Genome-wide patterns of genetic diversity, population structure and demographic history in mānuka (*Leptospermum scoparium*) grown on indigenous Māori land

Emily Koot[1], Elise Arnst[2], Melissa Taane[1], Kelsey Goldsmith[3], Peri Tobias[4,5], Eleanor Dormont[6], Tate Hancox[6], Kate Delaporte[6], Amali Thrimawithana[7], Kiri Reihana[2], Santiago C. González-Martínez[8], Victor Goldsmith[3], Gary Houliston[2], David Chagné[1#]

1 The New Zealand Institute for Plant and Food Research Limited (Plant & Food Research), Private Bag 11600, Palmerston North 4442, New Zealand

2 Manaaki Whenua Landcare Research, Lincoln, New Zealand

3 Kaiaka Consulting Limited, Wellington, New Zealand

4 School of Life and Environmental Sciences, University of Sydney, Eveleigh, NSW 2015, Australia

5 Plant & Food Research Australia, Sydney, Australia

6 University of Adelaide, Australia

7 Plant & Food Research, Private Bag 92169, Auckland 1142, New Zealand

8 Univ. Bordeaux, INRAE, UMR BIOGECO, 33610 Cestas, France

#: author for correspondence: David.Chagne@plantandfood.co.nz

Abstract

Leptospermum scoparium J. R. Forst et G. Forst, known as mānuka by Māori, the indigenous people of Aotearoa (New Zealand), is a culturally and economically significant shrub species, native to New Zealand and Australia. Chemical, morphological and phylogenetic studies have indicated geographical variation of mānuka across its range in New Zealand, and genetic differentiation between New Zealand and Australia. We used pooled whole genome re-sequencing of 76 *L. scoparium* and outgroup populations from New Zealand and Australia to compile a dataset totalling ~2.5 million SNPs. We explored the genetic structure and relatedness of *L. scoparium* across New Zealand, and between populations in New Zealand and Australia, as well as the complex demographic history of this species. Our population genomic investigation suggests there are five geographically distinct mānuka gene pools within New Zealand, with evidence of gene flow occurring between these pools. Demographic modelling suggests three of these gene pools have undergone expansion events, whilst the evolutionary histories of the remaining two have been subjected to contractions. Furthermore, mānuka populations in New Zealand are genetically distinct from populations in Australia, with coalescent modelling suggesting these two clades diverged ~9 –12 million years ago. We discuss the evolutionary history of this species and the benefits of using pool-seq for such studies. Our research will support the management and conservation of mānuka by landowners, particularly Māori, and the development of a provenance story for the branding of mānuka based products.

Keywords: Myrtaceae; single nucleotide polymorphism; pool sequencing; population structure; demographic history; mānuka; provenance story

Introduction

According to Māori lore, all trees in Aotearoa are the children of Tāne Mahuta, God of the forest (Grey, 1904; Gudgeon, 1905). Māori have many uses for the taonga (culturally significant/treasure) tree species mānuka (*Leptospermum scoparium* J. R. Forst et G. Forst) including for medicine (Best, 1904; Riley, 1994), food (Colenso, 1868; Crowe, 2001), hunting (Beattie, 1920), fishing (Best, 1929; Colenso, 1868), weaponry (Colenso, 1868; Hiroa, 1949; Palmer, 1957) and as a building material (Colenso, 1868). More recently, mānuka has been used to produce high value honey, including by Māori-owned agribusinesses, pushing the export value of the New Zealand honey industry to $348 million (NZD) in 2018 (Ministry for Primary Industries, 2018; J. Stephens, Molan, & Clarkson, 2005). Indigenous communities and businesses across New Zealand are seeking knowledge about the genetic variation and evolutionary history of mānuka as it is a culturally, ecologically and economically important species. This is so that an authentic and higher value honey industry based on indigenous plants of known New Zealand provenance can be established, and so that this genetic resource can be managed more sustainably for the future.

