Specialty Wood Products Research Partnership

FOREST GROWERS RESEARCH

Technical Report

Genetic structure and diversity in the NZDFI *Eucalyptus bosistoana* and *E. argophloia* breeding populations

-

Unexamined PhD chapter

Clemens Altaner on behalf of

Seoljong Kim supervised by **Pieter Pelser, Luis Apiolaza, Tammy Steeves, Clemens Altaner**

Date: 1st June 2023

Report: SWP-T168

TABLE OF CONTENTS

supported by forestgrowers commodity levy

Disclaimer

This report has been prepared by University of Canterbury for Forest Growers Research Ltd (FGR) subject to the terms and conditions of a research services agreement dated 1 July 2015.

The opinions and information provided in this report have been provided in good faith and on the basis that every endeavour has been made to be accurate and not misleading and to exercise reasonable care, skill and judgement in providing such opinions and information.

Under the terms of the Services Agreement, University of Canterbury's liability to FGR in relation to the services provided to produce this report is limited to the value of those services. Neither University of Canterbury nor any of its employees, contractors, agents or other persons acting on its behalf or under its control accept any responsibility to any person or organisation in respect of any information or opinion provided in this report in excess of that amount.

EXECUTIVE SUMMARY

Preamble

This report on SWP Work Plan WP126 is of preliminary nature. It is an unexamined thesis chapter from Seoljong Kim's PhD thesis, developed at the University of Canterbury under the supervision of Pieter Pelser, Luis Apiolaza, Tammy Steeves and Clemens Altaner. The thesis is scheduled to be submitted in August 2023, shortly after the SWP programme has finished. Once successfully defended, the PhD thesis will be publicly available online through the University of Canterbury's library (likely by the end of 2023).

Abstract

Eucalyptus bosistoana is a key breeding species of the NZDFI, which aims to establish forestry plantations of ground-durable, high-value timber hardwoods in warmer dry environments of New Zealand's east coast regions. Seeds of plants collected in 2008-2012 from trees identified as *E. bosistoana* in Australia, and of *E. argophloia*, a reputedly closely related species of secondary interest and to NZDFI, were grown in breeding trials in New Zealand. To inform the NZDFI breeding program, leaf samples of 221 breeding families of both species were genotyped using a *Eucalyptus* 72kSNP Axiom array to

1) identify patterns of genetic structure among *E. bosistoana* populations;
2) assess the genetic diversity of populations of *E. bosistoana* and *E. argophloia*; and
3) determine the taxonomic identity of breeding families that are morphologically deviating from other *E. bosistoana* families or that were grown from seed collected outside the known distribution area of *E. bosistoana*.

The findings are best understood by studying Figure 12.

Despite the initial suspicion on the morphologically deviating plants that might be hybrids between *E. argophloia* and *E. bosistoana*, genetic structure study showed that plants of *E. argophloia* within the NZDFI breeding trials were apparently different from *E. bosistoana* identified individuals. Some populations of *E. bosistoana* labelled individuals within the breeding programme were instead identified as *E. melliodora*.

Although our samples originated from a wide geographic range of localities in Australia, STRUCTURE analysis suggests only weak genetic structure among *E. bosistoana* collection sites. However, isolation by distance among the collecting sites was statistically significant. Additionally, evidence of hybridization between *E. bosistoana* and *E. melliodora* was found in some populations.

The level of genetic diversity was similar among the species, while *E. argophloia* showed higher level of inbreeding.

INTRODUCTION

Eucalyptus bosistoana is an emerging plantation species that produces naturally ground-durable timber (Class 1 and 2 Australian Standard, AS5606-2005) and is drought tolerant (NICHOLAS AND MILLEN 2012). Its timber is suitable to supply domestic and global markets for agricultural posts, outdoor joinery and engineered wood products. The NZDFI has established an *E. bosistoana* breeding programme that also included *E. argophloia*, a species closely related with *E. bosistoana* (THORNHILL et al. 2019), which has colored heartwood for its potential to hybridize with *E. bosistoana* as this may increase its commercial value in the market.

The NZDFI has established breeding trials of *E. bosistoana* and *E. argophloia* in New Zealand since 2009. Seeds collected in Australia of known progeny (i.e. information of mother trees) are being used for these trials (Appendix 2). In these trials, the trees have been phenotyped for traits that are of commercial interest, including growth and form, heartwood quantity and quality, and growth strain (DAVIES *et al.* 2017; LI AND ALTANER 2018; MISHRA *et al.* 2019; SHWE AND LEUNG 2020). This allowed to identify *E. bosistoana* families with superior characteristics for propagation and deployment in plantations. The first improved *E. bosistoana* plants became commercially available in 2021. However, the genetic relationship of the trees within and between families is unknown, introducing inaccuracy into breeding. A genomic approach can quantify the genetic relationship within breeding populations of the species, allowing the calculation of more accurate breeding values and management of the genetic diversity for the breeding programme. Joining the Eucalyptus 72kSNP Axiom array production and deployment initiative (ESAI) in 2018 gave access to a high-throughput customized SNP genotyping system for this study.

The effectiveness of a breeding program largely depends on the process of choosing certain individuals to be parents of the next generation based on their desirable traits (i.e. selection) and genetic diversity of its gene resource population (breeding stock) (ALLIER *et al.* 2020). The analysis of genetic structure among and diversity within natural populations can provide information for plant breeders to select and manage the gene resource with the goal of maximizing the breeding potential and genetic diversity of the population. To optimize genetic gain, it is crucial to map genetic patterns of population structure using molecular technologies, as this knowledge can inform and enhance breeding strategies. In this context, investigating patterns of genetic structure and diversity has been carried out to enhance tree species for forestry in various genera including *Acacia* (BAIRU *et al.* 2021), *Eucalyptus* (YEOH *et al.* 2012), *Pinus* (CHHATRE *et al.* 2013) and *Populus* (CASTIGLIONE *et al.* 2010) for last decade.

Eucalyptus bosistoana, also referred to as 'Coast Grey box' or 'Bosisto's box, is native to the eastern coast of Australia, specifically in New South Wales and Victoria. Most of the habitat of *E. bosistoana* is located within the Lowland Grassy Woodland ecosystem in the South Coast region of New South Wales. This is an endangered ecological community under the Threatened Species Conservation Act 1995 due to fragmentation and loss of ecological connectivity. *Eucalyptus argophloia*, often referred to as ‘Queensland western white gum’ or ‘Queensland white gum’, is native to a highly restricted area in Chinchilla, Queensland, where only one wild population is known to exist. This species is classified as 'vulnerable' under the Queensland Nature Conservation Act (1992). Given the ongoing conservation challenges that both species are facing, examining the genetic structure and diversity of these species not only provides opportunities to inform their commercial breeding program, but also to inform and guide conservation management efforts.

In the NZDFI breeding trials, some individuals grown from seeds from mother plants identified in situ as *E. bosistoana* and located in two collecting areas in the northern part of its distribution area (referred to as P1 and P2 throughout this study; Fig. 1.8-2) exhibit distinct morphological differences from most other plants of the same species in the breeding trials. These plants resemble *E. argophloia* in having lanceolate leaves during their juvenile stage, whereas young leaves of *E. bosistoana* are typically known to be oblong to elliptical to ovate in shape (SLEE *et al.* 2015). This could be explained by human errors such as mislabelling during seed collection and propagation or misidentification while gathering seeds. Alternatively, the morphological diversity could potentially be explained by the high adaptability and phenotypic plasticity that some *Eucalyptus* species display (GRATTAPAGLIA *et al.* 2012). It is also possible that the morphologically deviating plants are natural hybrids of *E. bosistoana* and *E. argophloia* (or hybrids with other

species). Furthermore, when seeds were collected for the NZDFI project near Waterloo, approximately 9 km north of the town of Beaufort, Victoria in 2010, some mother trees were identified as *E. bosistoana*. This area was however not recorded as an area where populations of this species occur (SLEE *et al.* 2015) and these plants may therefore have been misidentified. Knowing the taxonomic identity of a species is essential for ensuring genetic purity in breeding programs. If a species is misidentified, there is a risk of unintentionally introducing genetic variability from another species or hybridization with other species, which can compromise the integrity of the breeding program.

In this chapter, I aim to understand genetic patterns of the species to inform the NZDFI breeding program. Together with *E. bosistoana*, the genomic data of a population of *E. argophloia* is generated. These datasets will be used for understating patterns of genetic structure and diversity of the species for breeding program. In addition, ecological implications will be discussed as these patterns can contribute to knowledge for conservation and management of the species in the wild. The specific objectives of this chapter are: 1. to identify patterns of genetic structure among *E. bosistoana* populations; 2. to assess the genetic diversity of populations of *E. bosistoana* and *E. argophloia*; 3. to determine the taxonomic identity of breeding families that are morphologically deviating from other *E. bosistoana* families or that were grown from seed collected outside the known distribution area of *E. bosistoana*.

METHODS

Sampling sites

Leaf samples for DNA extraction and voucher specimens of plants labelled as *E. bosistoana* in the NZDFI breeding population were collected from the Avery and Dillion trials in December 2020, the Avery, Cravens, Dillon, and Lawson trials in March 2020, the Ngaumu trial in November 2019 and the Martin trial in October 2019 (Appendix 2). Additionally, samples of *E. argophloia* were collected from the Avery, Cuddons, and Dillons trials in March and December 2020. All these trials are common garden experiments of the NZDFI breeding programme in the North and South Island of New Zealand. These sampling sites were chosen to include plants from as many breeding families as possible in this study.

Plant collecting

Leaf samples were taken from 1353 plants labelled as *E. bosistoana* representing 178 breeding families used in the NZDFI programme. In addition, 102 samples were collected from *E. argophloia*. These represented 35 breeding families. The number of collected samples per family varied between 1 to 39. This number varied for several reasons: 1) in terms of studying patterns of genetic structure it requires only one sample per family, 2) for the study of the mating system (not part of this chapter), it requires as multiple samples per family for estimating pairwise relatedness within families and the number of samples is depending on availability in the NZDFI breeding trials at the times of fieldwork, 3) certain families are required to have more than 30 individuals sampled to compare the performance of the trees (i.e. poorly and well growing trees) (not part of this chapter).

Depending on the size of the leaves of each tree, 2-5 leaves per tree were collected and placed in a plastic bag labelled with the trial block, tree, and family NZDFI identification numbers, and the sampled trees were labelled and tagged in the same way as the bags for reference tracking. The bags with collected leaf samples were stored in a portable freezer at -17˚C at the end of each day and later transferred into the freezer in the laboratory of the School of Biological Sciences at UC for longer storage -80˚C. All samples were completely desiccated at 35˚C in a drier at the University of Canterbury Herbarium (CANU) for about four days for prior to DNA extraction.

Voucher specimens consisting of leafy twigs of 130 individuals from 66 families identified as *E. bosistoana* and 27 individuals from 15 families of *E. argophloia* were pressed and stored in a dry area during the field work and put into the drier at the CANU for about four days for complete desiccation.

DNA extraction and quality check

All collected leaf samples were stored in a -80˚C freezer for a minimum of 72 hours immediately prior to DNA extraction to make the leafy tissue brittle and easier to break open for cell lysis. Around 20 mg of tissue of each leaf sample was subsequently cut with scissors into small fragments and placed in a 1.7 ml centrifuge tube with three metal beads. This was shaken for 2 minutes in an Oscillating Mill MM400 (Retsch GmbH, Haan, Germany) to pulverize the tissue. DNA extraction was subsequently conducted using either the DNeasy ® Plant Mini Kit DNA (Qiagen) or the Genomic DNA Mini Kit (Plant) (DNAture). DNA extraction followed each manufacturer's protocol, but with a modification of a longer incubation times of up to one hour.

Following DNA extraction, DNA quality (purity) and concentration were measured using a Nanodrop ® ND-1000 Spectrophotometer and Qubit 2.0 fluorometer (ThermoFisher Scientific (TF), Waltham, MA, USA). DNA quality (purity) was recorded by documenting the OD260/OD280 and OD260/OD230 ratios. As recommended by the genotyping service provider that was used for this study (Thermo Fisher Scientific), only genomic samples with a OD260/OD280 ratio between 1.8 and 2.0 and a OD260/OD230 ratio greater than 1.5 were considered to be of sufficient quality for genotyping. The DNA samples failed to meet this criteria were reattempted for the extraction with the same method until they make acceptable quality. The DNA concentration of genomic samples measured using a Qubit and selected for genotyping ranged between 3.59-161 ng/ul.

SNP Genotyping and mitigation of ascertainment bias

SNP data were generated for 1536 DNA samples (including 83 sample repeats for reproducibility calculation and samples that were failed to be genotyped) with the *Eucalyptus* 72K Axiom array (TF) Samples of *E. argophloia* were prepared and sent to TF in separate plate from those used for *E. bosistoana* to optimize customized probeset for each species for efficient genotyping that is described in the following paragraph.

