig	$F_{1,14}$	Sig
Cellulose ^{13}C	364.9	***	3.0	ns	1.3	ns
Density	86.3	***	2.4	ns	1.2	ns
Leaf thickness	45.5	***	47.4	***	12.7	**
SLA	126.6	***	29.8	***	11.3	**
Leaf size	23.7	***	36.4	***	8.7	*
Leaf N_{mass}	26.6	***	0.0	ns	1.4	ns
Leaf N_{area}	8.7	*	10.8	**	2.0	ns
CSA	0.4	ns	2.6	ns	10.2	**
Height	18.5	***	1.9	ns	9.8	**

Figure Captions

Figure 1. Locations of the nine study populations of *Eucalyptus tricarpa* (numbers 1-9) and the two common garden provenance trials at Huntly (open star) and Lake Tyers (solid star) in southeastern Australia. Population numbering corresponds to that in Table 1.

Figure 2. Ordination of *Eucalyptus tricarpa* DArTseq data. (A) Full DArTseq data set (6,544 DArTseq markers); (B) Outlier data set (94 outlier markers); (C) Canonical analysis of Principal Coordinates (CAP) showing the directionality of 15 climatic variables (from ANUCLIM), of which 10 are labelled (see Supporting Information 1). Note that the axes of (C) are not scaled by their respective eigenvalues (Anderson et al. 2008).

Figure 3. Association between mean leaf size and the adaptive genetic index (CAP1) for each provenance of *Eucalyptus tricarpa* grown in common gardens at the dry site, Huntly (black diamonds, black regression line), and the wet site, Lake Tyers (grey squares, grey regression line).

Figure 4. Association of performance (mean total cross sectional area of stems) of *Eucalyptus tricarpa* provenance material in the field trials with the CAP-derived Aridity Index (AI) at the site of origin of each provenance. Provenances from wetter areas (more negative AI) grew better in the wetter field trial (Lake Tyers, grey squares and regression line), while provenances from drier areas (more positive AI) grew better in the drier field trial (Huntly, black diamonds and regression line). The AI at each trial site is indicated by an arrow (Lake Tyers, grey arrow; Huntly, black arrow).

Figure 5. Locations of DArTseq markers (grey crosses) from *Eucalyptus tricarpa* along the 11 major linkage groups of *E. grandis*. Of the 94 outlying markers, 48 could be mapped to the eleven main linkage groups of the *E. grandis* genome. The 43 outlying markers that were strongly associated with climate are shown as black squares; other outliers are shown as white triangles.

Figure 6. Predicted surface of climate adaptation across the entire range of *Eucalyptus tricarpa*. The map is derived from AI values calculated for all reliable records of *E. tricarpa* in the Atlas of Living Australia. Provenances used in this study are represented by numbers in white circles. Common garden trials are indicated by squares (HL – Huntly (dry) trial; LT – Lake Tyers (wet) trial). Inset shows the distribution of *E. tricarpa* in relation to the whole of Australia.

Supporting Information 1. ANUCLIM climatic variables and abbreviations.

Supporting Information 2. Correlations between ANUCLIM climatic variables at sampling sites of *Eucalyptus tricarpa* provenances.

Supporting Information 3. Linear regression of soil variables vs CAP1 in wild populations of *Eucalyptus tricarpa*.

Supporting Information 4. F_{st} distribution and outlier selection.

Supporting Information 5. ANOVA of functional traits in *Eucalyptus tricarpa* in wild populations and in two field trials.

Supporting Information 6. Linear regression of climate variables vs CAP1 in wild populations of *Eucalyptus tricarpa*.

Supporting Information 7. Linear regression of CAP1 vs functional traits in wild populations of *Eucalyptus tricarpa*.

Supporting Information 8. Linear regression of CAP1 vs functional traits in two common garden trials of *Eucalyptus tricarpa*.

Supporting Information 9. Outlier DArTseq markers from *Eucalyptus tricarpa*, associations with climate, CAP1, Aridity Index and functional traits in the two provenance trials; number of locations that each marker could be placed on the *E. grandis* reference genome; the scaffold number (scaffolds 1-11 correspond to the 11 chromosomes of *E. grandis*) to which the marker was mapped (with best score), position on the linkage group or scaffold and the first 60 bp of the DNA sequences of the DArTseq marker.