Current understanding of the population structure within New Zealand mānuka is predominantly based on studies of its antibacterial metabolic compounds and morphological variation (Douglas et al., 2004; Perry et al., 1997; Porter & Wilkins, 1999; Ronghua, Mark, & Wilson, 1984). Perry et al. (1997) were able to establish via common garden experiments that regional differences in mānuka's triketone chemotypes are not determined by abiotic factors, demonstrating there is a genetic contribution to the regional differentiation of essential oil compounds in mānuka. Additionally, Perry et al. (1997) identified distinct chemotypic variation between *L. scoparium* samples from Australia and New Zealand, and recommended a taxonomic revision of this species. Expanding on this earlier work, Douglas et al. (2004) carried out extensive chemotypic sampling of mānuka across New Zealand, elucidating 11 chemotypes that displayed degrees of geographic association. Unique compounds are

also found in the nectar of mānuka (dihydroxyacetone) and its honey (methylglyoxal) (Adams et al., 2008; Adams, Manley-Harris, & Molan, 2009; Atrott, Haberlau, & Henle, 2012; Mavric, Wittmann, Barth, & Henle, 2008), leading to the development of several medical wound care products in recent years, as well as demand as a health food. Studies by Williams et al. (2014) and J. M. C. Stephens (2006) indicate regional variation in the dihydroxyacetone content of mānuka's nectar and the non-peroxide antibacterial activity of mānuka honey, respectively, across New Zealand. Additionally, morphological and phenological variation of mānuka across New Zealand was found to correlate with geographic and environmental factors, and common garden experiments indicated that differences were likely linked to genetic variation (Ronghua et al. (1984).

Recently, a molecular study of mānuka was published (Buys et al., 2019), revealing the phylogenetic grouping of New Zealand mānuka into three clades – a Northland clade, a central and southern North Island clade and a South Island clade. However, as an inadequate sample size prevented Buys et al. (2019) from carrying out population genetic analyses, the population genetic structure of mānuka across New Zealand remains unexplored. Similarly, gaps remain in our knowledge regarding the demographic and evolutionary history of *L. scoparium* in New Zealand and the relationships between New Zealand and Australian populations of the species. The continent Zealandia split from Gondwana (and present-day Australia) ~80 Mya, and has since been subject to global fluctuations in climate, as well as geological disturbance as a result of straddling the Australian and Pacific continental plates (Lamb, Mortimer, Smith, & Turner, 2016). It is hypothesised that *L. scoparium* arrived in New Zealand from Australia (either from the mainland or from Tasmania) during the Miocene (23.03 – 5.33 Mya) (Thompson, 1989; Thornhill, Ho, Külheim, & Crisp, 2015). During this time, New Zealand has been subjected to geological events (mountain building, volcanism) (Kamp, 1986; Trewick & Bland, 2012) and glacial cycling (12,000 – 5 Mya) (McCulloch, Wallis, & Waters, 2010; Patterson et al., 2014) that may have shaped the diversity and distribution of mānuka following its establishment in New Zealand.

The use of pooled sequencing (pool-seq) to obtain genome-wide variants provides a relatively economical means to study genomic variation at a population scale (Dennenmoser, Vamosi, Nolte, & Rogers, 2017; Fischer et al., 2013; Futschik & Schlötterer, 2010; Mimee et al., 2015; Roda et al., 2013; Schlötterer, Tobler, Kofler, & Nolte, 2014). Equal quantities of high-quality DNA from multiple individuals (preferably n= >25, (Fracassetti, Griffin, & Willi, 2015)) are pooled for each population and sequenced using next-generation sequencing (NGS). Mapping to a high-quality reference genome reveals genetic variants in the NGS data – commonly single nucleotide polymorphisms (SNPs). Provided that read coverage across sites is high (>100x) to ensure reliability (Fracassetti et al., 2015; Lynch, Bost, Wilson, Maruki, & Harrison, 2014), various analyses can then be carried out, including: population genomic analyses (Corander, Majander, Cheng, & Merilä, 2013; Ferretti, Ramos-Onsins, & Pérez-Enciso, 2013; Mimee et al., 2015); genome wide association studies (Bertelsen et al., 2016; Endler, Betancourt, Nolte, & Schlötterer, 2016); gene environment association studies (Fischer et al., 2013; Turner, Bourne, Von Wettberg, Hu, & Nuzhdin, 2010); and the modelling of evolutionary histories (Christe et al., 2017; Leblois et al., 2018). Pool-seq is an increasingly popular method, and provided the correct quality control steps are taken, validation studies have proven it to be a valuable and informative tool (Anderson, Skaug, & Barshis, 2014; Lynch et al., 2014; Rellstab, Zoller, Tedder, Gugerli, & Fischer, 2013; Schlötterer et al., 2014).

