Genomic evidence of introgression and adaptation in a model subtropical tree species, *Eucalyptus grandis*

Marja Mirjam Mostert-O'Neill

Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria 0028, South Africa

marja.oneill@up.ac.za

Sharon Melissa Reynolds

Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria 0028, South Africa

melissa.reynolds@up.ac.za

Juan Jose Acosta

Camcore, Department of Forestry and Environmental Resources, North Carolina State University, PO Box 7626, Raleigh, NC, NC 27695, USA

jjacosta@ncsu.edu

David John Lee

Forest Industries Research Centre, University of the Sunshine Coast, Locked Bag 4, Maroochydore DC, QLD 4558, Australia

dlee@usc.edu.au

Justin O. Borevitz

Research School of Biology and Centre for Biodiversity Analysis, ARC Centre of Excellence in Plant Energy Biology, Australian National University, Canberra, ACT 0200, Australia

justin.borevitz@anu.edu.au

Alexander Andrew Myburg

Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria 0028, South Africa

zander.myburg@fabi.up.ac.za

CORRESPONDING AUTHOR:

Prof. Alexander A. Myburg

Email: zander.myburg@fabi.up.ac.za

Telephone: +27 (0)12 420 4945

Fax: +27 (0)12 420-3947

ABSTRACT

The genetic consequences of adaptation to changing environments can be deciphered using population genomics, which may help predict species' responses to global climate change. Towards this, we used genome-wide SNP marker analysis to determine population structure and patterns of genetic differentiation in terms of neutral and adaptive genetic variation in the natural range of *Eucalyptus grandis*, a widely cultivated subtropical and temperate species, serving as genomic reference for the genus. We analysed introgression patterns at subchromosomal resolution using a modified ancestry mapping approach and identified provenances with extensive interspecific introgression in response to increased aridity. Furthermore, we describe potentially adaptive genetic variation as explained by environment-associated SNP markers, which also led to the discovery of what is likely a large structural variant. Finally, we show that genes linked to these markers are enriched for biotic and abiotic stress responses.

KEY WORDS: introgression, ancestry mapping, population genomics, environmental association

1 INTRODUCTION

Hidden within extant genomes are the consequences of natural selection in response to millennia of abiotic and biotic stress. With population genomics, we can decipher genomic signatures of selection and combine them with environmental data to predict how a species will likely respond to climate change and *ex situ* environments. This approach has gained interest with recent advances in high-resolution environmental and genome-wide genotyping technologies, particularly for non-model organisms (Bragg *et al.* 2015; Rellstab *et al.* 2015), to aid in conservation, assisted migration and ecosystem restoration efforts (Ahrens *et al.* 2019; Gugger *et al.* 2018; Jordan *et al.* 2016; Landguth *et al.* 2017; Shryock *et al.* 2015; Supple *et al.* 2018). Because population genomics can reveal hidden adaptive genetic variation, it allows for management of genetic resources beyond simple species demography (Luikart *et al.* 2003).

Eucalyptus, frequently used in forest and woodlands restoration (Bennett 2016; Jordan *et al.* 2016; Pryor & Johnson 1981; Supple *et al.* 2018), is globally the most widely planted hardwood genus (Myburg *et al.* 2007) covering over 20 million ha worldwide (Iglesias-Trabado & Wilstermann 2008). Differentiation among the more than 700 eucalypt species spans the speciation continuum, with extensive evidence of reticulate evolution, hybridization and adaptive introgression (Bayly *et al.* 2013; Jones *et al.* 2016; McKinnon *et al.* 2004; Pollock *et al.* 2013; Rutherford *et al.* 2018). Obscured species delineation can complicate management of genetic resources, which are evaluated at the species level. Population genomics could resolve some species boundaries within this genus. Additionally, it could enrich our understanding of hybrid speciation and the role of adaptive introgression in response to changing environments.

Eucalyptus grandis serves as a model species for the genus. Although the wealth of common garden data (Arnold *et al.* 1996; Burgess 1988; Darrow 1983; Langat & Kariuki 2004; Matheson & Mullin 1987; Skolmen 1986), genomic (Myburg *et al.* 2014; Silva-Junior *et al.* 2015) and transcriptomic

(Mangwanda *et al.* 2015; Mizrachi *et al.* 2010; Oates *et al.* 2015; Vining *et al.* 2015) resources available for *E. grandis* stems from its commercial significance (Harwood 2011), it is also an important foundation species along the east coast of Australia. Its natural range encompasses eight Köppen climate classification zones (Köppen 2011; Köppen 1884) including subtropics in Queensland, and temperate areas in New South Wales (http://www.ala.org.au. Accessed 28 November 2019; see **Table S1** in Supplementary Information). Despite its ecological and commercial relevance, an in-depth analysis of potentially valuable adaptive genetic variation in relation to abiotic factors is still lacking. Outstanding questions include: What are the dominant environmental factors affecting the realised niche? What will most likely be the mode of response to niche changes? Do we see evidence of local adaptation as proposed from historic common garden experiments (Arnold *et al.* 1996; Burgess 1988; Darrow 1983; Langat & Kariuki 2004; Matheson & Mullin 1987; Skolmen 1986)? To address these questions, we first need to understand the population structure and genetic diversity in wild *E. grandis*.

With available genomic and environmental data (Belbin & Williams 2016) resources, we have begun to decipher population differentiation and genetic diversity in wild populations of *E. grandis* in terms of demography and potentially adaptive genetic variation. Our hypothesis is that genomic loci associated with environmental factors would be enriched for adaptive genetic variation. Furthermore, following the rational of Steane *et al.* (2014), population stratification derived from the adaptively enriched genetic space would correlate with environmental clines as opposed to the demography of the species. This is because regions of the genome affected by selection for adaptive traits could have different patterns of inheritance compared to selectively neutral regions. We report the use of Latent Factor Mixed Model (LFMM) analyses to identify SNP markers associated with environmental clines to serve as adaptively enriched marker sets for population differentiation analysis. Population stratification is investigated using Principal Component Analysis (PCA) and Sparse Non-negative Matrix Factorization (sNMF). Generalized Dissimilarity Modelling (GDM) was used to refine which environmental factors are most predictive of the population differentiation observed for adaptively

enriched SNP sets. Finally, we present a modified approach for ancestry mapping that does not require prior knowledge of the identity of the introgressing species. This study generates new hypotheses regarding adaptive population differentiation in *E. grandis*, to be tested in common garden trials.

2 MATERIALS & METHODS

2.1 Study population, DNA isolation and genotyping

Seed stocks collected from 193 wild growing trees (Bean 1989; Gunn 2001) were imported from the Australian Tree Seed Centre (Commonwealth Scientific and Industrial Research Organisation; Gunn 2001) and a Queensland Government collection managed by D. Lee (University of the Sunshine Coast, Australia), and germinated in Pretoria, South Africa, under greenhouse conditions. The collection represents most of the latitudinal range, from Windsor Tablelands (16°11'S) in Queensland to Bulahdelah (32°20'S) in New South Wales, with 33 provenances represented (**Table 1** and **Fig. 1**). Since most provenances were represented by fewer than 10 individual mother trees (**Table S1**), multiple half-siblings were selected in each family to capture pollen parent variation (**Appendix S1**). All seedlings analysed were established in an *ex situ* gene conservation park in the KwaZulu-Natal Midlands, South Africa.

Leaves were collected from six-week-old seedlings for DNA isolation using the NucleoSpin DNA extraction kit (Machery-Nagel GmbH & Co. KG, Düren, Germany) and SNP genotypes were generated for 596 seedlings using the EUChip60K SNP chip (Silva-Junior *et al.* 2015) by the genomics service provider, GeneSeek (Neogen, Lincoln, NE, USA). Intensity data were first assessed using GenomeStudio V2011.1 (Illumina Inc., San Diego, CA, USA) to recluster genotypic classes as described by Silva-Junior *et al.* (2015), followed by data filtering to retain high-confidence, informative SNPs (minor allele frequency > 0.02 and genotyped in at least 90% of samples) and samples with less than 10% missing data using the SNP & Variation Suite™ v8.x (SVS8; Golden Helix, Inc., Bozeman, MT). Identity by decent analysis was used to confirm sibship and detect cryptic relatedness among individuals (**Appendix S1** and **Fig. S1**). To avoid over-representation of any

individual family, a balanced subset of 362 samples with no more than two half-siblings per family was compiled for population differentiation and environmental association analyses.