Upon arrival at TF, sample Quality Control (QC) was conducted using PicoGreen (TF) analysis at the Applied Biosystems Microarray Research Services Laboratory (MRSL) of TF to verify the DNA concentration of the submitted samples. The probesets of the Euc72K Axiom array were hybridized and eventually scanned. To produce better quality genotyping data for the samples and to mitigate potential ascertainment bias of the array (a systematic deviation caused by the non-random selection of SNPs in genotyping arrays that can lead to inaccurate representation of genetic diversity), new and separate custom genotyping probeset lists were specifically created for our *E. bosistoana* and *E. argophloia* samples. At the end of this process, SNP data extracted from probes on the Euc72K Axiom array were stored as CEL files, data files created by Affymetrix DNA microarray image analysis software. These CEL files were converted to Variant Call Format (VCF) files.

Data Filtering

Removing outliers and sample repeats and subsetting the dataset

Samples in the unfiltered VCF file were relabeled with the sample numbering system used in this study with the BCFtools version 1.9 software (DANECEK *et al.* 2021) that was also used for further data filtering processes. Duplicate samples (83 sample repeats including samples that were failed to be genotyped from the previously submitted plates) were used to test the reproducibility of genotyping and removed from the dataset. Next, 11 outliers defined as having notably different Principal Coordinate Analysis (PCoA) coordinates (GenAlEx 6.5 (PEAKALL AND SMOUSE 2006; PEAKALL AND SMOUSE 2012)) from those of samples belonging to the same putative population were removed. These might be samples that were mislabeled during seed collection, handling, planting, or at some stage during the field and lab work for this project. Only one randomly selected individual from each family was chosen (as multiple individuals belonging to same family included in the original dataset that can cause bias in population genetic structure analysis) to represent the family's genetic makeup across all the examined populations.

Eucalyptus bosistoana families were grouped into 25 subpopulations according to their geographical proximity (defined in Figure 1.8-2). Three separate datasets comprising P1-P25, P3-P25, and P3-P17 & P22-P25, respectively, were used for the analysis of population genetic structure and diversity.

The post genotyping optimization pipeline used in this study is described in Figure 1.

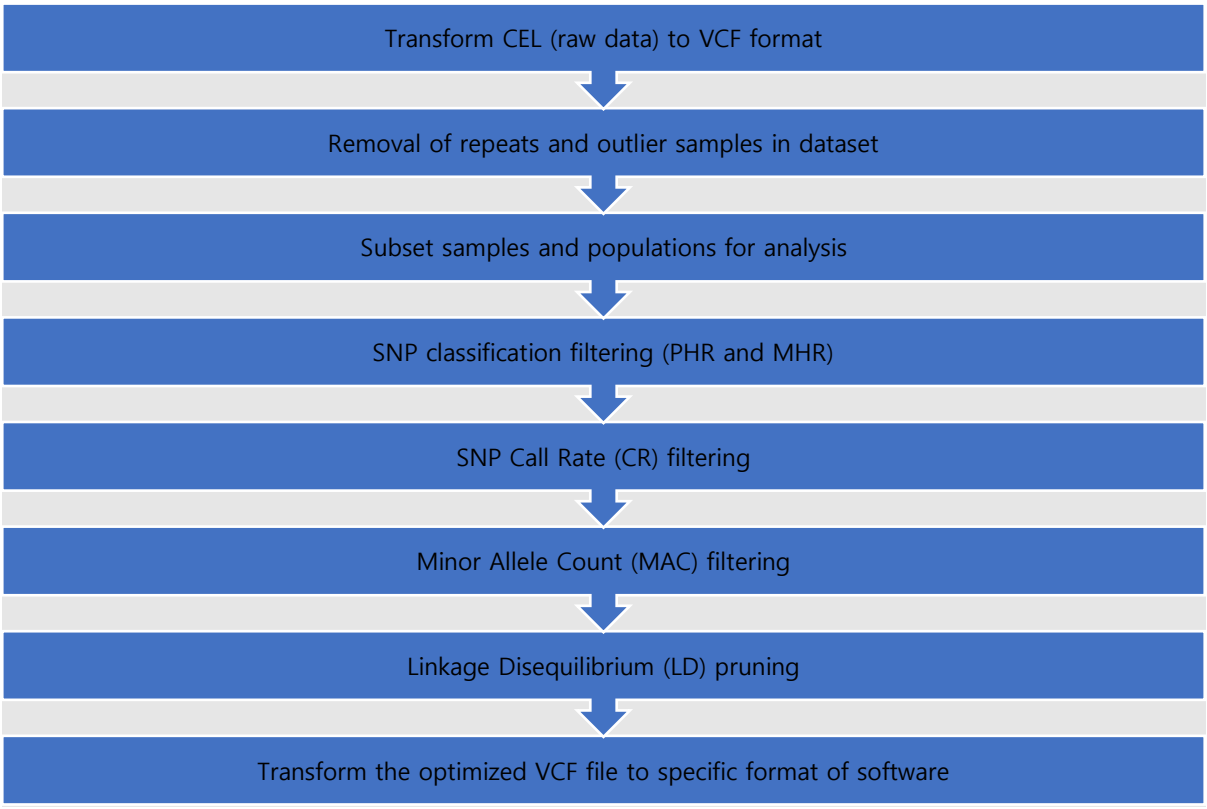


Figure 1 Post genotyping optimization pipeline used in this study. All optimization process was done using BCFtools. Transforming CEL to VCF format was done using Axiom Analysis Suite version 5.1.1.1 (TF). Transforming the optimized VCF file to specific format of software was done using PGDSpider version 2.1.1.5 for STRUCTURE and vcfR2genind function in vcfR package for GenAlEx formats.

Composition of dataset

SNPs of the PolyHighResolution (PHR) category are markers of key interest in this study because those SNPs are polymorphic markers that show variation in DNA which enables the estimation of genetic structure and diversity (DUVAL *et al.* 2023). However, when genetic diversity estimations (e.g. heterozygosity) are based solely on polymorphic markers (PHR), the results are often biased by the overall sample size (N), with smaller N producing higher estimations of heterozygosity (SCHMIDT *et al.* 2021). On the other hand, when both monomorphic (MonoHighResolution; MHR) and polymorphic (PHR) genotype data are taken into account in the calculations, genetic estimates are not affected by the sample size and are less biased. Thus, only PHR SNPs were used in the genetic structure analyses whereas both PHR and MHR SNPs were used in genetic diversity analyses. SNPs identified as Off-Target Variants (OTV), CallRateBelowThreshold (CRBT) and Others were not used for downstream analyses. Although NoMinorHom (NMH) SNPs might be useful but less stringent compared to PHR for a population genetics study, they were not used in the downstream analysis. Table 1 summarizes datasets used and the parameters applied for data filtering of each dataset used in each analysis in this chapter.

Table 1 Summary of datasets & parameters used for each analysis. The first column shows the type of genetic study, and the second column shows the specific analyses used. SNP classification columns indicate what category of SNPs are included in each type of analysis (PHR: PolyHighResolution, MHR: MonoHighResolution). Threshold/Filtering columns indicate the threshold values used for each analysis (CR: Call Rate, MAC: Minor Allele Count, LD: Linkage Disequilibrium, w: window size).

Dataset & Parameters applied for each analysis		SNP classification		Threshold/Filtering			Analysis	
		PHR	**MHR**	**CR**	**MAC**	**LD pruning**	**Specific Software**	**Parameters**
Genetic Structure	**STRUCTURE**	O	X	97.32%	MAC≥3	w-10k, r^2-0.8	STRUCTURE Threader	• Burnin: 100k • MCMC: 100k • Iterations: 20
	PCoA, AMOVA, IBD						GenAlEx	N/A
	(Putative) Outlier			97%	N/A	N/A	pcadapt (R package), GenAlEx	N/A
Genetic Diversity	**Heterozygosity**	O	O	97%	N/A	w-10k, r^2-0.8	GenAlEx	N/A

Data filtering for analyses

For the genetic structure analyses, only SNPs with a Call Rate (CR) over 97.32% were used to reduce the number of SNP loci to a size that can be used in GenAlEx 6.5 (i.e. a maximum of 8,191 codominant SNP loci). The same CR threshold was used for the STRUCTURE analyses to make it easier to compare the results of the various analyses. SNPs with a minimum CR of 97% were used in the genetic diversity analyses and for testing (putative) adaptive units (AUs).

SNP loci with a Minor Allele Count (MAC, the number of minor alleles in a given population) of 1 and 2 were removed from the data sets used for genetic structure analyses following the recommendation of SCHMIDT *et al.* (2021) as including singletons and doubletons can confound model-based inferences of genetic structure (LINCK AND BATTEY 2019). For testing AUs, however, MAC filtering was not applied because the dataset used for this specific analysis consisted of a very small number of SNPs (i.e. 68 outlier loci) and therefore the number of SNP loci was not reduced further.

Linkage Disequilibrium (LD) pruning of the dataset was carried out using BCFtools (DANECEK *et al.* 2021). Different values of r^2 (0.2, 0.4, 0.6, and 0.8) and w (500, 1k, 10k) were used to determine if the results of the genetic structure analyses (i.e. PCoA) would change if different LD filters were used.

Genetic structure analyses

For this research chapter, the 178 breeding families of *E. bosistoana* were each assigned to one of 25 putative populations (P1-P25) based on their geographical proximity and distribution in Australia to help understand patterns of genetic structure among natural populations. A single sample from each 178 *E. bosistoana* family groups was used to represent the genetic pattern of each group.

Genetic structure analyses were only conducted for samples of breeding families identified as *E. bosistoana* because all samples of *E. argophloia* belong to the only known natural population of this species. To infer genetic structure among the 25 putative *E. bosistoana* populations, STRUCTURE analyses were conducted using Structure_threader (PRITCHARD *et al.* 2000; PINA-MARTINS *et al.* 2017). The filtered VCF data files were converted to STRUCTURE format using PGDSpider version 2.1.1.5 (LISCHER AND EXCOFFIER 2012). Structure_threader was run remotely on a High Performance Computing (HPC) System at the University of Canterbury utilizing multi core CPU systems. Population location information was provided using the “indfile” option in Structure_threader using the 25 geographical putative populations. The specific parameters used for this analysis were: burnin-100k and MCMC-100k. Each analysis was run 20 times. The method of EVANNO *et al.* (2005) was used to determine the optimal K value which is defined as the highest Delta K value in the STRUCTURE analysis results. The estimated optimal K value (Delta K) was

further compared with the second-order rate of change in the likelihood of the data (LnP(K)) with respect to K to identify the optimal K value where there is a clear "elbow" in the curve (ROSENBERG *et al.* 2002) (Appendix 1). Graphs of both Delta K and LnP(K) estimates were plotted using Microsoft Excel. Finally, membership probability plots for each K value of interest were generated using “plot” mode in Structure_threader (PINA-MARTINS *et al.* 2017).

In addition to STRUCTURE analyses, PCoA was used to study genetic structure. This was done using GenAlEx 6.5 (PEAKALL AND SMOUSE 2006; PEAKALL AND SMOUSE 2012). To convert the VCF file to a GenAlEx format, the *vcfR2genind* function in the vcfR package was used to load VCF file as a genind object in RStudio version 1.4.1106 (KNAUS AND GRUNWALD 2017; R CORE TEAM 2021; RSTUDIO TEAM 2021). Markers with no scored alleles were removed when the VCF files were transformed to files in GenAlEx format using the vcfR package in R. The genind object was then reformatted for GenAlEx using the *genind2genalex* function in the poppr package (KAMVAR *et al.* 2014; KAMVAR *et al.* 2015). Grouping of individuals from *E. bosistoana* populations in PCoA analysis was made based on the genetic structure pattern of the species identified by STRUCTURE analysis (Figs. 2-4).

To further infer genetic differentiation among populations, Analysis of Molecular Variance Analysis (AMOVA) was conducted with 999 permutations in GenAlEx 6.5. A species-level Fst estimate and pairwise population Fst values were calculated. The *p*-values of these pairwise Fst values were corrected using B-Y correction to avoid Type Ⅰ and Type Ⅱ errors (NARUM 2006).

Isolation by distance (IBD) analysis was conducted to detect a possible correlation between geographic and genetic distances among populations. GPS coordinates of sourced *E. bosistoana* individuals in Australia were acquired from the NZDFI database. A Nei genetic distance matrix and a logarithmic geographic distance matrix were generated in GenAlEx 6.5. Mantel tests with 999 permutations were performed with these matrices to test for IBD (MANTEL 1967).