A complete assembly of the mānuka genome was recently developed, scaffolded into the expected 11 pseudo-chromosomes, which are syntenic to the related Myrtaceae model species *Eucalyptus grandis* (Thrimawithana et al., 2019). In the present study we applied pooled whole genome re-sequencing of mānuka and kānuka (*Kunzea robusta* de Lange et Toelken, used as outgroup) collected from naturally regenerating stands on indigenous Māori owned land across New Zealand and in native stands on public land in Australia to identify genome-wide DNA variants. With this data, we

differentiate local provenances of *L. scoparium* within New Zealand, and between New Zealand and Australian populations, as well as provide insights into the complex demographic and evolutionary history of this species.

Materials and Methods

Indigenous considerations and data sovereignty

Māori, the indigenous people of New Zealand, currently have a claim under the Waitangi Tribunal (the mechanism for redress via Te Tiriti o Waitangi (The Treaty of Waitangi) – the founding document of New Zealand), relating to intellectual property pertaining to native flora and fauna (Waitangi Tribunal, 2011). This claim has not been settled to date, but current convention in New Zealand research is to recognise Māori connection and guardianship over native flora and fauna, acknowledging that this claim is still before the tribunal. Therefore, before samples for genetic analysis were collected from natural stands of mānuka grown on Māori freehold land, mana whenua (Māori who have historical and territorial rights over the land) were individually approached and were briefed about the objectives of the research study. The privilege to access taonga material was granted for the purpose of this research and further research on these samples would require consent, including access to more samples, and further analysis or disclosure of the exact identity of the trees. For this reason, our sampling sites are named according to their region of origin and the exact location of the samples and identity of the people who contributed has purposefully been occulted to address the sensitivity of this indigenous intellectual property. For further comments on the involvement and interaction of Māori and science see Morgan et al. (2019) and Hudson et al. (2020).

Study species

The genus *Leptospermum* J. R. Forst et G. Forst of the Myrtaceae family comprises approximately 87 species, predominantly distributed throughout south-east Australia (Victoria, New South Wales, Tasmania); species are also present in South East Asia, New Guinea, Rarotonga and New Zealand (Bean, 1992, 2004; Cheeseman, 1925; Lyne, 1993; Lyne & Crisp, 1996; Thompson, 1989). Specifically, one native species resides in New Zealand – *L. scoparium*. *Leptospermum scoparium*, indigenously known in New Zealand as mānuka or kahikātoa, is a widespread, woody species, distributed from coastal to sub-alpine environments (1800m above sea level) (J. Stephens et al., 2005). Found across all New Zealand's major islands, including the Chatham Islands, mānuka is able to establish in adverse environments (e.g. low fertility soils, peat swamps, volcanic soils, geothermal areas, exposed coastal, sub alpine) and is a primary coloniser of recently disturbed habitats (e.g. post-fire and deforestation) (Allan, 1961; Ronghua et al., 1984).

Sampling

In total, samples from 2,325 individual mānuka (n = 2,265) and kānuka (*Kunzea robusta*) (n = 60) trees were collected from 70 sites (mānuka = 68 sites, kānuka = two sites) around New Zealand and six sites in Australia (Supplementary material. 1; Table 1). For each tree, fifteen young expanding leaves were sampled in duplicate, with at least 50m between trees to ensure they were unlikely siblings, and metadata (e.g. latitude, longitude) were recorded. Samples were kept on silica beads in 2 mL screw cap tubes before being stored at -80°C. For all New Zealand samples DNA was extracted using a modified CTAB protocol (Gardiner et al., 1996) and quantified using the Quant-iT™ PicoGreen™ dsDNA Assay kit (Invitrogen, Burlington, ON, Canada) and a SpectraMax® Gemini EM Microplate Reader (Molecular Devices, San Jose, California, USA). Australian samples were extracted using the Qiagen DNeasy Plant Mini Kit on the QIAGEN QIAcube, the ISOLATE II Plant DNA Kit (Bioline) using buffer PA1 to manufacturer's protocol and extending the lysis incubation to 40 minutes, or by using the Machery-Nagel Nucleospin Plant II Kit with the PL2/PL3 buffer system. Results were examined using SoftMax®

Pro Software (Molecular Devices, San Jose, California, USA). DNA of each sample was normalised to the same concentration and then pooled for each collection site to obtain two micrograms of total DNA per site.