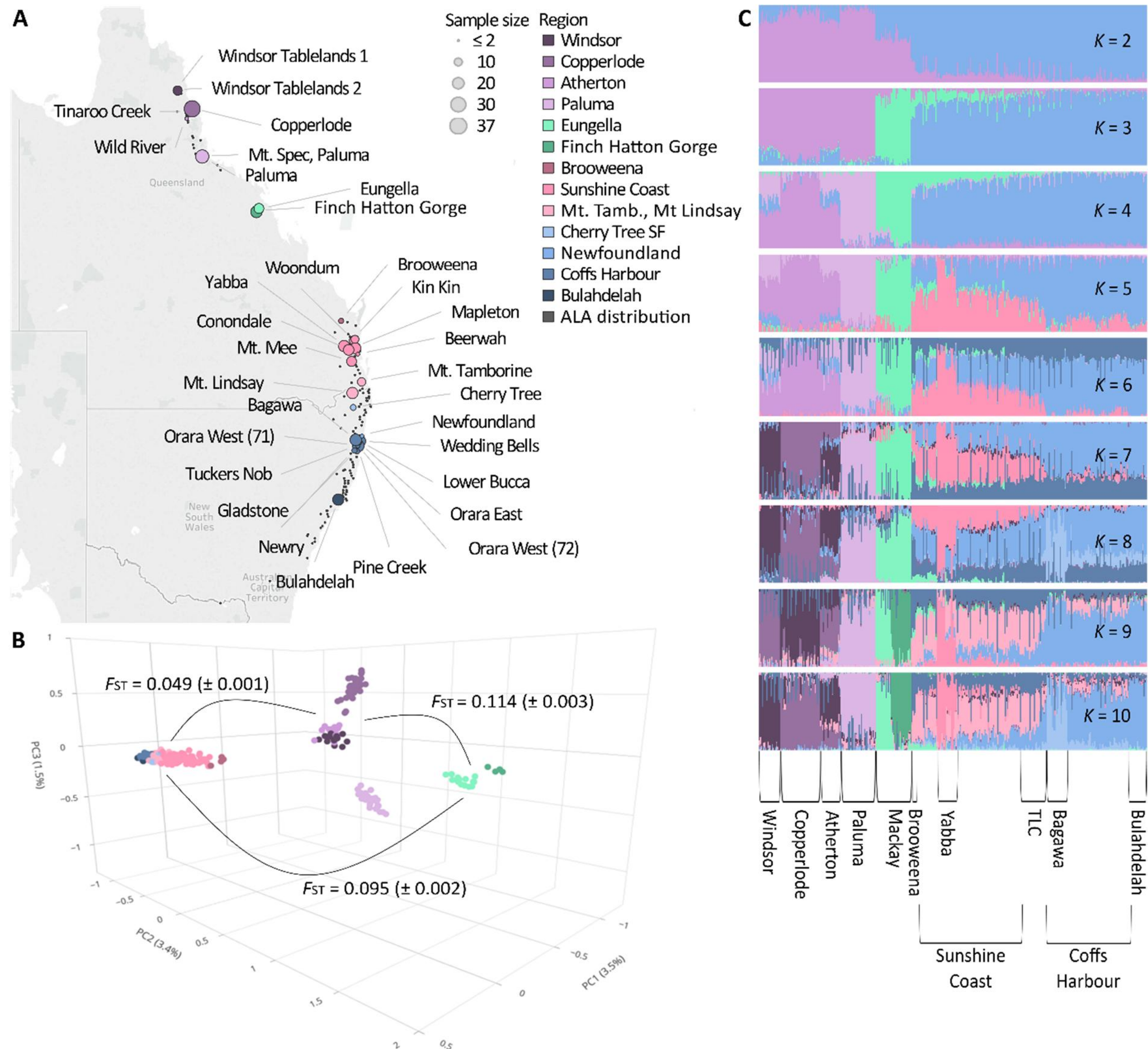


Figure 1 Population structure of *E. grandis*. **A**) Source locations (coloured circles) with circle sizes indicating numbers of samples. Broad geographic regions of the Northern and Mackay provenances are indicated in shades of purple and green, respectively. Southern regions are coloured in shades of pink and blue. The natural distribution according to Atlas of Living Australia (occurrence data download at http://www.ala.org.au. Accessed 28 November 2019) is given in dark grey dots. **B**) Population structure is given based on the first three components of the PCA and F_{ST} estimates are indicated with 95% confidence intervals in parentheses. **C**) sNMF outcome indicating per individual (x-axis) the genomic proportion assigned to each ancestral population (y-axis) for different values of *K*. Geographic regions and key provenances are indicated along the x-axis; Mackay – Eungella and Finch Hatton Gorge, TLC – Mt. Tamborine, Mt. Lindsey and Cherry Tree.

Table 1. Seed stocks and study population

Location/provenance	Region	Subpop.[a]	Latitude	Longitude	Elevation (m)
Windsor Tableland 1	Windsor	North	−16.183	144.983	1,161
Windsor Tableland 2	Windsor	North	−16.200	145.017	1,134
Copperlode	Copperlode	North	−16.967	145.667	368
Tinaroo Creek Road	Atherton	North	−17.083	145.567	925
Tinaroo Dam	Atherton	North	−17.133	145.567	893
Wild River, QLD	Atherton	North	−17.367	145.400	927
Mt. Spec, Paluma	Paluma	North	−18.933	146.117	822
Paluma	Paluma	North	−19.000	146.150	796
Eungella	Eungella	Mackay	−21.075	148.632	294
Finch Hatton Gorge	Finch Hatton Gorge	Mackay	−21.197	148.504	753
Brooweena	Brooweena	South	−25.550	152.267	210
Kin Kin[b]	Sunshine Coast	South	−26.263	152.876	142
Woondum	Sunshine Coast	South	−26.300	152.817	306
Yabba	Sunshine Coast	South	−26.533	152.400	475
Mapleton	Sunshine Coast	South	−26.600	152.867	325
Conondale	Sunshine Coast	South	−26.667	152.600	498
Beerwah	Sunshine Coast	South	−26.850	153.000	20
Mt. Mee	Sunshine Coast	South	−27.133	152.717	368
Mt. Tamborine	Mt. Tamborine & Mt Lindsay	South	−27.917	153.183	396
Mt. Lindsay	Mt. Tamborine & Mt Lindsay	South	−28.350	152.750	352
Cherry Tree	Cherry Tree	South	−28.900	152.817	236
Newfoundland	Newfoundland	South	−29.946	153.162	52
Bagawa	Coffs Harbour	South	−30.117	152.900	299
Wedding Bells	Coffs Harbour	South	−30.167	153.117	95
Lower Bucca	Coffs Harbour	South	−30.178	153.074	108
Orara East	Coffs Harbour	South	−30.225	153.068	195

Location/provenance	Region	Subpop.[a]	Latitude	Longitude	Elevation (m)
Orara West (71)	Coffs Harbour	South	−30.332	152.973	189
Orara West (72)	Coffs Harbour	South	−30.341	153.011	353
Tuckers Nob	Coffs Harbour	South	−30.372	152.951	259
Pine Creek	Coffs Harbour	South	−30.414	153.009	18
Gladstone	Coffs Harbour	South	−30.509	152.838	242
Newry	Coffs Harbour	South	−30.536	152.956	21
Bulahdelah	Bulahdelah	South	−32.338	152.149	119

[a] Subpopulation assignment based on geographic and genetically distinct groupings.

[b] GPS coordinates provided for Kin Kin (−26.200;153.167) appeared improbable, as such the coordinates for Kin Kin the town were used as proxy environment.

2.2 Population structure and differentiation

Population structure was assessed using sNMF (Frichot *et al.* 2014) with the *LEA* R package (Frichot & François 2015). The number of ancestral populations, *K*, tested in each admixture analysis was one to ten with 20 repetitions, and cross-entropy was visually inspected to determine the most plausible value for *K* either represented as a clear "knee" in the plot or the lowest cross-entropy value. Population differentiation was also assessed using PCA (Price *et al.* 2006) using the standard deviation of each marker for normalization in SVS8. Differentiation between subpopulations, regions and provenances (defined in **Table 1**) was quantified in terms of the *F*-statistic, F_{ST}, as described by Weir and Cockerham (1984), with 95% confidence intervals in SVS8. Hierarchical population structure analyses, as recommended in Evanno *et al.* (2005), was done to find within-group structure by repeating sNMF and PCA on subsets of the population defined by the assignment of individuals at the upper-most hierarchical level, i.e. to identify subtle structure within the Northern and Southern subpopulations.

Elevated F_{ST} estimates for Eungella, Finch Hatton Gorge (the two provenances from the Mackay region) and Brooweena, raised the suspicion that introgression might have occurred in these populations. The *Efficient Inference of Local Ancestry* (*EILA*) R package (Yang *et al.* 2013) was used to test ancestral origins of chromosomal segments. In short, the *EILA* method consisted of three steps: First, each SNP was assessed in sampled individuals and assigned a probability score for ancestry to one of the three possible ancestral populations. Ancestral population 1 (anc1) consisted of 168 *E. grandis* individuals (one individual of each of the 193 families, excluding those from Eungella, Finch Hatton Gorge and Brooweena). Ancestral population 2 (anc2) consisted of genetically diverse individuals from the section *Maidenaria* (10 *E. nitens*, 10 *E. viminalis* and 12 *E. dunnii* individuals), and ancestral population 3 (anc3) represented non-*E. grandis* individuals from the section *Latoangulatae* (10 *E. saligna* and 18 *E. urophylla* individuals). SNP data for anc2 and anc3 were obtained from Silva-Junior *et al.* (2015) and a subset of 7736 SNPs with no missing data and MAF >

0.02 across all reference and query samples was used. Species delineation was illustrated by PCA (**Fig. S2**). The second step in the *EILA* process involved identification of breakpoints along the chromosome where ancestry switched from one ancestral population to another. Different values for the breakpoint penalty (λ) were tested (**Appendix S1** and **Fig. S3**). Thirdly, each chromosomal segment was assigned to one of the three ancestral populations based on the cumulative ancestry assignment of all the SNPs in that segment. If the segments were too small (breakpoint penalty too low), there would not be enough SNPs for confident ancestry assignment, while larger breakpoint values resulted in reduced resolution along the chromosomes. For this purpose, we report $\lambda = 15$ and $\lambda = 30$ for suspected introgressed provenances and $\lambda = 30$ for all other provenances. With each provenance tested, the individuals from that provenance were removed from the anc1 reference set and all 596 seedlings genotyped were analysed. Ancestry assignment scores were graphically presented using Tableau Professional Edition (Tableau Software Inc., Seattle, WA, USA).