P3-P17 and P22-P25 *pcadapt* package (LUU *et al.* 2017) in RStudio version 1.4.1106 (RSTUDIO TEAM 2021). Candidate adaptive outlier loci were identified using a Bonferroni correction (i.e. returning a p-value lower than 0.01/number of tests) to correct for multiple testing (CAPBLANCQ AND FORESTER 2021). PCoA analyses were performed separately for outlier and neutral loci in GenAlEx 6.5 and their results were compared with each other and those obtained for the data set that contained all loci.

Data analysis for genetic diversity

To quantify the genetic diversity of each population, the percentage of allelic richness (N_a), number of effective alleles (N_e), percentage of polymorphic loci (P), observed heterozygosity (H_o), expected heterozygosity (H_e), and inbreeding coefficient (F_{is}) were calculated in GenAlEx 6.5.

RESULTS

SNP Genotyping result and quality check

In total, 1379 out of 1433 DNA samples (including 81 sample repeats, 1 sample omitted out of 1434 total submitted genomic samples during genotyping process from TF due as it failed to produce a scan) (96.23%) identified as *E. bosistoana* passed the sample QC (Table 2). The average cluster call rate (a metric used to evaluate the quality of genotyping data obtained from Axiom genotyping arrays) was 99.51% and sample reproducibility was 99.82% for these species, as reported by TF. A total of 100 samples out of 102 submitted (including 2 sample repeats) DNA samples (98.04%) of *E. argophloia* passed the QC. The average cluster call rate was 99.67% and sample reproducibility was 99.84% for *E. argophloia*.

Table 2 Genotyping QC summary of samples identified as E. argophloia *and* E. bosistoana.

Sample QC summary	*E. bosistoana*	*E. argophloia*
Total number of samples	1,433	102
Passed samples	1,379	100
Failed samples	54	2
% Passed samples	96.23%	98.04%
% Samples meeting concentration spec. passed	96.29%	97.96%
Average Cluster Call Rate	99.51%	99.67%
Sample Reproducibility	99.82%	99.84%

Table 3 shows the number and percentage of loci for each SNP classification category (i.e. PHR, NMH, MHR, CRBT, OTV, and Other). 11,088 (16.29%) loci were of the PHR category and 22,735 (33.41%) were MHR loci for plants labelled as *E. bosistoana*. *Eucalyptus argophloia* had 7,234 (10.63%) PHR and 34,285 (50.38%) MHR loci.

Table 3 SNP QC summary of genotyping samples identified as E. argophloia *and* E. bosistoana

	E. bosistoana		*E. argophloia*	
	# of markers	% of markers	# of markers	% of markers
PolyHighResolution (PHR)	11,088	16.29	7,234	10.63
NoMinorHom (NMH)	9,799	14.40	6,299	9.26
MonoHighResolution (MHR)	22,735	33.41	34,285	50.38
CallRateBelowThreshold (CRBT)	2,226	3.27	1,731	2.54
OffTargetVariant (OTV)	2,633	3.87	3,076	4.52
Other	19,574	28.76	15,430	22.67
Total	**68,055**	**100.00**	**68,055**	**100**

Optimization of SNP dataset

Analyses to determine appropriate settings for LD pruning showed that increasing the window size did not have much impact on the number of SNPs retained after filtering, whereas higher values of r^2 resulted in a substantial reduction of number of SNPs (Table 4). Regardless, genetic structure analyses using various combinations of r^2 and window size values resulted in very similar genetic structure patterns (results not shown). Thus, a window size of 10k was chosen as it is the most conservative value tested and an r^2 value of 0.8 was decided to retain a large number of SNPs for the analyses.

Table 4 Number of removed PHR SNPs for different values of r^2 *and Window Size (bp – base pair) in LD pruning using BCFtools. Parameter values in underlined bold characters were chosen for the final dataset in this study. The dataset used for this test contained PHR filtered samples form the* E. bosistoana *populations with as CR97.32%, having 8154 SNPs.*

Window Size (bp)	LD r^2	Removed # SNPs
500	0.2	3418
	0.4	1876
	0.6	1276
	0.8	1023
1000	0.2	3773
	0.4	2057
	0.6	1370
	0.8	1108
10000	0.2	3790
	0.4	2064
	0.6	1372
	0.8	1110

The dataset for *E. bosistoana* consisted of 8154 PHR SNPs after filtering by CR and MAC for the genetic structure analyses. After subsequent LD pruning and excluding loci with non-scored alleles, the number of SNPs in the final dataset for analysis further decreased to 6895 SNPs. For the genetic diversity study of the populations identified as *E. bosistoana*, 7106 SNPs (incl. 12 MHR SNPs) were used, while the final dataset of *E. argophloia* consisted of 4645 SNPs (incl. 12 MHR SNPs).

Patterns of genetic structure of *E. bosistoana* populations

STRUCTURE

To understand the patterns of genetic structure among putative populations labelled as *E. bosistoana* in the NZDFI breeding population (P1-P25) STRUCTURE analyses were carried out. The first run of the analysis used a dataset consisting of samples from all populations (P1-P25). The Delta K value for this data set was highest at K=2, indicating that the genetic variation of breeding families labelled as *E. bosistoana* is best structured into two genetic clusters (Appendix 1-A). Individuals of populations P1 and P2 were placed with high probability in one cluster, whereas P3-17 and P22-25 were assigned to the other cluster (Figure 2). P18-P21 showed evidence of admixture, which is strongest for P20-P21.

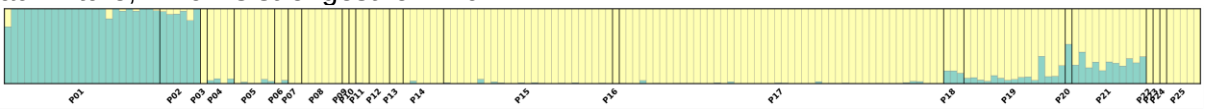


Figure 2 Population structure plot when K=2 of *E. bosistoana* populations of P1-P25.

Substructure within the P3-P25 cluster was studied in a second STRUCTURE analysis. The Delta K value for this data set was highest at K=2 (Appendix 1-B). The pattern obtained suggested that individuals from P3-P17 and P22-P25 are assigned with high probability to one cluster and those of P20 and P21 and two samples of P19 to the other cluster (Figure 3). P18 and P19 show evidence of admixture.

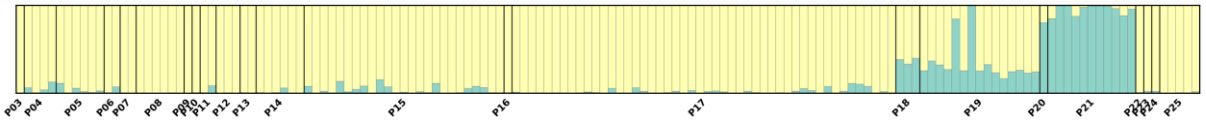


Figure 3 Population structure plot when K=2 of *E. bosistoana* populations of P3-P25.

Finally, to determine the genetic substructure of the P3-25 populations that did not display the admixture observed between the P1-P2 and the P3-17 and P22-25 clusters, another STRUCTURE analysis was performed using only the latter set of populations. The genetic structure was best explained with K=3 (Appendix 1-C). This revealed groups of populations with various levels of admixture (Figure 4).

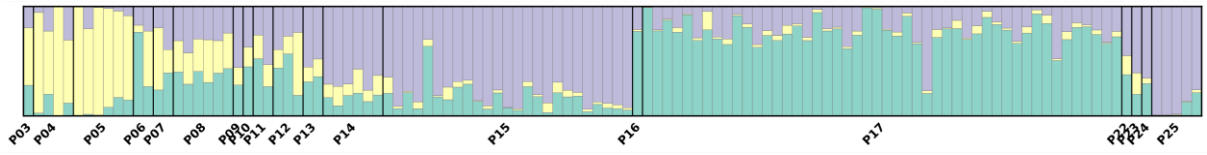


Figure 4 Population structure plot when K=3 of *E. bosistoana* populations of P3-P17 and P22-25.

PCoA

To further explore the patterns of genetic structure as revealed by STRUCTURE, PCoA analyses were conducted with the three previously mentioned STRUCTURE datasets (e.g. P1-P25, P3-P25, P3-P17 & P22-25). The PCoA patterns (Figs. 5 – 7) were congruent with those identified by STRUCTURE for the same individuals (Figs. 2 – 4).

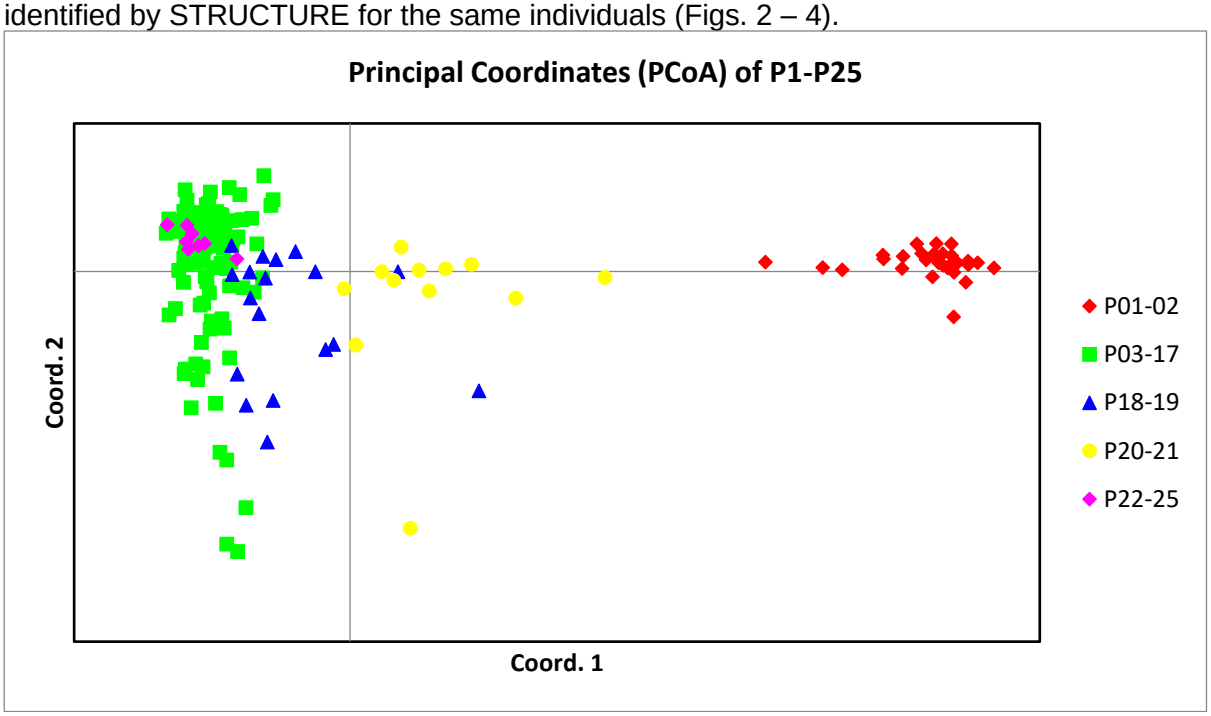


Figure 5 PCoA plot of P1-P25 *E. bosistoana* populations generated by using GenAlEx 6.5. Percentage of variation explained by the axes were: Cumulative-8.74% Coord1-7.22%; Coord2-1.52%.