Pooled sequencing and variant calling

All 76 DNA pools were indexed using Illumina (Illumina, Inc., San Diego, USA) commercial indices, following the manufacturer's protocol. DNA pools were sequenced using the Illumina NovaSeq 6000 System utilising S2 flow cell technology and seven lanes of 2x150 base pair (bp) output at the Australian Genome Research Facility (AGRF, Melbourne, Australia). Raw fastq sequencing files from the 76 pools were separated by indices and quality control was carried out using md5 checksum v3.3 (Tung et al., 2017) and FastQC v0.11.8 (Andrews, 2016). Read mapping was performed against the reference genome of mānuka 'Crimson Glory' (Thrimawithana et al., 2019) using Bowtie2 v3.4.3 (Langmead & Salzberg, 2012), applying the --end-to-end parameter for each site independently. SAMtools v1.7 (Li et al., 2009) was utilised to convert output SAM files into BAM files, which were sorted and indexed against 'Crimson Glory'. Coverage statistics for each BAM file were acquired by applying the SAMtools flagstat utility. BCFtools v1.9 (Li, 2011) was used to detect variants for each pseudo-chromosome of the reference genome, implementing the –m and –v options of BCFtools. Filtering of raw variants into high confidence SNPs was carried out using VCFtools v0.1.14 (Danecek et al., 2011), applying the following parameter settings: a minor allelic frequency (MAF) of 5% and 2% (--maf 0.05 and --maf 0.02); no indels (--remove-indels); a minimum mean read depth of 100X for each site (--min-meanDP 100); and no missing data (--max-missing 1). We created six SNP datasets in total, consisting of three different population datasets (New Zealand only populations (n=70), New Zealand and Australian *L. scoparium* populations (n=74), and all populations including kānuka (n=76)) with two different MAF filtering parameters applied to each – 0.02 and 0.05. Different datasets were required for carrying out various population genomic analyses.

Population genetics and demographic modelling

Allele frequencies were calculated following SNP calling in the R v3.5.0 (R Core Team, 2013) package vcfR v1.8.0 (Knaus & Grünwald, 2017). Population structure was examined via a Discriminant Analysis of Principal Components (DAPC) using the R package Adegenet v2.1.1 (Jombart, 2008), applying the find.clusters function to determine the optimal k and therefore number of clusters within the data. Population pairwise Fixation Indices (*F*st) were calculated utilising PoolFstat v1.0.0 (Hivert, Leblois, Petit, Gautier, & Vitalis, 2018). Neighbour-net networks were implemented in SplitsTree4 v4.14.8 (Huson & Bryant, 2005) based on the aforementioned *F*st genetic distance matrix. Diversity statistics and neutrality tests (segregating sites (S), nucleotide diversity (π), Watterson's θ and Tajima's D) were carried out using NPStats v1 (Ferretti et al., 2013). An allele count dataset was utilised for TreeMix v1.13 analyses (Pickrell & Pritchard, 2012) where eleven migration events (0-10) were explored in 1000 SNP blocks. Covariance data were extracted from the TreeMix output files and an equation provided by Pickrell and Pritchard (2012) (fraction *f*, page 5) was used to calculate the explained variance of each migration run.

To explore the demographic history within and between New Zealand and Australian gene pools (established in our DAPC, *F*st and TreeMix analyses), we applied a coalescent simulation-based method in fastsimcoal v2.6.0.3 (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013). Rare variants were called in VCFtools using the settings: MAF 0.00, missing = 0, depth = 100, resulting in a dataset of 5,503,881 SNPs. The minor allele frequencies of these SNPs were averaged across the populations within each gene pool, to fully encapsulate the genetic diversity and structure of each. Unfolded Site Frequency Spectra (SFS) and two-dimensional SFS (2D-SFS) (using P011 (kānuka) as an outgroup) were calculated for each gene pool using SweepFinder2 (DeGiorgio, Huber, Hubisz, Hellmann, & Nielsen, 2016) and a custom R script, respectively. For every demographic model tested,

100,000 coalescent simulations were applied in fastsimcoal2, with maximum likelihood estimates calculated based on differences between the input observed SFS and the output expected SFS. Models were repeated 100 times, and a global maximum likelihood estimate was obtained from these independent runs and Akaike's Information Criterion (AIC) calculated for model comparison and selection.