2.3 Population structure in relation to environmental clines

Environmental Association Analysis (EAA) was conducted to identify loci associated with abiotic environmental clines, using LFMM analysis, as implemented in the *LEA* R package (Frichot & François 2015). Environmental predictor variables were obtained from the Atlas of Living Australia (Belbin & Williams 2016) environmental data layers (http://www.ala.org.au. Accessed 28 November 2019; **Fig. S4** and **Table S2**). Environmental factors were selected to include variables reported for source locations and/or trial sites in historic provenance and growth trials of *E. grandis* (Arnold *et al.* 1996; Burgess 1988; Skolmen 1986; Van den Berg *et al.* 2016). These included five climate variables: cold period minimum temperature (Bio06), warm period maximum temperature (Bio05), dry period average precipitation (Bio14), wet period average precipitation (Bio13), and mean annual aridity index. In addition to the geographic variables elevation and distance to coast, four edaphic variables were included namely soil depth, soil pedality (grade between one and five, predictive of water retention ability; McKenzie *et al.* 2000), pre-European soil nitrogen levels and pre-European soil phosphorus levels. First, PCA was performed to reduce dimensionality among correlated

environmental variables for the entire species range and two subpopulations, respectively (**Appendix S1**). Principal Components (PCs) were retained for downstream analysis if eigenvalues ≥ 1, and environmental variables showed significant correlation (loadings ≥ 0.7) with the PC to enable meaningful interpretation. LFMM analyses were then used to find loci associated with the PCs and with individual environmental variables. Due to the potential presence of interspecific introgression in Eungella and Finch Hatton Gorge provenances (Mackay region), differences in correlations among environmental variables in the Northern (latitudes -16.183 to -19.000) and Southern (latitudes -25.550 to -32.338) subpopulations (**Table S3**), and distinct gene flow patterns in the two subpopulations, we performed EAA on each subpopulation separately. For LFMM the K value with the minimum cross-entropy in the sNMF analysis was used to correct for populations structure. This was $K = 4$ in the Northern subpopulation EAA and $K = 5$ in the Southern subpopulation EAA (**Fig. S7**). Lastly, population differentiation, calculated as F_{ST} estimates, derived from the full genetic space was compared with that derived from the environment-associated genetic space. Missing genotypic data can result in increased false positives (Frichot & François 2015), therefore, EAA was performed using SNPs with less than 10% missing data in each of the subpopulations and population differentiation analysis, described below, was done using environment-associated SNPs with no missing data.

Correlations between genetic differentiation, geographic distance and environmental clines were tested using GDM. This makes use of F_{ST} population differentiation estimates, calculated for the full genetic space or environment-associated genetic spaces, as response variable. Although small sample sizes can affect F_{ST} estimates, the large number of markers used in our analyses helped to circumvent this (Willing *et al.* 2012). Still, provenances represented by single families were excluded from the models. First, models were generated using the *gdm* R package (Ferrier *et al.* 2007; Manion *et al.* 2018) for the entire balanced population (model A), and the population excluding introgressed provenances, Eungella, Finch Hatton Gorge and Brooweena (model B). Models were also generated and compared for the Northern and Southern subpopulations using population differentiation as given by the total genetic space and environment-associated SNP sets. For each model generated, three *I*-spline basis

functions (Ferrier *et al.* 2007; Ramsay 1988) were used to fit the relationship of each predictor variable, environmental factor or geographic distance, with the response variable, genetic distance as given by F_{ST} estimate. Plots of the *I*-spline function with the maximum value for each predictor variable were used to determine the relative importance of each predictor variable, given by the maximum value of the plot, and the rate of change in the response variable along the environmental gradient of the predictor variable as given by the slope of the plot (Fitzpatrick *et al.* 2013). Permutation tests were done to quantify the importance of each predictor variable calculated as the change in deviance explained by the full model compared to that explained by the model where the variable was permuted. In addition, significance of each predictor variable was calculated as bootstrapped *P*-values.

2.4 Identification of chromosomal rearrangements

The population structure analysis by PCA using SNPs associated with North-PC1 separated all provenances into three classes along the y-axis (**Fig. 3C**). A similar pattern was observed by Faria *et al.* (2019) while investigating chromosomal rearrangements. This raised the questions: is there evidence of structural variants in the North PC1-associated SNP data set? Lack of recombination between loci within chromosomal rearrangements, due to lack of sequence homology, can manifest as genomic regions with high linkage disequilibrium (LD). As such, the North PC1-associated SNPs were interrogated to identify genomic regions of high LD, measured as the squared correlation (R^2) between allelic values at two loci using SVS8. Subsequently, SNPs on chromosome 11, including a large region where $R^2 > 0.8$, were used for population structure analysis to corroborate the putative presence of a structural variant.

2.5 Functional dissection of environment-associated genomic regions

To identify enriched biological processes in the environment-associated gene space compared to the total SNP-captured gene space, genes overlapping a 4 kb window around SNPs (2 kb up- and

downstream of each locus) were extracted using *bedtools* (Quinlan & Hall 2010). This was based on previously determined genome-wide LD decay estimates for *E. grandis* (Silva-Junior & Grattapaglia 2015). Gene annotation and functional enrichment were performed as described by Pinard *et al.* (2019), with the exception that functional enrichment was tested for each environment-associated gene set against the total SNP-captured gene set instead of all annotated genes in the reference genome. Enrichment was also tested for a combined set of genes in linkage with SNPs that had association with at least one of the environmental clines tested in each of the subpopulations ('North all environment-associated' and 'South all environment-associated'). The enrichment test was also repeated for genes in LD with a random selection of 513 informative SNP loci ('Random') and SNPs that did not show association with any of the environmental variables in the Northern subpopulation ('North non-associated').

RESULTS

3.1 Population structure, differentiation and demography

Of the 64 639 SNP markers assayed across 596 seedlings representing 193 families, 25 099 were informative in at least 90% of the samples with minor allele frequency (MAF) > 0.02. Three samples were excluded with more than 10% missing data and 15 samples did not cluster with samples from the same provenance or region in hierarchical PCA. These PCA outliers were excluded from population structure and environmental association analyses, but retained for ancestry mapping. The 578 samples were reduced to 362 so that each family was represented by no more than two half-siblings. Of the 193 families, 24 were represented by only one seedling (**Fig. S1**).

Based on PCA and sNMF, $K = 3$ was the best supported number of subpopulations in the first tier population structure analysis (**Fig. 1**). Population differentiation estimates by F_{ST} (**Fig. S5**) supported clustering observed in the PCA with the highest level of differentiation observed for the Mackay subpopulation (**Fig. 1B**). The Northern subpopulation, consisting of 109 samples from eight source

locations, appeared to be fragmented into four broad regions (K = 4; **Fig. 1C, Fig. S6** and **Fig. S7**). In contrast, the Southern subpopulation structure was less fragmented as illustrated by a lack of distinct clustering in the PCA and sNMF, and lower F_{ST} estimates for the seven geographic regions represented by 220 seedlings (**Fig. 1C**, **Fig. S6** and **Fig. S7**). Only one provenance, Brooweena, appeared clearly distinct from the rest of the subpopulation (**Fig. S6B**).

3.2 Evidence of introgression in three provenances

To test the hypothesis that the high F_{ST} estimates observed for Eungella, Finch Hatton Gorge and Brooweena were due to interspecific introgression, *EILA* was used to test for *E. grandis* and non-*E. grandis* ancestry of chromosomal segments in each seedling. Different introgression patterns were observed in all individuals from the three provenances presumed to be introgressed (**Fig. 2** and **Fig. S8-S10**). Chromosomes 7 and 8 had the least non-*E. grandis* assignment. The 15 samples excluded from population structure and association analyses as they did not group with the expected provenance, had ancestry profiles indicative of more recent interspecific hybridization, e.g. Eungella_20 (**Fig. 2** and **Fig. S8**).

3.3 Population structure in relation to environmental clines

3.3.1 Generalized dissimilarity modelling across the species range

The effect of geographic distance and environmental variables on population differentiation across the species range was modelled using GDM. Model A, which included introgressed provenances, explained 71.4% of deviance. Exclusion of the introgressed provenances (model B) increased the explained deviance to 87.6% (**Table S4**). Geographic distance between provenances was the only variable that explained significant variation in genetic distance in either model (**Fig. S11**).