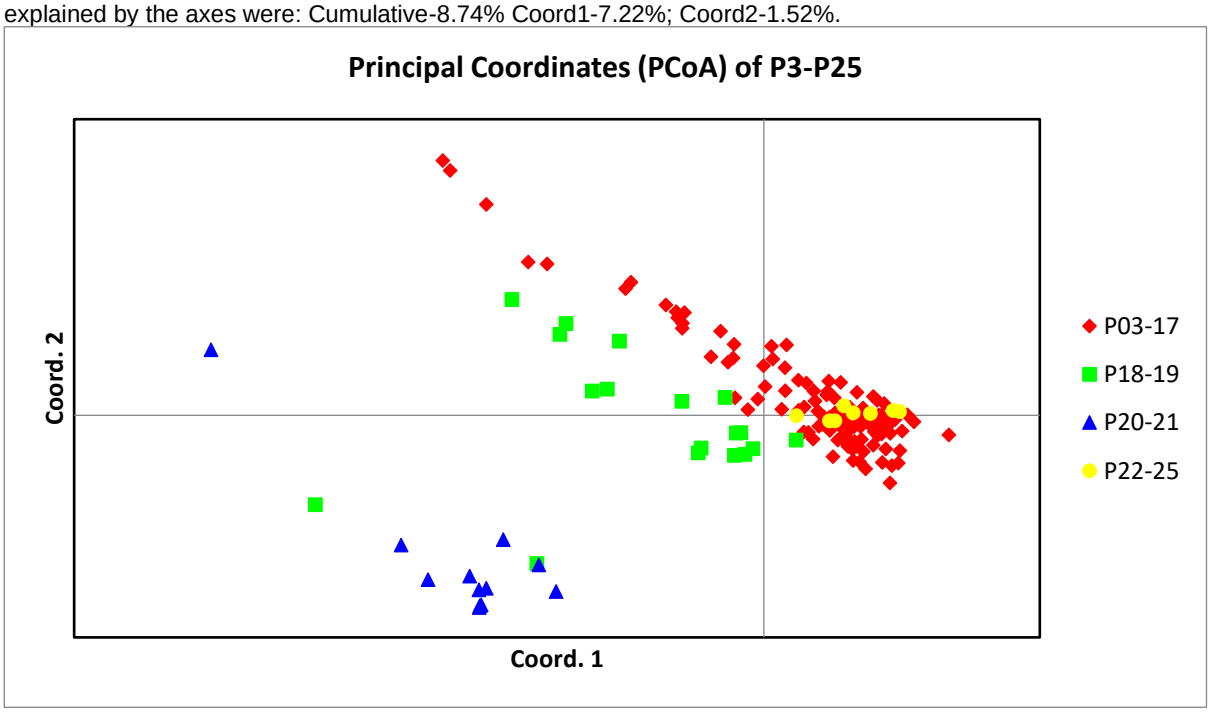


Figure 6 PCoA plot of P3-P25 *E. bosistoana* populations generated by using GenAlEx 6.5. Percentage of variation explained by the axes were: Cumulative-3.88% Coord1-2.03%; Coord2-1.85%.

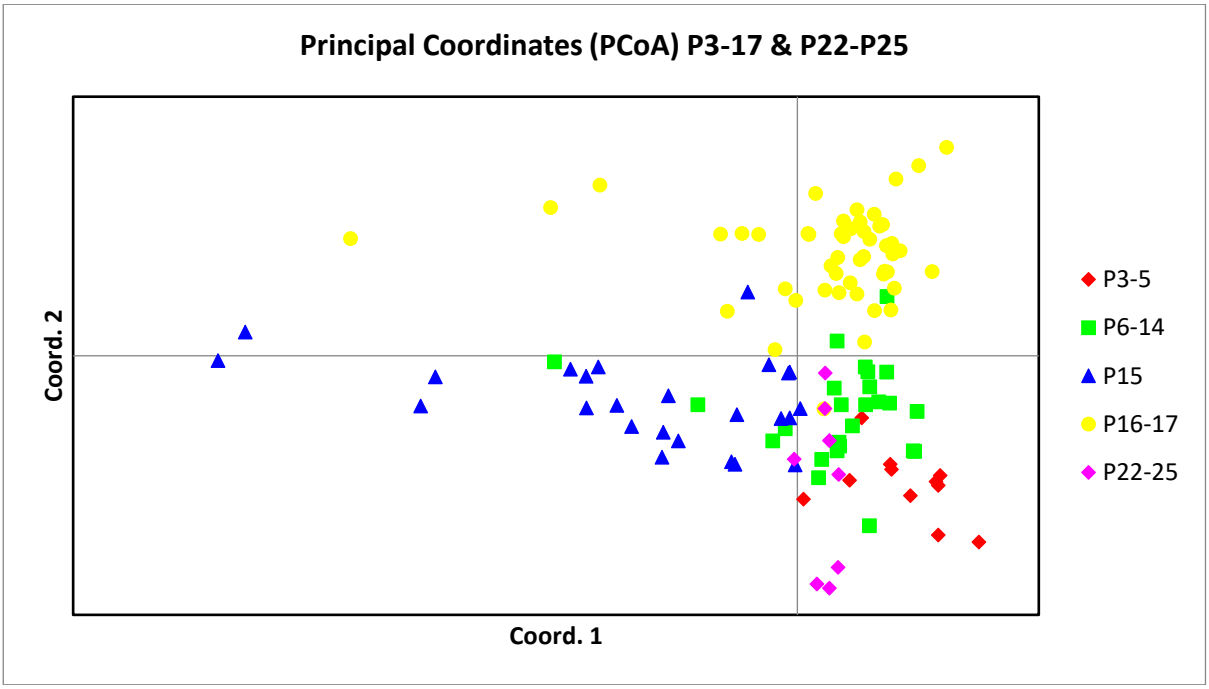


Figure 7 PCoA plot of P3-P17&P22-25 *E. bosistoana* populations generated by using GenAlEx 6.5. Percentage of variation explained by the axes were: Cumulative-3.84% Coord1-2.14%; Coord2-1.69%.

AMOVA

In an AMOVA of the P1-P25 data set, 9% of the variation was found among the five different population groups recognized in the STRUCTURE and PCoA analyses and 38% was among individuals. The overall Fst (0.126) was statistically significant at p=0.001 and indicates moderate genetic differentiation. Pairwise Fst values showed statistically significant differentiation for all group pairs (Table 5). These values show the greatest genetic differentiation between P1-P2 and the other groups (0.110-0.175). The pairwise Fst values for the other pairs of population groups were relatively low.

Table 5 Pairwise Fst values for *E. bosistoana* between P1-P2, P3-P17, P18-P19, P20-P21 and P22-P25. All values are significant at p=0.0219 after B-Y method correction (NARUM 2006).

P1-P2	P3-P17	P18-19	P20-21	
0.158				P3-17
0.139	0.019			P18-19
0.110	0.049	0.040		P20-21
0.175	0.019	0.039	0.069	P22-25

The result of the AMOVA with the data set that only consisted of samples from populations P3-P17 and P22-P25 revealed that 2% of the variation was found among population groups and that were 41% among individuals. For this analysis five population groups were recognized based on the STRUCTURE (Figure 4) and PCoA results (Figure 7). P22-P25 was treated as distinct from P15 because of the geographic distance between them. The overall Fst value (0.026) was significant at p=0.001 and may be considered as indicating little overall genetic differentiation for this subset of population groups. Pairwise Fst values between P15 and P16-P17, and between P6-P14 and all the other groups were not significant after B-Y correction (Table 6). The highest pairwise Fst was found between the P3-P5 and P22-P25 groups. The lowest pairwise Fst was found between P15 and P22-P25.

Table 6 Pairwise Fst values for *E. bosistoana* between five groups of P3-P5, P6-P14, P15, P16-P17 and P22-P25. *Significant pairwise comparisons after B-Y correction, otherwise all significant (NARUM 2006).

P3-P5	P6-P14	P15	P16-P17	
0.020				P6-P14
0.036*	0.014			P15
0.030*	0.010	0.017		P16-P17
0.047*	0.018	0.025*	0.027*	P22-P25

Isolation by distance (IBD)

To test for isolation by distance among the P3-P17 & P22-P25 populations, a Mantel test was performed. This showed a significant positive correlation (r= 0.344, p=0.01) between genetic and geographic distance suggesting isolation by distance (Figure 8).

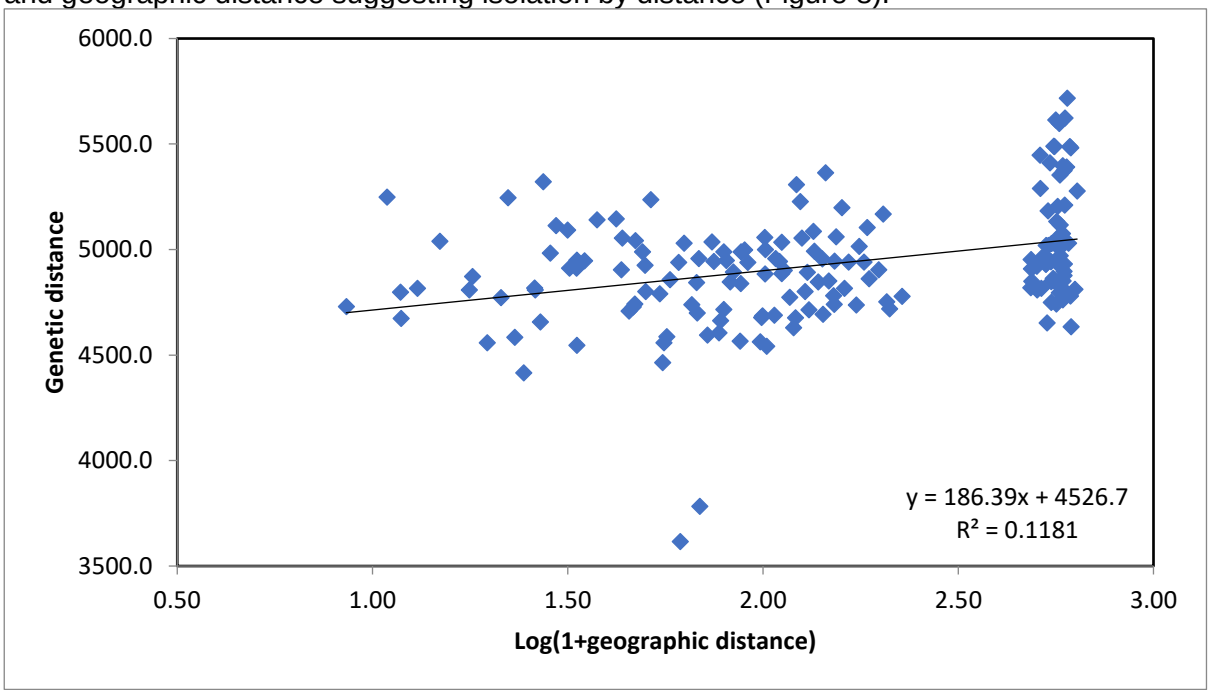


Figure 8 Mantel correlation between pairwise genetic distances of *E. bosistoana* individuals from populations P3-P17&P22-P25 and log (1+ geographic distances). The R value of 0.344 was significant (p=0.01).

Additional Mantel tests were also performed for population groups P3-P17 and P22-P25 separately to obtain more information about the geographical scale at which isolation by distance is present (Figures 9 and 10). The p-values of each test were not significant (0.170 for P3-P17 and 0.250 for P22-P25), meaning that no isolation by distance was found for these datasets.

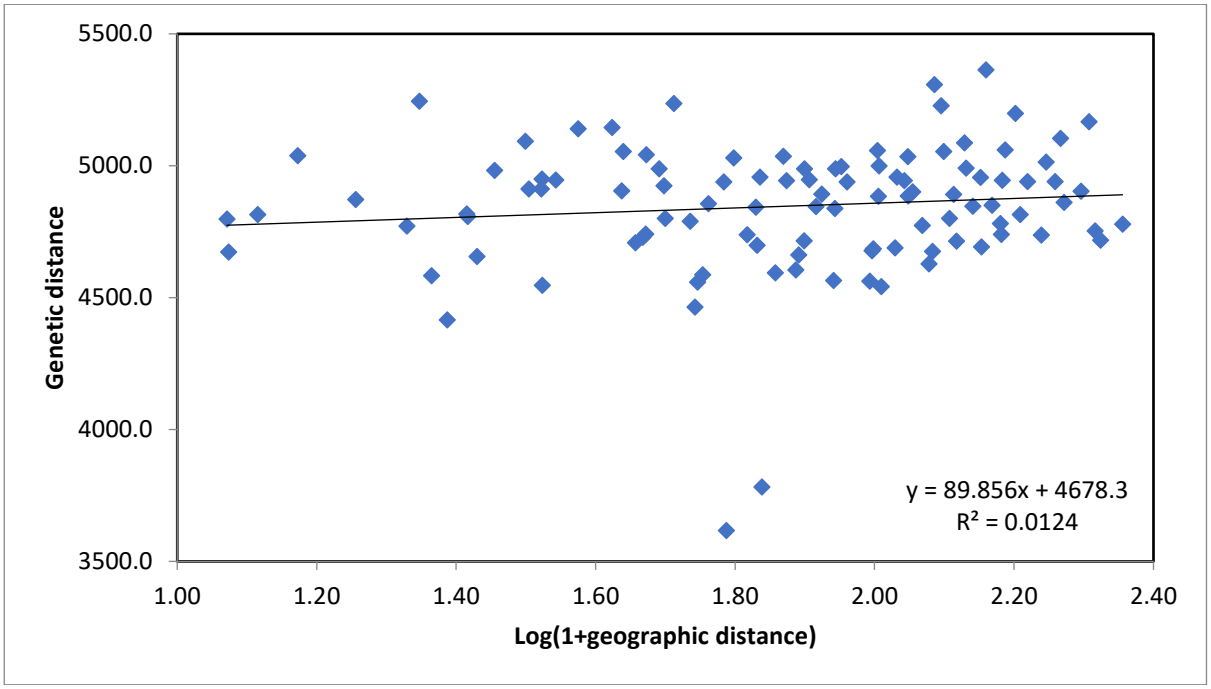


Figure 9 Mantel correlation between pairwise genetic distances of *E. bosistoana* individuals from populations P3-P17 and log (1+ geographic distances). The R value of 0.111 was not significant (p=0.170).