Individual demographic models were implemented for each gene pool in fastsimcoal2, with neutral, bottleneck, expansion and contraction models explored. Multi-gene pool models were then carried out. Additionally, a 2D unfolded SFS matrix between the Australian gene pool and a representative New Zealand gene pool (Northland) was calculated to explore the demographic history of divergence and gene flow between New Zealand and Australian *L. scoparium*. Five models were tested: divergence followed by isolation (D-I); divergence followed by continuous gene flow (D-CGF); divergence with ancestral gene flow followed by isolation (D-AGF-I); divergence with only recent gene flow (D-RGF); and divergence with continuous gene flow as well as contraction for Australia and expansion for New Zealand (D-CGF-CE). The structure of the final model was based on findings of the initial individual demographic models.

Results

Whole genome re-sequencing

In total, samples from 2,325 individual mānuka (n = 2,265) and kānuka (*Kunzea robusta*) (n = 60) trees were collected from 70 sites (mānuka = 68 sites, kānuka = 2 sites) around New Zealand and six sites in Australia (Supplementary material 1; Table 1). The DNA of 30 trees was pooled for each New Zealand site and 15 to 30 trees from each Australian site. For some populations we were unable to extract adequate DNA from all 30 samples, thus pools for some populations comprised fewer than 30

individuals (Table 1). From the Novaseq 6000 sequencing runs a total of 5811.53×10^9 nucleotidic bases ((1×10^9 = Gb) from 19.27×10^9 sequencing reads were generated, with an average 253.61 billion sequencing reads and 76.46 Gb per collection site. The quality of the sequencing reads was high, with most sequences having a quality score greater than Q30 across the 150 bp reads (data not shown). Read mapping to the reference genome was consistent among populations and ranged between 50 and 64% of the *L. scoparium* reads mapping to the reference genome of *L. scoparium* 'Crimson Glory' and 26% of kānuka reads mapping to the mānuka reference genome (Supplementary material 2). The remaining sequences (on average 41% of reads) were from fungal and bacterial associated microbiome and not used for further analysis in this study. The equivalent coverage obtained from reads mapped had a mean depth of 169X across the genome (Supplementary material 3). Base pair coverage for the complete 76 population dataset ranged from 150,419,718 bp (for kānuka population P011, equivalent to 60% of the reference genome assembly) to 247,091,534 bp (population P003, equivalent to 99.5% of the reference genome assembly) (Supplementary material 4). When this was filtered to include only nucleotide bases with a read depth greater than 100X, the range in base pair coverage reduced to 65,445,232 (for kānuka population P008, equivalent to 26.2% of the reference genome assembly) to 196,036,101 bp (for population P060, equivalent to 79% of the reference genome assembly). Kānuka populations (P008 and P011) had the lowest coverage, as expected for a relative in a different genus.

Variant detection

A total of 10.14 million candidate variants (SNPs and indels) were detected across the 76 DNA pools and throughout the genome prior to filtering. A stringent filtering protocol was applied to the raw set of variants to remove sequencing errors, rare variants and variants occurring in low depth regions, only keeping SNPs with an average read depth greater than 100X and no missing data, resulting in 5,503,881 SNPs being detected across the reference assembly of 'Crimson Glory'. When further

filtering was applied using minor allele frequencies (MAF) of 5% and 2%, 1,498,405 and 2,513,694 SNPs were detected across all 76 collection sites (including both kānuka and Australian *L. scoparium* samples), respectively (Table 2). When both kānuka populations were excluded (i.e. only *L. scoparium* populations taken into account), a total of 2,580,451 and 4,049,649 SNPs were detected using MAF of 5% and 2%, respectively. When only the 68 New Zealand mānuka populations were included (i.e. Australian *L. scoparium* and kānuka populations excluded), a total of 2,526,589 and 3,270,864 SNPs were detected using MAF of 5% and 2%, respectively.