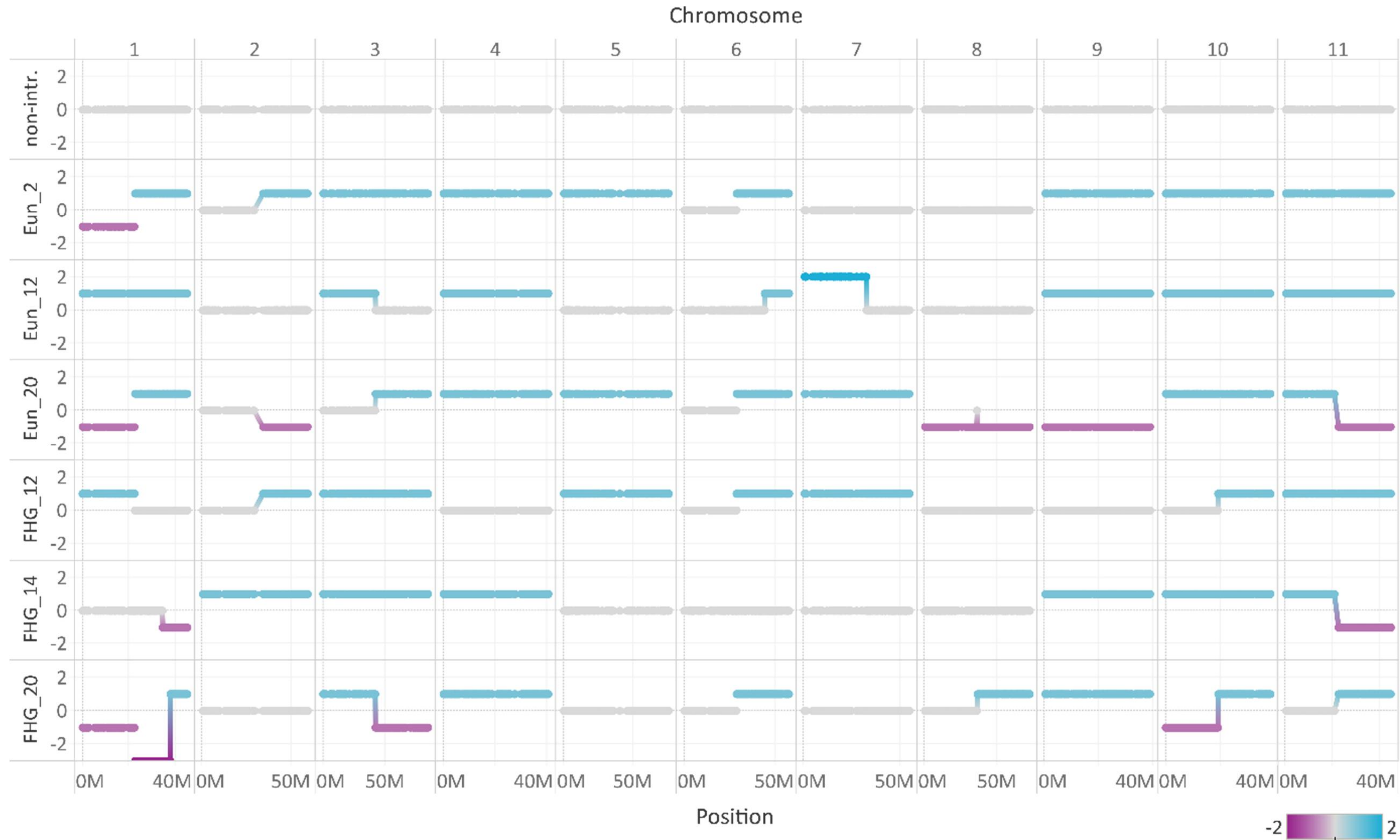
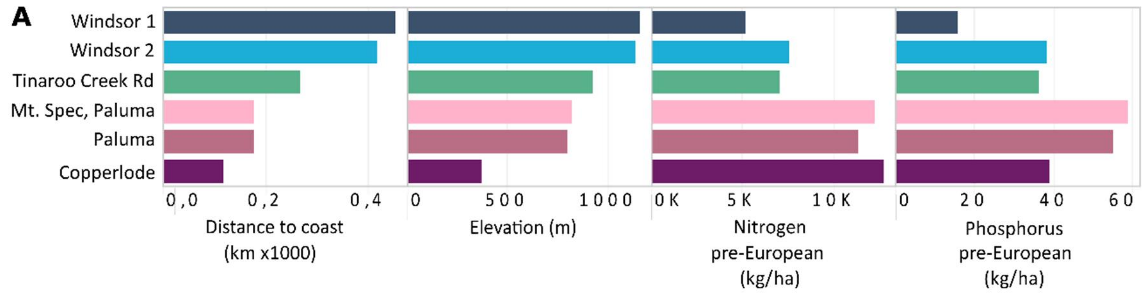


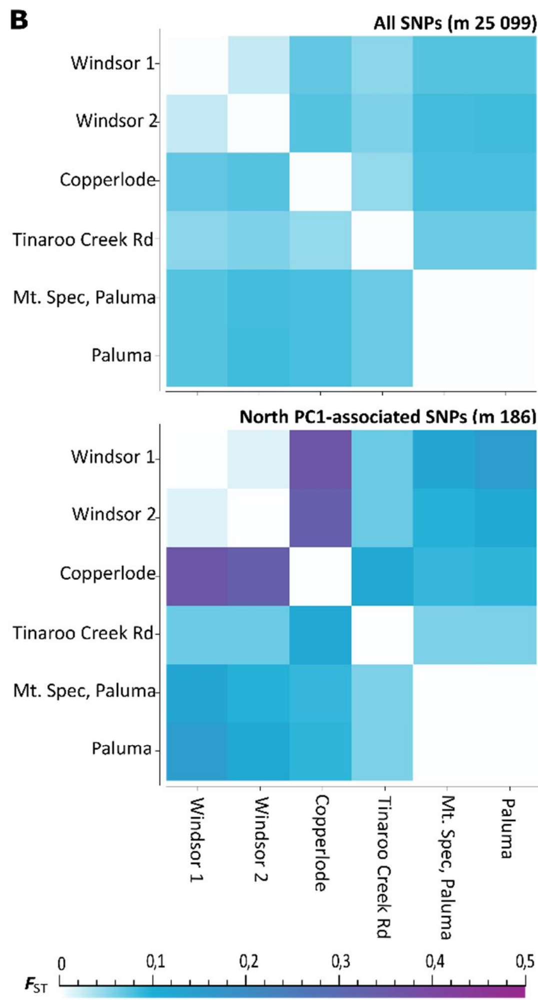
Figure 2 Ancestry assignment of chromosomal segments. Assignments are shown to *E. grandis* (anc1, grey), non-*E. grandis Latoangulatae* (anc3, cyan) or *Maidenaria*-like (anc2, purple) using *EILA* for one seedling from the Copperlode provenance with no introgression (non-intr.), and selected individuals from Eungella (Eun) and Finch Hatton Gorge (FHG). Each colour dot represents a SNP (7736 SNPs shown along the 11 chromosome with Mbp positions on the x-axis). Both alleles assigned to *E. grandis* (anc1) = 0; one allele assigned to *E. grandis* the other to anc2 = -1, both alleles assigned to anc2 = -2, one allele assigned to *E. grandis* and the other to anc3 = 1, both alleles assigned to anc3 = 2, one allele assigned to anc2 and the other to anc3 = -2.5. Breakpoint penalty $\lambda = 30$.

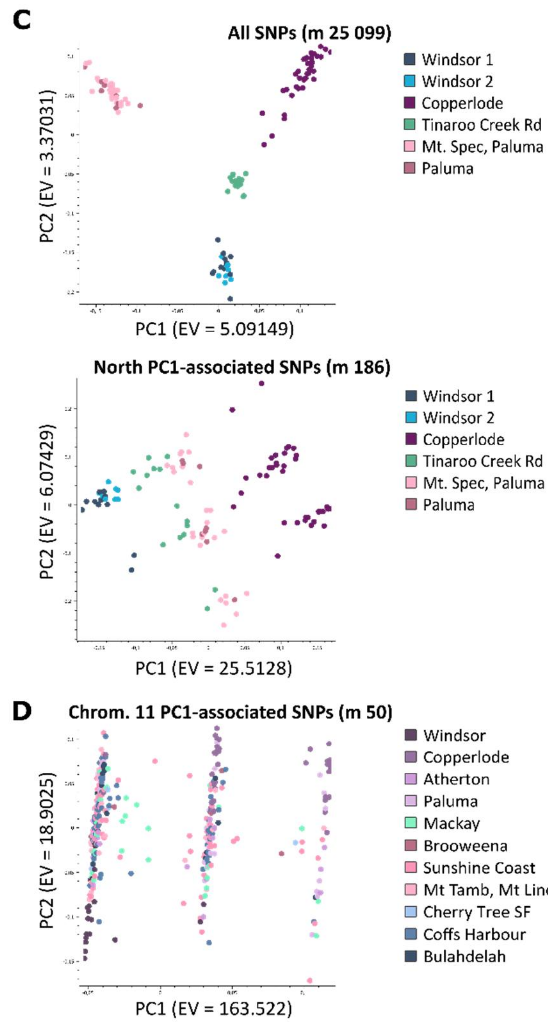
3.3.2 Environment Association Analysis (EAA)

Correlation patterns among environmental variables differed between the Northern and Southern subpopulations (**Table S4**). Furthermore, fewer SNPs were associated with environmental clines in the South than in the North (**Table S4**). Although GDM was successful for Southern subpopulation environment-associated SNP sets (**Fig. S12**), the models did not perform as well as those for the North.