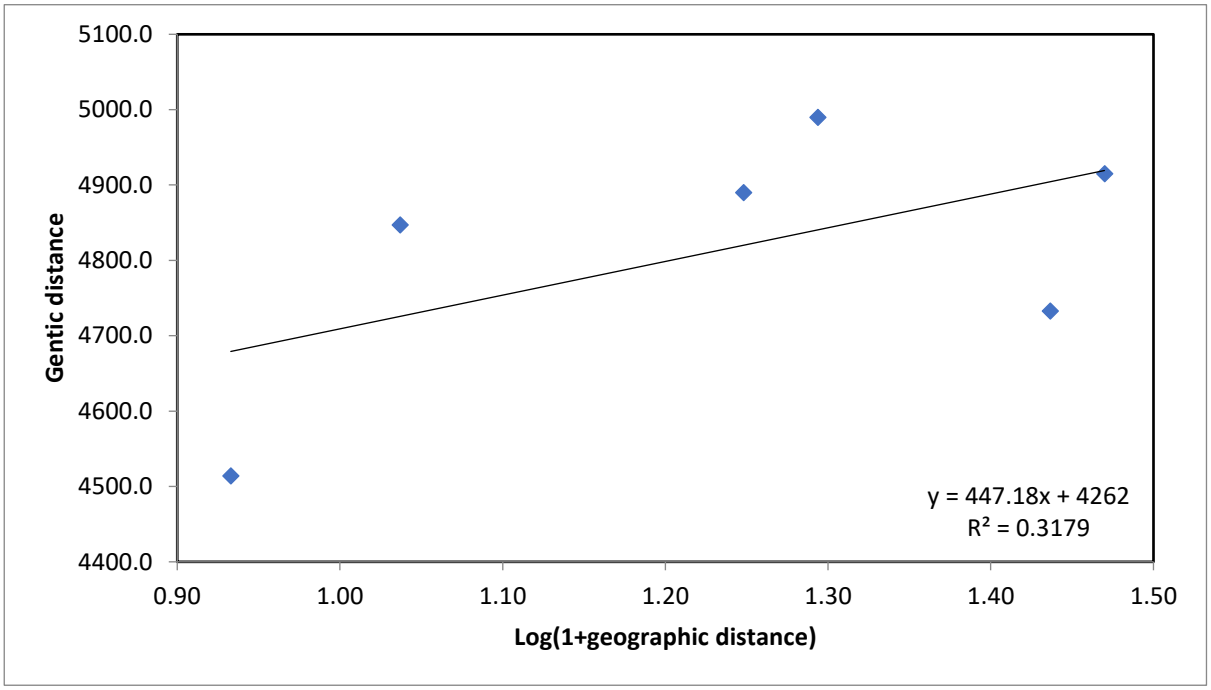


Figure 10 Mantel correlation between pairwise genetic distances of *E. bosistoana* individuals from populations P22-P25 and log (1+ geographic distances). The R value of 0.564 was not significant (p=0.250).

Genetic adaptation

Individuals from population P3-P17 & P22-P25 were tested for outliers to identify possible genetic adaptaion. Using the *pcadapt* package, 68 outlier loci were identified after the Bonferroni correction. PCoA analyses of these 68 putative adaptive loci and the dataset composed of the 6867 non-outlier (putative neutral) loci are shown in Figure 11. In general, patterns of genetic structure of each dataset were similar to each other and to those obtained for the entire dataset (Figs. 2-4).

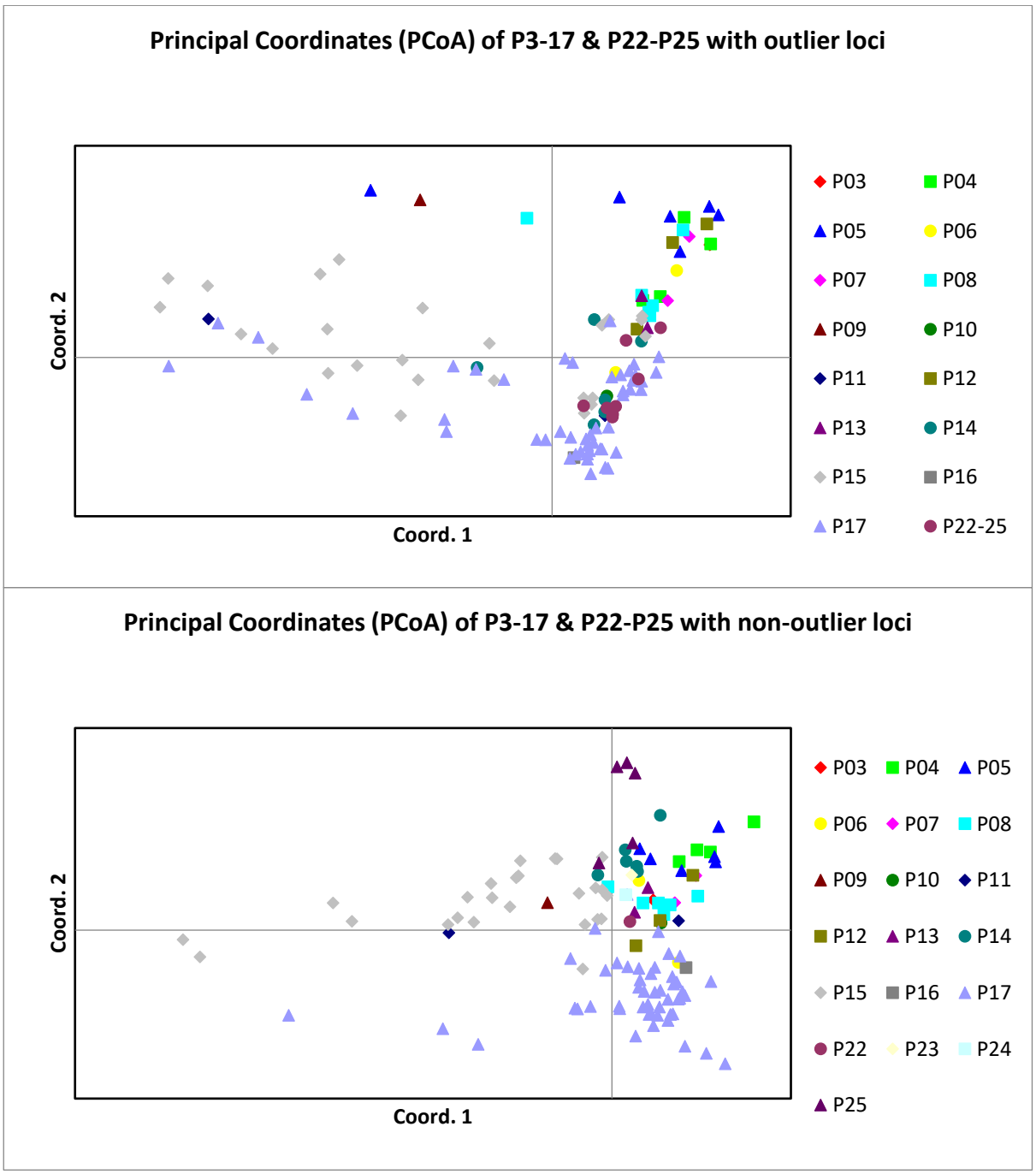


Figure 11 PCoA plots of P3-P17 & P22-25 *E. bosistoana* populations with 68 outlier (top) and 6867 non-outlier (bottom) loci. Percentage of variation explained by the axes were: Cumulative-30.47%, Coord1-18.48%, Coord2-11.99% (top); Cumulative-3.73%, Coord1-2.07%, Coord2-1.67% (bottom).

Patterns of genetic diversity

To understand the patterns of genetic diversity, the allelic richness (N_a), number of effective alleles (N_e), percentage of polymorphic loci (P), observed heterozygosity (H_o), expected heterozygosity (H_e), and inbreeding coefficient (F_{is}) were calculated. The genetic diversity patterns of populations that were labelled as *E. argophloia* and *E. bosistoana* in the NZDFI breeding programme are compared in Table 7 The estimated values of N_a, P, N_e, H_o, H_e, and F_{IS} were higher in *E. argophloia* compared to populations of *E. bosistoana*.

Samples labelled as *E. bosistoana* were grouped into six subpopulations (i.e. P1-P2, P3-P5, P6-P14, P15, P16-P17, and P22-P25) based on the genetic differentiation analysis (Tables 5 and 6). Overall, groups with small number of samples (e.g. P3-P5 and P22-P25) got lower N_a, P, H_e, and F_{IS}. However, estimation of those indices in population groups with high number of

samples (more than 25 samples) of P1-P2, P6-P14, P15, and P16-P17 showed similar values in N_a, N_e, H_o, and H_e. The values of N_a and P were the highest in the group of P16-P17 where the number of samples were also the highest.

Table 7 Standard genetic diversity indices for species of E. argophloia *and plants identified as* E. bosistoana *that used single sample per family, and for populations of* E. bosistoana *in 6 clustering groups. Column labels: # SNPs – number of SNPs used for analysis, # Samples - number of samples included in the populations group, N_a - allelic richness corrected for sample size, N_e - number of effective alleles, P - percentage of polymorphic loci, H_o - observed heterozygosity, H_e - expected heterozygosity, and F_{IS} - inbreeding coefficient. Standard error values are in parenthesis.*

Species and populations	**# SNPs**	**# Samples**	**N_a**	**N_e**	**P**	**H_o**	**H_e**	**F_{IS}**
E. argophloia	4645 (4633 PHR, 12 MHR)	33	1.947 (0.003)	1.460 (0.005)	94.73%	0.195 (0.002)	0.279 (0.002)	0.320 (0.006)
***E. bosistoana* populations**	7106 (7094 PHR, 12 MHR)	**# Samples**	**N_a**	**N_e**	**P**	**H_o**	**H_e**	**F_{IS}**
P1-P2		29	1.809 (0.005)	1.346 (0.004)	80.93%	0.156 (0.002)	0.214 (0.002)	0.239 (0.005)
P3-P5		11	1.674 (0.006)	1.318 (0.004)	67.39%	0.151 (0.002)	0.196 (0.002)	0.188 (0.005)
P6-P14		36	1.829 (0.004)	1.330 (0.004)	82.90%	0.151 (0.002)	0.207 (0.002)	0.231 (0.005)
P15		25	1.814 (0.005)	1.331 (0.004)	81.40%	0.155 (0.002)	0.207 (0.002)	0.206 (0.005)
P16-P17		49	1.885 (0.004)	1.333 (0.004)	88.47%	0.153 (0.002)	0.209 (0.002)	0.231 (0.004)
P22-P25		8	1.602 (0.006)	1.304 (0.004)	60.34%	0.154 (0.002)	0.186 (0.002)	0.138 (0.006)

DISCUSSION

Eucalyptus argophloia and *E. bosistoana* are target breeding species of the NZDFI for the establishment of a durable hardwood plantation resource in New Zealand. Understanding the genetic structure and diversity of their populations has implications for their breeding strategy and can also be used for conservation management of these species (PORTH AND EL-KASSABY 2014). This study examined the genetic structure and diversity of *E. argophloia* and *E. bosistoana* populations across its native geographic range, using dataset of SNP markers generated with Eucalyptus 72K Axiom array 72k (TF).

Breeding programs can inadvertently reduce genetic diversity if only a small number of individuals or varieties are used as parents (SWARUP *et al.* 2021). This can lead to inbreeding depression and reduced adaptability in the population. Knowing the genetic diversity of the species, breeders can identify individuals or populations with unique or rare alleles that can help to maintain genetic diversity in the population (RESTOUX *et al.* 2022). Studying the genetic structure of a species also can inform tree breeding programs by identifying patterns of gene flow and population structure (CHARLES 1986). This information can be used to optimize breeding strategies, such as choosing parents from different geographic regions or populations to increase genetic diversity in the offspring.