Genetic differentiation and population structure

Discriminant Analysis of Principal Components (DAPC) (Jombart, Devillard, & Balloux, 2010) of allele frequencies generated from 1,498,405 SNPs from 76 populations (including Australian *L. scoparium* and kānuka) revealed that these populations group into seven clusters (Fig. 1) as determined by the optimized K-means clustering algorithm in Adegenet v2.1.1 (Jombart et al., 2010). The Adegenet find.clusters function was used to run 30 successive K-means analyses, with an optimal K of seven selected based on Bayesian Information Criterion (BIC) (Fig. 1A). Linear dimensions (LD) one, two, three and four (LD1, LD2, LD3 and LD4) of the DAPC (K = 7) explained 69.55%, 13.55%, 7.54% and 4.56% of the data variation, respectively (Fig. 1 B, D). New Zealand populations from the north-eastern South Island (NESI), south-western South Island (SWSI), northern North Island (NNI), central and southern North Island (CSNI) and the East Cape of the North Island (ECNI) each formed separate clusters, suggesting strong geographic structuring of mānuka populations into five distinct gene pools across New Zealand (Fig. 1C). There was also evidence of strong partitioning between *L. scoparium* populations in New Zealand and those sampled from Australia (including Tasmania), with the six populations of *L. scoparium* from Australia forming their own cluster.

289 The exploration of DAPC posterior-membership probabilities at five different K values (K = 5 to K = 9)
290 further demonstrated the strong geographic structuring of the New Zealand mānuka populations (Fig.
291 2). When K = 5 three New Zealand clusters were identified – the NNI cluster, a central, eastern and
292 southern North Island cluster and one cluster containing all South Island populations. When K = 6 the
293 central, eastern and southern North Island cluster split into two – into the CSNI cluster and the ECNI
294 cluster. When K = 7 (the optimal K value), the South Island cluster divided into the NESI and SWSI
295 clusters. The single Tasmanian population was identified as an additional cluster when K = 8, however
296 at K = 9 this cluster was lost, with the NNI and SWSI clusters both dividing in two. The NNI populations
297 that formed an additional cluster at K = 9 (P037, P040 and P041) are distributed at the northern tip of
298 the North Island, and the four SWSI populations that formed an additional cluster at K = 9 are all
299 distributed on the West Coast of the South Island, indicating additional geographical structuring. P050
300 is the only population to demonstrate mixture between clusters (NNI and CSNI), predominantly being
301 clustered with CSNI at K = 5, K = 6, K = 7 and K = 8, and is entirely clustered with NNI at K = 9. This
302 population occurs at the northern edge of the CSNI distribution, and is in close proximity to an NNI
303 population (P066).

304

305 Pairwise *F*st genetic distances were calculated in PoolFstat v1.0.0 (Hivert et al., 2018) using mpileup
306 and sync input files (Fig. 3A). Overall average *F*st distances for all New Zealand populations was 0.128
307 (SD = 0.057) (Table 3). Between the New Zealand gene pools, the lowest mean *F*st values were
308 observed between populations from CSNI and ECNI (average *F*st = 0.079; SD = 0.023), whilst highest
309 mean *F*st values were between populations from ECNI and NESI populations (average *F*st = 0.196, SD
310 = 0.019). Within the New Zealand gene pools, average *F*st values varied between 0.024 (SD = 0.016)
311 (EC NI) and 0.073 (SD = 0.024) (NESI). Pairwise *F*st distances between all populations can be found in
312 supplementary material (Supplementary material 5). Average *F*st distances between New Zealand and
313 Australian populations, New Zealand and Tasmanian populations, and Australian and Tasmanian

314 populations were 0.357 (SD = 0.02), 0.383 (SD = 0.011) and 0.308 (SD = 0.006) respectively. Similarly,
315 an unrooted, neighbour-joining network generated in SplitsTree4 v4.14.8 (Huson & Bryant, 2005)
316 using the pairwise *F*st distance matrix revealed strong structuring between the aforementioned gene
317 pools; however, there was also an apparent conflicting signal observed within the dataset, visible in
318 the boxes formed within the SplitsTree network (Fig. 3B).