Population differentiation patterns derived from PC1-associated markers in the North showed highest correlation with distance to coast, this was supported by the GDM where changes in distance to coast was most predictive of population differentiation patterns (**Fig. S13**) and the most important predictor variable in the permutation test. Distance to coast was also important in PC2-associated population differentiation (**Table S4**). As expected, dry period average precipitation (Bio14), soil depth, soil pedality and warmest period maximum temperature (Bio05) were significant predictors in GDMs for Bio14-, soil depth-, soil pedality- and Bio05-associated population differentiation patterns, respectively (**Fig. S13**). Markers associated with PC1 in the Northern subpopulation also revealed unexpected population differentiation patterns (**Fig. 3C**). Population structure PCA split each provenance into three groups along the y-axis. Since this pattern could be indicative of a segregating chromosomal rearrangement (Faria *et al.* 2019), a genome-wide scan was done for large regions of high LD, which revealed a ~700 kb block containing 17 SNP markers in high LD ($R^2 > 0.8$) on chromosome 11. Population structure across the species range based on North PC1-associated markers on chromosome 11 separated provenances into three clusters (**Fig. 3D**).


A
Windsor 1
Windsor 2
Tinaroo Creek Rd
Mt. Spec, Paluma
Paluma
Copperlode
Distance to coast (km x1000)
Elevation (m)
Nitrogen pre-European (kg/ha)
Phosphorus pre-European (kg/ha)


B
All SNPs (m 25 099)
North PC1-associated SNPs (m 186)
Windsor 1
Windsor 2
Copperlode
Tinaroo Creek Rd
Mt. Spec, Paluma
Paluma
F_{ST}


C
All SNPs (m 25 099)
PC2 (EV = 3.37031)
PC1 (EV = 5.09149)
North PC1-associated SNPs (m 186)
PC2 (EV = 6.07429)
PC1 (EV = 25.5128)
D
Chrom. 11 PC1-associated SNPs (m 50)
PC2 (EV = 18.9025)
PC1 (EV = 163.522)
Windsor
Copperlode
Atherton
Paluma
Mackay
Brooweena
Sunshine Coast
Mt Tamb, Mt Lind
Cherry Tree SF
Coffs Harbour
Bulahdelah

E
R^2

Figure 3 Environmental association analysis for PC1 of environmental variables in the Northern subpopulation of *E. grandis*. A) Northern subpopulation environmental factors contributing to PC1 (**Table S3**) with loadings > 0.7. Colour shades for different provenances in the Northern subpopulations match that in **C** and provenances are ordered as furthest from coast to closest to coast. Environmental data layers were obtained from Atlas of Living Australia (occurrence data download at http://www.ala.org.au. Accessed 28 November 2019; **Fig. S4**). **B**) Heatmap of F_{ST} estimates as determined using all 25 099 SNPs (top) and 186 PC1-associated SNPs (bottom). **C**) PCA for population structure based on all 25 099 SNPs (top) and 186 PC1-associated SNPs (bottom) with Eigenvalues indicated in parentheses. **D**) Species-wide population structure based on Northern subpopulation PC1-associated SNPs on chromosome 11 separates the population into three genotypic classes. **E**) Heatmap of pair-wise LD, measured as squared correlation (R^2), for 186 PC1-associated SNPs on chromosome 11 with an enlargement of the 736 kb LD block of 17 SNP markers.

Gene ontology term enrichment for biological processes was tested for 13 environment-associated SNP sets against the full SNP-captured gene space of 13 014 genes (**Table S5**). Defense response terms were enriched in all query sets and signalling processes were enriched for environment-associated loci in the Southern subpopulation (**Table S6**). No enrichment was detected for the control data sets.

4 DISCUSSION

We conducted a population genomics study on the *E. grandis* native species range. With 33 provenances genotyped, this study represents the most comprehensive analysis of the natural genetic diversity for the species (Jones *et al.* 2006; Silva-Junior & Grattapaglia 2015; Song *et al.* 2016). With the exception of Silva-Junior and Grattapaglia (2015) who used variants detected from pooled resequencing data of two provenances, this is also the best genome coverage achieved with one informative SNP for every 25.5 kb of the genome. Because we aim to test hypotheses generated in this study in comprehensive common garden trials, our sampling strategy was limited to available single-tree seedlot collections with sufficient seed for establishment of such trials, and seedlings were used as proxy of the genotypes of standing trees from which seeds were derived. Consequently, some provenances were not adequately represented to allow their inclusion for population differentiation estimates. Still, we were able to interrogate the genetic space for putatively adaptive genetic variation, and uncover, to our knowledge, the first genomic evidence of ongoing interspecific introgression in *E.*

grandis. We also describe the first observation of a potential chromosomal structural variant in the genus.

Suspected introgression was assessed at subchromosomal resolution by modifying the *EILA* approach (Yang *et al.* 2013) for mapping the ancestry of genomic segments. In most studies investigating genomic signatures of hybridization, the parental species are known and genotyped, (e.g. Christe *et al.*, 2017; Suarez-Gonzalez *et al.*, 2018). Since the introgressing species in our study was unknown, we devised a method to leverage the comprehensive species-wide genetic variation captured in our data set against interspecific genetic variation represented by non-*E. grandis* species as published by Silva-Junior *et al.* (2015). Power to detect non-*E. grandis* ancestry was limited in the closely related species, *E. saligna*, where genomic assignments were equivalent to one allele originating from *E. grandis* and the other from non-*E. grandis* species of section *Latoangulatae* (**Fig. S3**). This was in part due to the recent divergence of the two species, and partly due to the restricted number of *E. saligna* individuals represented in anc3. Despite this limitation, the modified approach performed well in the samples of known ancestry at breakpoint penalty 30 and higher.

Genome-wide ancestry mapping revealed extensive introgression in Finch Hatton Gorge and Eungella provenances in the Mackay subpopulation (**Fig. 2** and **Fig. S8-S9**). Introgressed genomic segments in seedlings from the Brooweena provenance, which is at the northern edge of the Southern subpopulation, were fewer and larger than introgressed segments seen in Finch Hatton Gorge and Eungella, even at $\lambda = 15$ (**Fig. S10**), suggesting that this represents a different introgression event. Changes in precipitation was most predictive of the population differentiation pattern of the species when the introgressed provenances were included in the GDM (model A), particularly at lower dry period average precipitation (Bio14) values (**Fig. S11**). The eastern coastline of Australia was wetter and greener than present, up to 5000 years ago, when the onset of El Niño Southern Oscillation conditions resulted in increased aridity and variable climates (Donders *et al.* 2007; Marx *et al.* 2009). Our hypothesis is that *E. grandis* occurred as a continuous population until the region between the

present-day Northern and Southern subpopulations became drier, and hybrid individuals began to outperform the pure species. In response to increased aridity, the Mackay provenances potentially fall within the continuum between introgression and hybrid speciation (Nieto Feliner *et al.* 2017).

Natural hybridization has been reported in several eucalypts (Griffin *et al.* 1988; McKinnon *et al.* 2001; Whitham *et al.* 1994; Willis *et al.* 2004). For *E. grandis*, hybrids with other species in section *Latoangulatae* have been reported based on morphological features (Jacobs 1981). However, as shown by Holman *et al.* (2003), morphological traits attributed to hybridization may in fact be the result of within-species adaptation to an environmental cline. This study provides the first molecular evidence of extensive interspecific introgression in *E. grandis* populations. Our study may also shed some light on the frequency of natural hybridization events in eucalypts. We observed up to 87 seedlings with non-*E. grandis* genomic segments (16% of genotyped individuals, excluding Eungella, Finch Hatton Gorge and Brooweena provenances) representing 58 families (**Table S7**). However, this does not seem to translate into pervasive introgression outside Mackay and Brooweena, suggesting reduced fitness in hybrids or rare hybridization events followed by neutral evolution of introgressed segments in the rest of the species' range. Differential introgression patterns in the introgressed provenances may underlie genomic regions containing genes involved in reproductive isolation and differential adaptation between *E. grandis* and the introgressing species (see review by Harrison and Larson, 2014), prompting further interrogation of non-introgressed vs introgressed genomic segments in the future.