Quality of genotype data and its optimization

The average cluster call rate (CR) for the genotyping data was high (as threshold CR is 97%), with over 99.5% of SNPs successfully called for all species in this study (Table 2). The genotyping results were highly reproducible (higher than 99.8%), with consistent results across multiple replicates. More than 11k and 7k polymorphic SNP markers (PHR) were obtained for samples labelled as *E. bosistoana* and *E. argophloia*, respectively (Table 3). Polymorphic markers (PHR) provide more genetic variation than monomorphic markers (MHR), as they have multiple

alleles with varying frequencies in the population. This increased variability is beneficial for genetic studies (WELLMANN AND BENNEWITZ 2019). It was reported that more than 65 random SNPs loci are required to detect population structure (TURAKULOV AND EASTEAL 2003). This requirement was easily met with the thousands of informative SNPs used for the population genetics study of *E. bosistoana* and *E. argophloia*. These SNP data were of high quality as they provided a significant amount of genetic variation and precision.

Identification of *E. melliodora* among *E. bosistoana* labelled populations

The population genetic structure pattern of P1-P2 were clearly separated from other *E. bosistoana* populations (P3-P25) (Figure 2 and 3). Investigation of voucher specimens of these genetically differing populations (i.e. P1 and P2) by Pieter Pelser revealed that they are *E. melliodora* A.Cunn.*,* also commonly known as yellow box, honey box, or yellow ironbark. As *E. bosistoana* and *E. argophloia*, this species is also belonging to section Adnataria and thereby having morphological similarity (THORNHILL *et al.* 2019).

The previously believed taxonomically confusing individuals from *E. bosistoana* labelled families in the NZDFI trials were part of P1 and P2. Therefore, the taxonomic confusion resulted from misidentifying some trees from which seed for the NZDFI breeding programme was collected due to their morphological similarity.

Evidence of hybridization between *E. bosistoana* and *E. melliodora*

Some of the *E. bosistoana* populations (P18-P21) showed evidence of hybridization with *E. melliodora* (P1 and P2) to various proportions (Figure 2 and 3). Interspecific hybridization within Eucalyptus is known to occur in Australia and is a common phenomenon within the genus (GRIFFIN *et al.* 1988). This is facilitated by many *Eucalyptus* species have overlapping ranges and grow in close proximity to each other, providing opportunities for hybridization to occur through natural cross-pollination. Habitats of *E. melliodora* indeed overlaps the area of *E. bosistoana* populations including P18-P21.

Population genetic structure of *E. bosistoana*

This is the first population genetics study of natural populations of *E. bosistoana*. The genetic structure analyses conducted using STRUCTURE, PCoA, and AMOVA revealed a common thread of connection among the natural populations of *E. bosistoana* (P3-P17 and P22-P25), suggesting that the observed patterns are robust and reliable. Cluster groups of P3-P5, P16-P17, and P25 showed a relatively clear genetic patterns with one large genetic component for each group in STRUCTURE analysis (Figure 4). Populations groups P6-P14, P15, and P22-P24 were of admixed genetic compositions of different proportions. PCoA analysis also showed that many individuals in those genetically weak-structured groups are all mixed into the rim of other relatively clear clusters of P3-P5, P16-P17, and P22-P25 (Figure 7).

The Fst values in AMOVA analysis were statistically significant between the groups with one large genetic component (i.e. P3-P5, P16-P17, and P25), while they were not significant for the groups with relatively equally dividend genetic components (i.e. P6-P14 and P15) (Table 6). This suggested no genetic differentiation for the *E. bosistoana* groups with mixed genetic components and, as the Fst values were less than 0.5, weak genetic differentiation for the groups with a dominant genetic component.

It is not uncommon to observe weak genetic differentiation within distant natural populations of Eucalyptus species (VON TAKACH DUKAI *et al.* 2019). For example, studies have found low levels of genetic differentiation between populations of *E. tricarpa* (ANDREW et al. 2010), *E. urophylla* (LU et al. 2018), and *E. camaldulensis* (DILLON *et al.* 2014). This is because many Eucalyptus species often have high levels of gene flow and low levels of genetic structuring, which can result in weak genetic differentiation among populations. Significant IBD was observed between two population clusters of P1-P17 and P22-P25 (Figure 8), although no IBD was present within each of these clusters (Figure 9 and Figure 10). This would be consistent with an environmental barrier that limits gene flow between these clusters (e.g. long geographic distance)

but not within each cluster leading to genetic similarity that is not dependent on geographic distance. Also, there was no evidence of genetic AU (putative outlier SNPs) found among all *E. bosistoana* populations (Figure 11).

Long-distance seed dispersal is often a factor that can account for the weak genetic differentiation observed among natural populations of trees (MARTINEZ-LOPEZ *et al.* 2020). Samples from P3-P17 of natural *E. bosistoana* populations were sourced from a long and continuous geographic range in the South-Eastern coast of Australia. There is limited information available about the extent of long-distance seed dispersal in *E. bosistoana*. While it is still possible that *E. bosistoana* may have mechanisms of seed dispersal that can facilitate movement over large distances, for example, by lightweight seeds that may be dispersed by wind, the distance of seed dispersal of Eucalyptus species are often limited within a kilometer (BARBOUR *et al.* 2005). Given the significant geographic distances among the populations and the low probability of rare events occurring in a large number of individuals included in this study (148 samples), it is less likely that long-distance seed dispersal is a major contributor to the weak genetic differentiation observed.

Some genetic variation observed in multiple clusters of the populations P3-P17 can be explained by the fact that the habitat area of populations P3-P14 is located in severely fragmented lowland grassy woodlands, which have experienced extensive grazing, clearing, and fires that could have limited gene flow (TOZER *et al.* 2010).

The boundaries of populations matching the genetic structure are illustrated Figure 12.

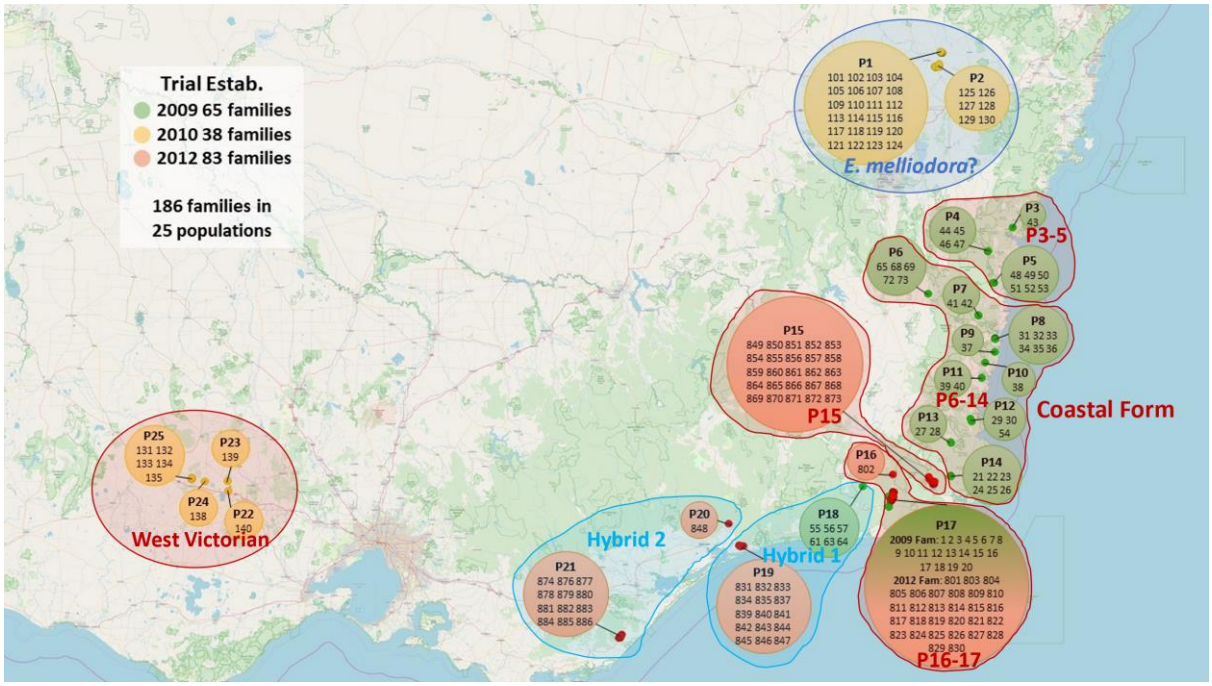


Figure 12 Population genetic structure of natural populations of E. bosistoana *based on STRUCTURE, PCoA, and AMOVA analyses conducted in this research chapter.*

Genetic diversity patterns of durable Eucalyptus

In general, genetic diversity measurements of durable Eucalyptus species made in this study using Euc72K Axiom array showed significantly lower estimates than other population genetics studies of Eucalyptus. As an example, the analysis showed that the average level of genetic diversity in H_e was 0.214 for *E. melliodora* in this study, whereas a study using microsatellite data found a much higher average of 0.62 (BROADHURST *et al.* 2015). The primary reason for the lower estimates of genetic diversity using SNP loci compared to microsatellite loci was due to the SNP loci having lower levels of polymorphism and heterozygosity at the population level (YANG *et al.* 2020). TELFER *et al.* (2015) thoroughly investigated the differences in estimates of genetic diversity using different markers in Eucalyptus. For example, their analysis showed that the mean level of H_e in *E. nitens* populations using SNP markers generated from EuCHIP60K ranged from 0.494 to 0.499, whereas microsatellite markers yielded a much higher range of 0.584 to 0.737.

The durable Eucalyptus species included in the genetic diversity study showed lower H_o than H_e, suggesting a deficiency of heterozygotes in the species. This can be caused by inbreeding or genetic drift, reducing genetic diversity (Table 7). This might be also a consequence of the mating system (not part of this chapter). When closely related individuals mate, the frequency of homozygotes (individuals with two copies of the same allele) can increase and the frequency of heterozygotes decrease. Genetic drift can occur when random fluctuations in allele frequency result in some alleles becoming more common while others become less common, which can also reduce heterozygosity over time (WILLOUGHBY *et al.* 2015).

E. argophloia is a species that is native to a very limited region in Chinchilla, Queensland, with only one known wild population. This population is considered 'vulnerable' under the Queensland Nature Conservation Act (1992), due to its small size and limited geographic range. Both *E. bosistoana* and *E. melliodora* are key species in the endangered Box-Gum Lowland Grassy Woodland ecosystem of southeastern Australia due to habitat loss and fragmentation. The decline in heterozygosity within the genome of the species facing critical ecological issues may be attributed to factors such as inbreeding and genetic drift (HOHENLOHE *et al.* 2021).

The higher F_{is} observed in *E. argophloia* compared to other species could be primarily attributed to the species' small population size. Both estimates of N_e and heterozygosity (H_o and H_e) were also higher in *E. argophloia*, suggesting that the species has a higher level of genetic diversity. This could be due to a variety of factors, such as historical population size, patterns of migration or gene flow, or selection (KUANG *et al.* 2020).

Most of genetic diversity estimates were similar in the 5 cluster groups of *E. bosistoana* (Table 7). However, estimates of N_a and P and were significantly lower in groups of P3-P5 and P22-P25. This might be due to the small number of samples included in the analysis. With a smaller sample size, there may be a lower probability of detecting rare alleles, which could lead to an underestimation of the total number of alleles present in the population (SANCHEZ-MONTES *et al.* 2017).

Based on the observed values of N_e and H_o in the analysis, it is reasonable to conclude that all clusters of populations have comparable levels of genetic diversity.

Implications for breeding strategy of *E. bosistoana*

Although it was possible to define boundaries of natural populations of *E. bosistoana* (Figure 12), genetic differentiation was low and population structure was weak among the populations. Low genetic differentiation can limit the potential for genetic improvement through traditional breeding programs, as there may be fewer desirable alleles available for selection (RESTOUX *et al.* 2022). Narrow genetic diversity limits potential of genetic gain in a breeding programme and adaptation potential to changing environmental conditions. It may be possible to broaden the genetic base of the breeding population by incorporating individuals from unknown populations, if they exist, (ALLIER *et al.* 2020) or alternatively hybridization with a related species. The latter was envisaged for the NZDFI *E. bosistoana* breeding programme with *E. argophloia* but unintentionally realized with *E. melliodora*. However, it should be kept in mind that a planned hybrid programme would utilize the best genotypes of the relevant species. This is not possible for *E. melliodora* at the current stage as a systematic collection and phenotyping of the species has not been conducted.

To maximize the benefits of currently identified genetic structure the breeders can select individuals from multiple populations to reduce the risk of inbreeding or loss of genetic diversity, ensuring the long-term viability of the breeding population (RUTKOSKI *et al.* 2022).

Conservation management of *E. bosistoana*

The majority of the *E. bosistoana* populations (P3-P14) are located within the Lowland Grassy Woodland ecosystem in southeastern Australia, which is currently experiencing significant fragmentation and a reduction in ecological connectivity (MANNING *et al.* 2020). While there was indication of inbreeding and genetic drift in the natural populations of the species, possibly due to habitat fragmentation, as reflected by lower estimates of H_o compared to H_e, the low genetic differentiation and weak population structure among populations suggested high level of gene flow occurring between them. Nevertheless, the aggregating habitat fragmentation can continue to

increase genetic drift and erode genetic diversity within populations increasing the risk of extinction due to reduced fitness (SCHLAEPFER *et al.* 2018). Therefore, maintaining the genetic diversity of populations is crucial for effective conservation management.

Maintaining ex-situ gene pools

The natural populations of *E. bosistoana* in Australia are essential gene pools and resources for future genetic selection, but they are susceptible to hybridization, as demonstrated in this study with *E. melliodora*. Many eucalypt species are well known for their hybridizing behaviour between species within subgenera (VAN DIJK *et al.* 2020). Interspecific hybridization in Eucalyptus presents a challenge for conservation management as it can result in genetic contamination, loss of genetic diversity, and the creation of new hybrid species that may displace native species and impact ecosystem functioning (CHAN *et al.* 2019; QUILODRAN *et al.* 2020). Hybridization can also lead to the spread of invasive traits, making it difficult to manage introduced eucalyptus populations (FIELD *et al.* 2011). Conservation efforts require monitoring of hybridization events and the development of strategies to preserve genetic diversity, including controlled breeding programs and the establishment of protected areas to prevent mixing of different eucalyptus species (PFEILSTICKER *et al.* 2021).

Maintaining ex-situ gene pools of populations, such as the breeding populations of NZDFI in New Zealand, supports conservation management. It serves as a safety net for threatened or endangered species, preventing the loss of genetic diversity and ensuring their survival (ELDRIDGE 1990; XIAO *et al.* 2021). Additionally, ex-situ gene pools can act as a source of genetic material for restoration and reintroduction programs, allowing for the enhancement of desired traits and increasing the adaptive potential of populations (VALBUENA-URENA *et al.* 2017).

REFERENCES

Allier, A., S. Teyssedre, C. Lehermeier, L. Moreau and A. Charcosset, 2020 Optimized breeding strategies to harness genetic resources with different performance levels. BMC Genomics 21**:** 349.

Andrew, R. L., I. R. Wallis, C. E. Harwood and W. J. Foley, 2010 Genetic and environmental contributions to variation and population divergence in a broad-spectrum foliar defence of Eucalyptus tricarpa. Ann Bot 105**:** 707-717.

Bairu, M. W., A. B. Amelework and W. G. Coetzer, 2021 Genetic diversity and population structure of six South African Acacia mearnsii breeding populations based on SSR markers. J Plant Res 134**:** 1243-1252.

Barbour, R. C., B. M. Potts and R. E. Vaillancourt, 2005 Pollen dispersal from exotic eucalypt plantations. Conservation Genetics 6**:** 253-257.

Broadhurst, L. M., G. Fifield, B. Vanzella and M. Pickup, 2015 An evaluation of the genetic structure of seed sources and the maintenance of genetic diversity during establishment of two yellow box (Eucalyptus melliodora) seed-production areas. Australian Journal of Botany 63**:** 455-466.

Capblancq, T., and B. R. Forester, 2021 Redundancy analysis: A Swiss Army Knife for landscape genomics. Methods in Ecology and Evolution 12**:** 2298-2309.

Castiglione, S., A. Cicatelli, R. Lupi, G. Patrignani, T. Fossati *et al.*, 2010 Genetic structure and introgression in riparian populations of Populus alba L. Plant Biosystems 144**:** 656-668.

Chan, W. Y., A. A. Hoffmann and M. J. H. van Oppen, 2019 Hybridization as a conservation management tool. Conservation Letters 12**:** e12652.

Chhatre, V. E., T. D. Byram, D. B. Neale, J. L. Wegrzyn and K. V. Krutovsky, 2013 Genetic structure and association mapping of adaptive and selective traits in the east Texas loblolly pine (Pinus taeda L.) breeding populations. Tree Genetics & Genomes 9**:** 1161-1178.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan *et al.*, 2021 Twelve years of SAMtools and BCFtools. Gigascience 10.

Davies, N. T., L. A. Apiolaza and M. Sharma, 2017 Heritability of growth strain in Eucalyptus bosistoana: a Bayesian approach with left-censored data§. New Zealand Journal of Forestry Science 47.

Dillon, S., R. McEvoy, D. S. Baldwin, G. N. Rees, Y. Parsons *et al.*, 2014 Characterisation of adaptive genetic diversity in environmentally contrasted populations of Eucalyptus camaldulensis Dehnh. (river red gum). PLoS One 9**:** e103515.

Duval, H., E. Coindre, S. E. Ramos-Onsins, K. G. Alexiou, M. J. Rubio-Cabetas *et al.*, 2023 Development and Evaluation of an Axiom(TM) 60K SNP Array for Almond (Prunus dulcis). Plants (Basel) 12.

Eldridge, K. G., 1990 CONSERVATION OF FOREST GENETIC RESOURCES WITH PARTICULAR REFERENCE TO EUCALYPTUS SPECIES. The Commonwealth Forestry Review 69**:** 45-53.

Evanno, G., S. Regnaut and J. Goudet, 2005 Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14**:** 2611-2620.

Field, D. L., D. J. Ayre, R. J. Whelan and A. G. Young, 2011 Patterns of hybridization and asymmetrical gene flow in hybrid zones of the rare Eucalyptus aggregata and common E. rubida. Heredity (Edinb) 106**:** 841-853.

Grattapaglia, D., R. E. Vaillancourt, M. Shepherd, B. R. Thumma, W. Foley *et al.*, 2012 Progress in Myrtaceae genetics and genomics: Eucalyptus as the pivotal genus. Tree Genetics & Genomes 8**:** 463-508.

Griffin, A. R., I. P. Burgess and L. Wolf, 1988 Patterns of Natural and Manipulated Hybridisation in the Genus Eucalyptus L'hérit.-1 A Review. Australian Journal of Botany 36.

Kamvar, Z. N., J. C. Brooks and N. J. Grunwald, 2015 Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet 6**:** 208.

Kamvar, Z. N., J. F. Tabima and N. J. Grunwald, 2014 Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2**:** e281.

Knaus, B. J., and N. J. Grunwald, 2017 vcfr: a package to manipulate and visualize variant call format data in R. Mol Ecol Resour 17**:** 44-53.

Li, Y., and C. Altaner, 2018 Predicting extractives content of Eucalyptus bosistoana F. Muell. Heartwood from stem cores by near infrared spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc 198: 78-87.
Linck, E., and C. J. Battey, 2019 Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. Mol Ecol Resour 19**:** 639-647.
Lischer, H. E., and L. Excoffier, 2012 PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28**:** 298-299.
Lu, W., R. J. Arnold, L. Zhang and J. Luo, 2018 Genetic Diversity and Structure through Three Cycles of a Eucalyptus urophylla S.T.Blake Breeding Program. Forests 9**:** 372.
Luu, K., E. Bazin and M. G. Blum, 2017 pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Ecol Resour 17**:** 67-77.
Mantel, N., 1967 The detection of disease clustering and a generalized regression approach. Cancer Res 27**:** 209-220.
Martinez-Lopez, V., C. Garcia, V. Zapata, F. Robledano and P. De la Rua, 2020 Intercontinental long-distance seed dispersal across the Mediterranean Basin explains population genetic structure of a bird-dispersed shrub. Mol Ecol 29**:** 1408-1420.
Mishra, G., A. Garrill and C. M. Altaner, 2019 Bioactivity of ethanol extracts from Eucalyptus bosistoana F. Muell. heartwood. iForest - Biogeosciences and Forestry 12**:** 467-473.
Narum, S. R., 2006 Beyond Bonferroni: Less conservative analyses for conservation genetics. Conservation Genetics 7**:** 783-787.
Nicholas, I., and P. Millen, 2012 Durable eucalypt leaflet series: Eucalyptus bosistoana. NZDFI, Ed.
Peakall, R., and P. E. Smouse, 2012 GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics 28**:** 2537-2539.
Peakall, R. O. D., and P. E. Smouse, 2006 genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6**:** 288-295.
Pfeilsticker, T. R., R. C. Jones, D. A. Steane, P. A. Harrison, R. E. Vaillancourt *et al.*, 2021 Expansion of the rare Eucalyptus risdonii under climate change through hybridization with a closely related species despite hybrid inferiority. Annals of Botany 129**:** 1-14.
Pina-Martins, F., D. N. Silva, J. Fino and O. S. Paulo, 2017 Structure_threader: An improved method for automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU systems. Molecular Ecology Resources.
Pritchard, J. K., M. Stephens and P. Donnelly, 2000 Inference of Population Structure Using Multilocus Genotype Data. Genetics 155**:** 945–959.
Quilodran, C. S., J. I. Montoya-Burgos and M. Currat, 2020 Harmonizing hybridization dissonance in conservation. Commun Biol 3**:** 391.
R Core Team, 2021 R: A language and environment for statistical computing., pp. R Foundation for Statistical Computing, Vienna, Austria.
Restoux, G., X. Rognon, A. Vieaud, D. Guemene, F. Petitjean *et al.*, 2022 Managing genetic diversity in breeding programs of small populations: the case of French local chicken breeds. Genet Sel Evol 54**:** 56.
Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd *et al.*, 2002 Genetic structure of human populations. Science 298**:** 2381-2385.
RStudio Team, 2021 RStudio: Integrated Development Environment for R., pp. RStudio, PBC, Boston, MA.
Rutkoski, J. E., M. R. Krause and M. E. Sorrells, 2022 Breeding Methods: Population Improvement and Selection Methods, pp. 83-96 in *Wheat Improvement: Food Security in a Changing Climate*, edited by M. P. Reynolds and H.-J. Braun. Springer International Publishing, Cham.
Sanchez-Montes, G., A. H. Arino, J. L. Vizmanos, J. Wang and I. Martinez-Solano, 2017 Effects of Sample Size and Full Sibs on Genetic Diversity Characterization: A Case Study of Three Syntopic Iberian Pond-Breeding Amphibians. J Hered 108**:** 535-543.
Schlaepfer, D. R., B. Braschler, H.-P. Rusterholz and B. Baur, 2018 Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. Ecosphere 9**:** e02488.

Schmidt, T. L., M. E. Jasper, A. R. Weeks and A. A. Hoffmann, 2021 Unbiased population heterozygosity estimates from genome-wide sequence data. Methods in Ecology and Evolution 12**:** 1888-1898.

Shwe, S. S., and D. W. M. Leung, 2020 Plant regeneration from Eucalyptus bosistoana callus culture. In Vitro Cellular & Developmental Biology - Plant 56**:** 718-725.

Slee, A., M. I. H. Brooker, S. Duffy and J. West, 2015 EUCLID, pp. Centre for Australian National Biodiversity Researc.

Telfer, E. J., G. T. Stovold, Y. Li, O. B. Silva-Junior, D. G. Grattapaglia *et al.*, 2015 Parentage Reconstruction in Eucalyptus nitens Using SNPs and Microsatellite Markers: A Comparative Analysis of Marker Data Power and Robustness. PLoS One 10**:** e0130601.

Thornhill, A. H., M. D. Crisp, C. Külheim, K. E. Lam, L. A. Nelson *et al.*, 2019 A dated molecular perspective of eucalypt taxonomy, evolution and diversification. Australian Systematic Botany.

Tozer, M., K. Turner, D. Keith, D. Tindall, C. Pennay *et al.*, 2010 Native vegetation of southeast NSW: a revised classification and map for the coast and eastern tablelands. Cunninghamia 11**:** 359-406.

Turakulov, R., and S. Easteal, 2003 Number of SNPS loci needed to detect population structure. Hum Hered 55**:** 37-45.

Valbuena-Urena, E., A. Soler-Membrives, S. Steinfartz, M. Alonso, F. Carbonell *et al.*, 2017 Getting off to a good start? Genetic evaluation of the ex situ conservation project of the Critically Endangered Montseny brook newt (Calotriton arnoldi). PeerJ 5**:** e3447.

van Dijk, K.-j., M. Waycott, J. Quarmby, D. Bickerton, A. H. Thornhill *et al.*, 2020 Genomic Screening Reveals That the Endangered Eucalyptus paludicola (Myrtaceae) Is a Hybrid. Diversity 12**:** 468.

von Takach Dukai, B., C. Jack, J. Borevitz, D. B. Lindenmayer and S. C. Banks, 2019 Pervasive admixture between eucalypt species has consequences for conservation and assisted migration. Evol Appl 12**:** 845-860.

Xiao, Z., T. Yang, S. Wang, X. Wei and M. Jiang, 2021 Exploring the origin and genetic representation of ex situ living collections of five endangered tree species established for 20–35 years. Global Ecology and Conservation 32**:** e01928.