We identified adaptively enriched SNP sets associated with environmental clines to expose subtle adaptively-enriched population structure not discernible using the total SNP set of 25 099 loci (enriched for 'neutral' variants). In light of the differences in gene flow patterns and differing correlations among environmental factors in the two subpopulations, we chose an EAA approach that would detect environment-associated loci given low levels of dispersal, variable demographic scenarios and varying selection intensity (De Villemereuil *et al.* 2014; Forester *et al.* 2016; Lotterhos & Whitlock 2015). We excluded introgressed provenances from the EAA, since the presence of

hybridization could result in spurious associations (Hoban *et al.* 2016). Our approach was modelled on that of Ingvarsson and Bernhardsson (2019) who used LFMM (Frichot & François 2015; Frichot *et al.* 2013) to identify loci associated with environmental clines, and performed GDM to compare the genetic change caused by isolation by distance to that caused by changes in climate. In the absence of common garden trials to test putatively adaptive genetic variation, our strategy was to find indirect support for our delineation of the adaptively enriched genetic space. Following the rational of Steane *et al.* (2014), our hypothesis was that adaptively enriched SNP sets would reveal population differentiation patterns that correlate with environmental clines and deviate from the demography of the species, as observed for North PC1-associated SNPs (**Fig. 3C**). Furthermore, we hypothesised that genes in LD with SNPs in the adaptively enriched sets would have functional enrichment for biological processes related to biotic and abiotic stress responses when compared against the total SNP-captured gene space.

Given the SNP density and predicted LD decay (Silva-Junior & Grattapaglia 2015), we did not remove the small proportion of SNPs that were in high LD. This fortuitous decision led to the discovery of a putative large structural variant on chromosome 11. Fluctuations in LD can be ascribed to the genomic architecture of, for example, centromeric regions; however, the abrupt decay from near perfect LD to no LD directly adjacent to the region on chromosome 11 supports the hypothesis that this is more likely a chromosomal rearrangement (Todesco *et al.* 2020). To our knowledge, this is the first report of such a structural polymorphism in *Eucalyptus*. Since all three karyotypes were observed across the species' range (**Fig. 3D**), we conclude that the polymorphism predates the split between the three subpopulations. Although it encompasses a genomic segment of approximately 700 kb (**Fig. 3E**), it still only represents a single variant, and thus we cannot exclude the possibility that the elevated allele frequency of the minor homokaryotype in Copperlode is due to neutral evolutionary processes such as genetic drift, as opposed to it conveying some adaptive advantage.

This study highlights the variable selection landscape in which a single species can evolve. We observed contrasting population stratification patterns between the Northern and Southern subpopulations in terms of neutral and potentially adaptive genetic variation. The fragmented distribution in the Northern subpopulation represents an aggregated landscape that may result in more discernible local adaptation (Forester *et al.* 2016). Loci associated with PC1 (**Fig. 3**) differentiated the subpopulation according to the correlated environmental cline. In contrast, we observed a lack of adaptive differentiation in the South. This could be due to the strength of local adaptation; in the presence of undisrupted gene flow and highly heterogeneous landscapes, as observed in the Southern subpopulation, local adaptation strength is predicted to be low (Forester *et al.* 2016). However, our inability to detect discernible adaptive population differentiation in the Southern subpopulation may also have been due to insufficient sampling depth for some Southern provenances. In addition, the high value of K derived from PCA and sNMF, compared to $K = 4$ obtained for the Northern subpopulation, may have resulted in over-correction for demography and increased false negatives (Forester *et al.* 2016).

Elevation and distance to coast are steady environmental variables in terms of accuracy in measurement and consistency over very long periods of time. All other factors, especially climatic variables, are subject to fluctuations within and between seasons. Distance to coast and elevation are also complex variables and perhaps represents a better proxy of the ecosystem in which the species is adapting. Since these two factors were important and significant in most of the GDM analyses that resulted in well-fitted models, they could serve as reliable proxies for the environmental cline shaping local adaptation in *E. grandis*. To pinpoint more accurately specific loci underlying adaptation to climates across seasons and years would require higher marker density, increased sample size, daily climate measures over several years and greater computation power than we could achieve in the present study (Harfouche *et al.* 2019).

Protein domains underlying adaptive variation have been shown to be highly expanded in eucalypts (Kersting *et al.* 2015). *E. grandis* also has among the highest prevalence of tandem gene duplications, known to harbour genes involved in adaptive processes (Butler *et al.* 2018; Hanada *et al.* 2008), of any sequenced plant genome (Myburg *et al.* 2014). This could complicate GO enrichment analysis for genes presumed to underlie adaptation. The problem is confounded when the sampled genetic space has been biased to single copy genomic regions as is the case for the informative SNPs used in this study (Silva-Junior *et al.* 2015). In spite of these potential difficulties, terms representing very large gene families in the *E. grandis* genome (e.g. 'protein phosphorylation' and 'signal transduction' terms amount to 3845 and 2544 genes, respectively) were enriched in one of the larger query sets, 'South all environment-associated'. Signal transduction is particularly relevant in fluctuating environments since proteins involved in these pathways are responsible for relaying intra- and extracellular stimuli detected by cellular receptors to responses in the organism (Huber 2007). Protein phosphorylation and signal transduction are also important in biotic stress response, which aligns with the 'defense response' GO term that was enriched in all environment-associated sets, suggesting that the adaptive gene space includes genomic regions underlying responses to abiotic and biotic stress.

5 CONCLUSIONS

We have shown that particular provenances of *E. grandis* may have captured adaptive genetic variation from other species through hybridization. Adaptive introgression, if not properly assayed, could reduce the accuracy of species distribution models under future climate conditions (Aitken *et al.* 2008). Hybrid success is generally low in environments to which the parental species are adapted (Rieseberg *et al.* 2003). Still, with low dispersal rates (Potts 1990) and long generation times, introgression may prove to be the leading mode for eucalypts to adapt to rapid climate change. Our modified ancestry mapping approach, using genus-wide genetic variation to identify genomic regions originating from alternative species, could prove valuable in studies of eucalypts and other species for which genus-wide genetic variation data are available. This is also a useful tool for breeding

populations where hybrids between geographically separated species with weak reproductive isolation barriers occur. The method permits an estimate of the extent of introgression and identification of broad genomic regions that tend to be retained preferentially, which could be used to decipher the genomic architecture underlying reproductive isolation and local adaptation.

The predominant environmental factors predictive of population differentiation patterns not explained by geographic separation, included distance to coast, elevation, wettest period precipitation (Bio14) and pre-European soil phosphorus content. Although we only considered abiotic environmental variables, GO enrichment analysis of genes in LD with associated SNPs provide support that these loci underlie adaptation to biotic as well as abiotic stresses. Other studies of adaptive genetic variation in the genus have identified abiotic factors affecting species distributions (Ahrens *et al.* 2019; Dillon *et al.* 2014; Jordan *et al.* 2017; Steane *et al.* 2014) and we hope to add some understanding regarding the complexity of the selection forces driving evolution in eucalypts. Due to the economic importance of the species, large common garden *E. grandis* trials are being established by Camcore (NC State University, Raleigh, NC) across various exotic environments, which will include 108 of the families represented in our study. Two of the sites established in South Africa will be expanded to include all of the families sampled here, with n ≈ 5000 seedlings, each. These trials will serve as experimental populations with which to test hypotheses generated from this discovery project.

ACKNOWLEDGEMENTS

The authors acknowledge B Smit and D Pinard (University of Pretoria) for technical assistance with seed germination and Gene Ontology enrichment analyses, respectively. Also, Sappi Forests for financial support towards seed acquisition, Mondi Forests South Africa for establishment of the gene conservation park, Camcore (North Carolina State University) and Department of Agriculture and Fisheries (Queensland) for assistance with phytosanitary certification and seed import. This work was funded in part by the National Research Foundation (NRF) of South Africa - Bioinformatics and

Functional Genomics Programme (BFG Grant UID 97911), the Technology and Human Resources for Industry Programme (THRIP Grant UID 96413), the Department of Science and Technology and Technology Innovation Agency (DST/TIA, Strategic Grant for the Eucalyptus Genomics Platform), the Forestry Sector Innovation Fund (FSIF *Eucalyptus* Genome Diversity Atlas grant) and by the Forest Molecular Genetics (FMG) Industry Consortium at the University of Pretoria. MO acknowledges PhD scholarship support from the NRF.

DATA ACCESSIBILITY

The genomic data analysed in this study is available online via the Dryad archives under accession https://doi.org/10.5061/dryad.p8cz8w9kp.

(URL for reviewers: https://datadryad.org/stash/share/D16t_bYvdKx8ljAI36TP_I75sBP0uXJvvMchwDKqCok)

AUTHOR CONTRIBUTIONS

MO and AM developed the idea. All authors contributed to the design of the study. DL contributed biological material. MO generated and analysed the data, and wrote the first draft of the manuscript. All authors read, edited and approved the final manuscript.