Xu, Y., P. Li, C. Zou, Y. Lu, C. Xie *et al.*, 2017 Enhancing genetic gain in the era of molecular breeding. J Exp Bot 68**:** 2641-2666.

Yang, H., H. Liao, W. Zhang and W. Pan, 2020 Genome-wide assessment of population structure and genetic diversity of Eucalyptus urophylla based on a multi-species single-nucleotide polymorphism chip analysis. Tree Genetics & Genomes 16**:** 1-11.

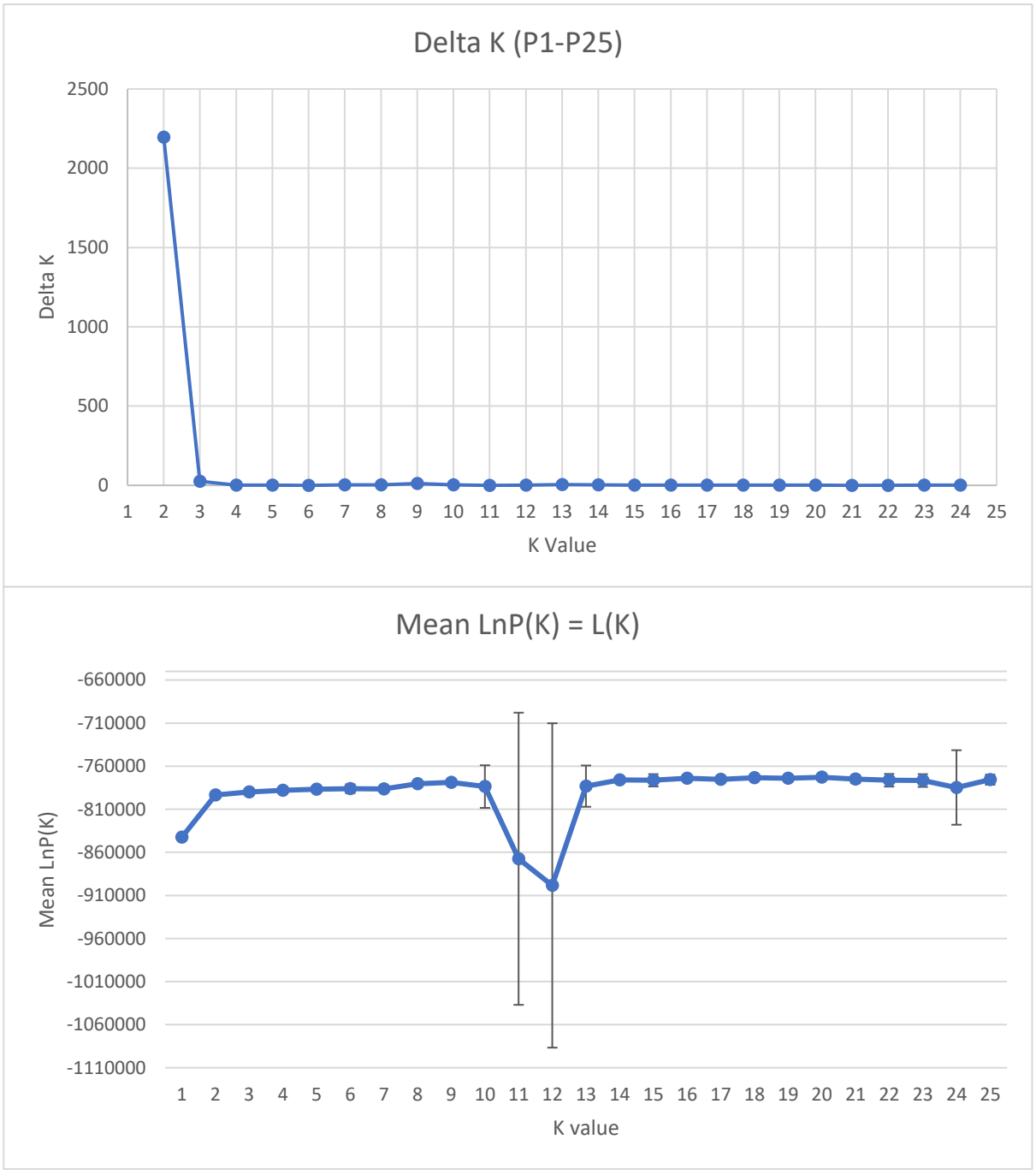
Yeoh, S. H., J. C. Bell, W. J. Foley, I. R. Wallis and G. F. Moran, 2012 Estimating population boundaries using regional and local-scale spatial genetic structure: an example in Eucalyptus globulus. Tree Genetics & Genomes 8**:** 695-708.

APPENDICES

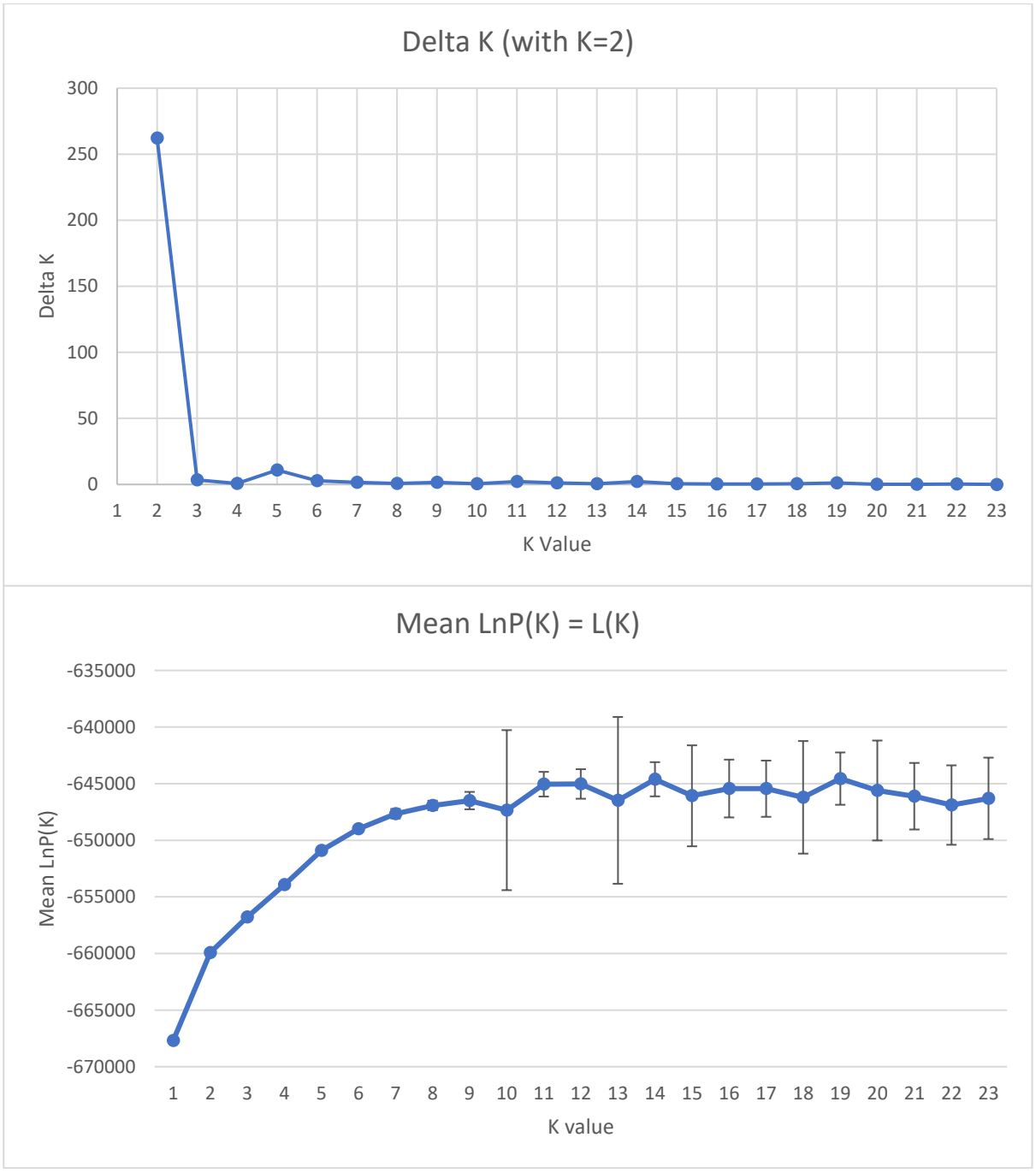
Appendix 1.

Delta K and LnP(K) graphs plotted in Microsoft Excel for three different datasets of P1-P25 (A), P3-P25 (B), P3-P17 & P22-P25 (C).

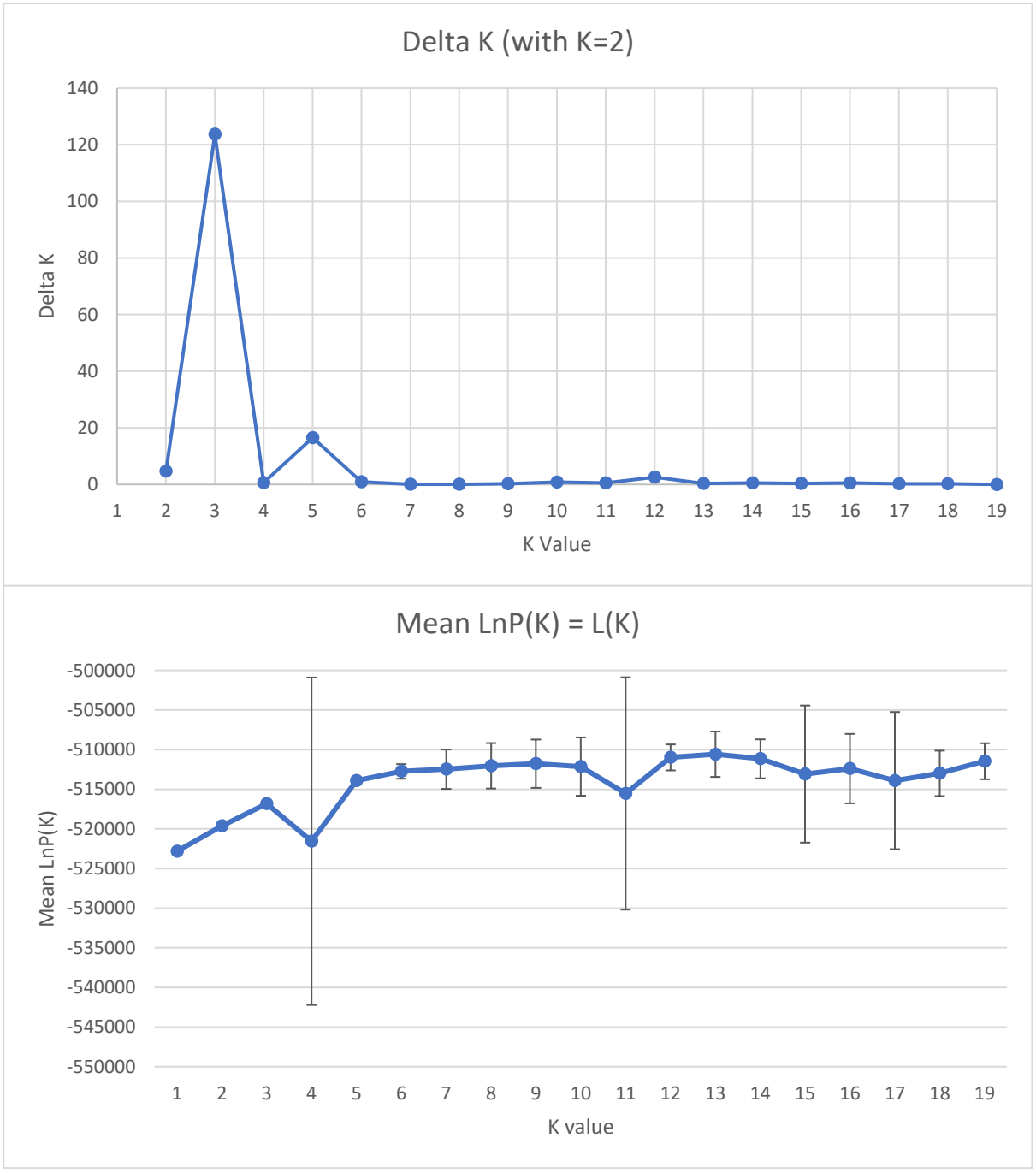
A. Delta K and LnP(K) graphs of P1-P25



B. Delta K and LnP(K) graphs of P3-P25



C. Delta K and LnP(K) graphs of P3-P17 & P22-P25



Appendix 2.

Locations of *E. bosistoana* trials in New Zealand as per the established year and of the *E. argophloia* trials.