REFERENCES

Ahrens, C. W., Byrne, M., & Rymer, P. D. (2019). Standing genomic variation within coding and regulatory regions contributes to the adaptive capacity to climate in a foundation tree species. *Molecular Ecology, 28*, 2502-2516. doi:10.1111/mec.15092

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: Climate change outcomes for tree populations. *Evolutionary Applications, 1*(1), 95-111. doi:10.1111/j.1752-4571.2007.00013.x

Arnold, R. J., Burgess, I. P., & Allender, E. B. (1996). *Eucalyptus grandis* seed source variation for growth and form in the southern Murray-Darling Basin. *Australian Forestry, 59*(2), 114-119. doi:10.1080/00049158.1996.10674676

Bayly, M. J., Rigault, P., Spokevicius, A., Ladiges, P. Y., Ades, P. K., Anderson, C., . . . Woodrow, I. E. (2013). Chloroplast genome analysis of Australian eucalypts – *Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). *Molecular Phylogenetics and Evolution, 69*(3), 704-716.

Bean, A. (1989). *The distribution of Queensland eucalypts.* Nambour, Queensland: A. R. Bean.

Belbin, L., & Williams, K. J. (2016). Towards a national bio-environmental data facility: Experiences from the Atlas of Living Australia. *International Journal of Geographical Information Science, 30*(1), 108-125.

Bennett, A. F. (2016). Eucalypts, wildlife and nature conservation: From individual trees to landscape patterns. *Proceedings of the Royal Society of Victoria, 128*(1), 71-86.

Bragg, J. G., Supple, M. A., Andrew, R. L., & Borevitz, J. O. (2015). Genomic variation across landscapes: Insights and applications. *New Phytologist, 207*(4), 953-967. doi:10.1111/nph.13410

Burgess, I. (1988). Provenance trials of *Eucalyptus grandis* and *E. saligna* in Australia. *Silvae Genetica, 37*(5-6), 221-227.

Butler, J. B., Freeman, J. S., Potts, B. M., Vaillancourt, R. E., Grattapaglia, D., Silva-Junior, O. B., . . . Shepherd, M. (2018). Annotation of the *Corymbia* terpene synthase gene family shows broad conservation but dynamic evolution of physical clusters relative to *Eucalyptus*. *Heredity, 121*(1), 87-104. doi:10.1038/s41437-018-0058-1

Christe, C., Stölting, K. N., Paris, M., Fraïsse, C., Bierne, N., & Lexer, C. (2017). Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Molecular Ecology, 26*(1), 59-76.

Darrow, W. K. (1983). Provenance-type trials of *Eucalyptus grandis* and *E. saligna* in South Africa: Eight year results. *South African Forestry Journal, 126*(1), 30-38.

De Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology, 23*(8), 2006-2019. doi:10.1111/mec.12705

Dillon, S., McEvoy, R., Baldwin, D. S., Rees, G. N., Parsons, Y., & Southerton, S. (2014). Characterisation of adaptive genetic diversity in environmentally contrasted populations of *Eucalyptus camaldulensis* Dehnh. (river red gum). *PLoS ONE, 9*(8), e103515. doi:10.1371/journal.pone.0103515

Donders, T. H., Haberle, S. G., Hope, G., Wagner, F., & Visscher, H. (2007). Pollen evidence for the transition of the Eastern Australian climate system from the post-glacial to the present-day ENSO mode. *Quaternary Science Reviews, 26*(11), 1621-1637. doi:http://dx.doi.org/10.1016/j.quascirev.2006.11.018

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology, 14*(8), 2611-2620.

Faria, R., Chaube, P., Morales, H. E., Larsson, T., Lemmon, A. R., Lemmon, E. M., . . . Butlin, R. K. (2019). Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Molecular Ecology, 28*(6), 1375-1393. doi:10.1111/mec.14972

Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions, 13*, 252-264. doi:10.1111/j.1472-4642.2007.00341.x

Fitzpatrick, M. C., Sanders, N. J., Normand, S., Svenning, J.-C., Ferrier, S., Gove, A. D., & Dunn, R. R. (2013). Environmental and historical imprints on beta diversity: Insights from variation in rates of species turnover along gradients. *Proceedings of the Royal Society B: Biological Sciences, 280*(1768), 20131201.

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology, 25*(1), 104-120. doi:10.1111/mec.13476

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution, 6*(8), 925-929.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics, 196*(4), 973-983. doi:10.1534/genetics.113.160572

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution, 30*(7), 1687-1699. doi:10.1093/molbev/mst063

Grattapaglia, D., & Kirst, M. (2008). *Eucalyptus* applied genomics: From gene sequences to breeding tools. *New Phytologist, 179*(4), 911-929.

Griffin, A., Burgess, I., & Wolf, L. (1988). Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'Hérit. - a review. *Australian Journal of Botany, 36*(1), 41-66. doi:https://doi.org/10.1071/BT9880041

Gugger, P. F., Liang, C. T., Sork, V. L., Hodgskiss, P., & Wright, J. W. (2018). Applying landscape genomic tools to forest management and restoration of Hawaiian koa (*Acacia koa*) in a changing environment. *Evolutionary Applications, 11*(2), 231-242.

Gunn, B. (2001). *Australian tree seed centre operations manual*: CSIRO Forestry and Forest Products.

Hanada, K., Zou, C., Lehti-Shiu, M. D., Shinozaki, K., & Shiu, S.-H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiology, 148*(2), 993-1003. doi:10.1104/pp.108.122457

Harfouche, A. L., Jacobson, D. A., Kainer, D., Romero, J. C., Harfouche, A. H., Scarascia Mugnozza, G., . . . Altman, A. (2019). Accelerating climate resilient plant breeding by applying next-generation artificial intelligence. *Trends in Biotechnology, 37*(11), 1217-1235. doi:https://doi.org/10.1016/j.tibtech.2019.05.007

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity, 105*(S1), 795-809.

Harwood, C. (2011). New introductions–doing it right. *Developing a eucalypt resource: learning from Australia and elsewhere: University of Canterbury. Christchurch, New Zealand: Wood Technology Research Centre*, 43-54.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., . . . Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *The American Naturalist, 188*(4), 379-397. doi:doi:10.1086/688018

Holman, J. E., Hughes, J. M., & Fensham, R. J. (2003). A morphological cline in *Eucalyptus*: A genetic perspective. *Molecular Ecology, 12*(11), 3013-3025. doi:10.1046/j.1365-294X.2003.01970.x

Huber, S. C. (2007). Exploring the role of protein phosphorylation in plants: From signalling to metabolism. *Biochemical Society Transactions, 35*(1), 28-32. doi:10.1042/bst0350028

Iglesias-Trabado, G., & Wilstermann, D. (2008). *Eucalyptus* universalis. Global cultivated eucalypt forests map 2008. *GIT Forestry Consulting's EUCALYPTOLOGICS: Information Resources on Eucalyptus Cultivation Worldwide,< http://www. git-forestry. com.*

Ingvarsson, P. K., & Bernhardsson, C. (2019). Genome-wide signatures of environmental adaptation in European aspen (*Populus tremula*) under current and future climate conditions. *Evolutionary Applications, 0*(0), 1-11. doi:10.1111/eva.12792

Jacobs, M. R. (1981). *Eucalypts for planting*. Rome: Food and Agriculture Organization of the United Nations.

Jones, M. E., Shepherd, M., Henry, R. J., & Delves, A. (2006). Chloroplast DNA variation and population structure in the widespread forest tree, *Eucalyptus grandis*. *Conservation Genetics, 7*(5), 691-703. doi:10.1007/s10592-005-9104-7

Jones, R. C., Nicolle, D., Steane, D. A., Vaillancourt, R. E., & Potts, B. M. (2016). High density, genome-wide markers and intra-specific replication yield an unprecedented phylogenetic reconstruction of a globally significant, speciose lineage of *Eucalyptus*. *Molecular Phylogenetics and Evolution, 105*, 63-85. doi:https://doi.org/10.1016/j.ympev.2016.08.009

Jordan, R., Dillon, S. K., Prober, S. M., & Hoffmann, A. A. (2016). Landscape genomics reveals altered genome wide diversity within revegetated stands of *Eucalyptus microcarpa* (Grey Box). *New Phytologist, 212*(4), 992-1006.

Jordan, R., Hoffmann, A. A., Dillon, S. K., & Prober, S. M. (2017). Evidence of genomic adaptation to climate in *Eucalyptus microcarpa*: Implications for adaptive potential to projected climate change. *Molecular Ecology, 26*(21), 6002-6020. doi:10.1111/mec.14341

Kersting, A. R., Mizrachi, E., Bornberg-Bauer, E., & Myburg, A. A. (2015). Protein domain evolution is associated with reproductive diversification and adaptive radiation in the genus *Eucalyptus*. *New Phytologist, 206*(4), 1328-1336. doi:10.1111/nph.13211

Köppen, W. (2011). The thermal zones of the Earth according to the duration of hot, moderate and cold periods and to the impact of heat on the organic world. *Meteorologische Zeitschrift, 20*(3), 351-360.

Köppen, W. P. (1884). *Die Wärmezonen der Erde, nach der Dauer der heissen, gemässigten und kalten Zeit und nach der Wirkung der Wärme auf die organische Welt betrachtet*: publisher not identified.

Landguth, E. L., Holden, Z. A., Mahalovich, M. F., & Cushman, S. A. (2017). Using landscape genetics simulations for planting blister rust resistant whitebark pine in the US Northern Rocky Mountains. *Frontiers in Genetics, 8*, 9.

Langat, D. K., & Kariuki, J. G. (2004). Variations in heartwood proportion and basic density between provenances of *Eucalyptus grandis* in Kenya. *Journal of Tropical Forest Science, 16*(2), 179-186.

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology, 24*(5), 1031-1046. doi:10.1111/mec.13100

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics, 4*(12), 981-994.

Mangwanda, R., Myburg, A. A., & Naidoo, S. (2015). Transcriptome and hormone profiling reveals *Eucalyptus grandis* defence responses against *Chrysoporthe austroafricana*. *BMC Genomics, 16*(1), 319. doi:10.1186/s12864-015-1529-x

Manion, G., Lisk, M., Ferrier, S., Nieto‐Lugilde, D., Mokany, K., & Fitzpatrick, M. (2018). *gdm*: Generalized dissimilarity modeling. *R package ver. 1.3. 8.*

Marx, S. K., McGowan, H. A., & Kamber, B. S. (2009). Long-range dust transport from eastern Australia: A proxy for Holocene aridity and ENSO-type climate variability. *Earth and Planetary Science Letters, 282*(1), 167-177. doi:http://dx.doi.org/10.1016/j.epsl.2009.03.013

Matheson, A., & Mullin, L. (1987). Variation among neighbouring and distant provenances of *Eucalyptus grandis* and *E. tereticornis* in Zimbabwean field trials. *Australian Forest Research, 17*(3), 233-250.

McKenzie, N., Jacquier, D., Ashton, L., & Cresswell, H. (2000). Estimation of soil properties using the Atlas of Australian Soils. *CSIRO Land and Water Technical Report, 11*(00), 1-12.

McKinnon, G. E., Jordan, G. J., Vaillancourt, R. E., Steane, D. A., & Potts, B. M. (2004). Glacial refugia and reticulate evolution: The case of the Tasmanian eucalypts. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 359*(1442), 275-284.

McKinnon, G. E., Vaillancourt, R. E., Jackson, H. D., & Potts, B. M. (2001). Chloroplast sharing in the Tasmanian eucalypts. *Evolution, 55*(4), 703-711. doi:10.1111/j.0014-3820.2001.tb00806.x

Mizrachi, E., Hefer, C. A., Ranik, M., Joubert, F., & Myburg, A. A. (2010). *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. *BMC Genomics, 11*, 681-681. doi:10.1186/1471-2164-11-681

Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., . . . Schmutz, J. (2014). The genome of *Eucalyptus grandis*. *Nature, 510*(7505), 356-362. doi:10.1038/nature13308

Myburg, A. A., Potts, B. M., Marques, C. M., Kirst, M., Gion, J.-M., Grattapaglia, D., & Grima-Pettenatti, J. (2007). Eucalypts. In C. Kole (Ed.), *Forest Trees* (pp. 115-160). Berlin, Heidelberg: Springer Berlin Heidelberg.

Nieto Feliner, G., Álvarez, I., Fuertes-Aguilar, J., Heuertz, M., Marques, I., Moharrek, F., . . . Villa-Machío, I. (2017). Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity, 118*(6), 513-516. doi:10.1038/hdy.2017.7

Oates, C. N., Külheim, C., Myburg, A. A., Slippers, B., & Naidoo, S. (2015). The transcriptome and terpene profile of *Eucalyptus grandis* reveals mechanisms of defense against the insect pest, *Leptocybe invasa*. *Plant and Cell Physiology, 56*(7), 1418-1428. doi:10.1093/pcp/pcv064

Pinard, D., Fierro, A. C., Marchal, K., Myburg, A. A., & Mizrachi, E. (2019). Organellar carbon metabolism is coordinated with distinct developmental phases of secondary xylem. *New Phytologist, 222*(4), 1832-1845. doi:10.1111/nph.15739

Pollock, L. J., Bayly, M. J., Nevill, P. G., & Vesk, P. A. (2013). Chloroplast DNA diversity associated with protected slopes and valleys for hybridizing *Eucalyptus* species on isolated ranges in south‐eastern Australia. *Journal of Biogeography, 40*(1), 155-167.

Potts, B. (1990). The response of eucalypt populations to a changing environment. *Tasforests, 2*, 179-193.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics, 38*(8), 904.

Pryor, L., & Johnson, L. (1981). *Eucalyptus*, the universal Australian. *Ecological biogeography of Australia, 1*, 500-536.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842. doi:10.1093/bioinformatics/btq033

Ramsay, J. O. (1988). Monotone Regression Splines in Action. *Statist. Sci., 3*(4), 425-441. doi:10.1214/ss/1177012761

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology, 24*(17), 4348-4370. doi:10.1111/mec.13322

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., . . . Lexer, C. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science, 301*(5637), 1211-1216. doi:10.1126/science.1086949

Rutherford, S., Rossetto, M., Bragg, J. G., McPherson, H., Benson, D., Bonser, S. P., & Wilson, P. G. (2018). Speciation in the presence of gene flow: Population genomics of closely related and diverging *Eucalyptus* species. *Heredity, 121*(2), 126.

Shryock, D. F., Havrilla, C. A., DeFalco, L. A., Esque, T. C., Custer, N. A., & Wood, T. E. (2015). Landscape genomics of *Sphaeralcea ambigua* in the Mojave Desert: A multivariate, spatially-explicit approach to guide ecological restoration. *Conservation Genetics, 16*(6), 1303-1317.

Silva-Junior, O. B., Faria, D. A., & Grattapaglia, D. (2015). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytologist, 206*(4), 1527-1540. doi:10.1111/nph.13322

Silva‐Junior, O. B., & Grattapaglia, D. (2015). Genome‐wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. *New Phytologist, 208*(3), 830-845.

Skolmen, R. G. (1986). Performance of Australian provenances of *Eucalyptus grandis* and *Eucalyptus saligna* in Hawaii. *USDA, Forest Service Research Paper PSW-181*.

Song, Z., Zhang, M., Li, F., Weng, Q., Zhou, C., Li, M., . . . Gan, S. (2016). Genome scans for divergent selection in natural populations of the widespread hardwood species *Eucalyptus grandis* (Myrtaceae) using microsatellites. *Scientific Reports, 6*, 34941. doi:10.1038/srep34941

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. doi:10.1111/mec.12751

Suarez-Gonzalez, A., Hefer, C. A., Lexer, C., Cronk, Q. C. B., & Douglas, C. J. (2018). Scale and direction of adaptive introgression between black cottonwood (*Populus trichocarpa*) and balsam poplar (*P. balsamifera*). *Molecular Ecology, 27*(7), 1667-1680. doi:10.1111/mec.14561

Supple, M. A., Bragg, J. G., Broadhurst, L. M., Nicotra, A. B., Byrne, M., Andrew, R. L., . . . Borevitz, J. O. (2018). Landscape genomic prediction for restoration of a *Eucalyptus* foundation species under climate change. *Elife, 7*, e31835.

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., . . . Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*. doi:10.1038/s41586-020-2467-6

Van den Berg, G. J., Verryn, S. D., Chirwa, P. W., & Van Deventer, F. (2016). Genetic parameters and genotype by environment interaction of *Eucalyptus grandis* populations used in intraspecific hybrid production in South Africa. *Southern Forests: a Journal of Forest Science*, 1-9. doi:10.2989/20702620.2016.1254900

Vining, K. J., Romanel, E., Jones, R. C., Klocko, A., Alves-Ferreira, M., Hefer, C. A., . . . Strauss, S. H. (2015). The floral transcriptome of *Eucalyptus grandis*. *New Phytologist, 206*(4), 1406-1422. doi:10.1111/nph.13077

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution, 38*(6), 1358-1370. doi:10.2307/2408641

Whitham, T. G., Morrow, P. A., & Potts, B. M. (1994). Plant hybrid zones as centers of biodiversity: The herbivore community of two endemic Tasmanian eucalypts. *Oecologia, 97*(4), 481-490. doi:10.1007/bf00325886

Willing, E.-M., Dreyer, C., & Van Oosterhout, C. (2012). Estimates of genetic differentiation measured by FST do not necessarily nequire large sample sizes when using many SNP markers. *PLoS ONE, 7*(8), e42649. doi:10.1371/journal.pone.0042649

Willis, K. J., Bennett, K. D., Walker, D., McKinnon, G. E., Jordan, G. J., Vaillancourt, R. E., . . . Potts, B. M. (2004). Glacial refugia and reticulate evolution: The case of the Tasmanian eucalypts. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359*(1442), 275-284. doi:doi:10.1098/rstb.2003.1391

Yang, J. J., Li, J., Buu, A., & Williams, L. K. (2013). Efficient inference of local ancestry. *Bioinformatics, 29*(21), 2750-2756.

[dataset] Mostert-O'Neill, M.M., Reynolds, S.M., Acosta, J.J., Lee, D.J., Borevitz, J.O. & Myburg, A.A. (2019). Genomic evidence of introgression and adaptation in a model subtropical tree species, *Eucalyptus grandis. Dryad.* https://doi.org/10.5061/dryad.p8cz8w9kp.