319

320 *Genetic diversity and demographic modelling*

321 Summary statistics generated by NPStats v1 (Ferretti et al., 2013) were calculated in non-overlapping
322 10 kb windows and results are presented as the weighted medians of these windows averaged per
323 gene pool (Table 4). Weighted medians by population can be found in supplementary material
324 (Supplemental material 6). The average weighted median for segregating sites (S) varied from 28 (SD
325 = 11) (for kānuka) to 118 (SD = 92.8) (for Australian populations, including Tasmania) across all 76
326 populations. Within the New Zealand gene pools, S varied from 39.83 (SD = 4.01) (ECNI) to 46.03 (SD
327 = 9.5) (CSNI). Watterson's θ varied from 0.0076 (SD = 0.0003) (NESI) to 0.0120 (SD = 0.002) (Australia)
328 across all populations, and from 0.0076 (SD = 0.0003) (NESI) to 0.0091 (SD = 0.0014) (CSNI) within the
329 New Zealand mānuka populations. Nucleotide diversity (π) ranged from 0.0067 (SD = 0.0002) (ECNI)
330 to 0.0108 (SD = 0.0012) (Australia), and from 0.0067 (SD = 0.0002) (ECNI) to 0.0075 (SD = 0.001) (CSNI)
331 within the New Zealand mānuka populations. Average Tajima's D weighted medians ranged from -
332 0.905 (SD = 0.142) (NNI) to -0.514 (SD = 0.362) (Australia) across populations, and from -0.905 (SD =
333 0.142) (NNI) to -0.522 (SD = 0.08) (NESI) within the New Zealand mānuka populations, indicating an
334 excess of low-frequency alleles.

335

336 An allele count dataset consisting of 74 populations (MAF of 0.05, and excluding kānuka) totalling
337 2,580,451 SNPs was used for the TreeMix (Pickrell & Pritchard, 2012) analysis. Based on the explained
338 variance of eleven different migration events, it was determined that five to six migration events best

339 explained the observed data used in the TreeMix model (Fig. 4A, B). Phylogenetic trees produced by
340 this analysis support the clustering of New Zealand mānuka into five gene pools (NNI, CSNI, ECNI, and
341 two South Island clades), with clear division also seen between New Zealand mānuka and Australian
342 *L. scoparium* (Fig. 4C). The five migration event analysis suggests two major migration events have
343 occurred between New Zealand and external gene pools: one between mainland Australia and
344 Northland (NNI), and another from an ancestral population in Australia or Tasmania (not sampled in
345 our study) into the NNI gene pool. Within New Zealand, there have been major migration events
346 between the SWSI gene pools and NNI, and between the CSNI gene pool and NESI. An additional
347 migration event is seen within the CSNI gene pool. When a sixth migration event was applied alongside
348 the aforementioned events, the model determined an additional event from NNI to the SWSI gene
349 pool (Fig. 4C). Migration events with the greatest weight are seen between the CSNI and NESI gene
350 pools, and between NNI and SWSI gene pools. The results suggest no substantial migration events
351 have occurred from New Zealand back to Australia.

352

353 SFS-based demographic models that explored four hypotheses (population stability, bottleneck,
354 expansion and contraction models) suggest that NNI, CSNI and SWSI gene pools have all undergone
355 expansion events, whilst the ECNI, NESI and Australian gene pools have all undergone contractions
356 (Fig. 5A; Table 5a). Parameter estimates and associated summary statistics inferred from 100
357 parametric bootstraps of the most suitable model for each gene pool, as well as output of the best
358 overall model based on AIC can be found in Supplementary material 7. In the case of the three gene
359 pools that have undergone population size contractions, results for bottleneck and contraction models
360 were very similar and in some instances the bottleneck model had better support. However, results
361 of the bottleneck models all indicated that the effective population size of the gene pools does not
362 recover following the bottleneck, and is less than the effective population size during the bottleneck
363 – effectively indicating that the population sizes have continued to contract (hence the contraction
364 model being selected as the better model). Here we report the results of the models with the lowest

AIC from 100 runs of the best model for each gene pool. NNI is estimated to have begun expanding ~243,602 generations ago, increasing effective population size (N_e) by ~2739.8%; the CSNI gene pool is estimated to have begun expanding ~23,721 generations ago, increasing N_e by ~941.35%; and the SWSI gene pool is estimated to have begun expanding ~362,230 generations ago, increasing N_e by ~458.03%. Whilst the ECNI gene pool is estimated to have begun contracting ~964 generations ago, with N_e decreasing by ~97.02%, the NESI gene pool is estimated to have begun contracting ~550 generations ago, decreasing N_e by ~95.62%, and the Australian gene pool is estimated to have begun contracting ~1,155 generations ago, decreasing N_e by ~98.98%.

The model that best explains the evolutionary history of Australian and New Zealand populations (represented by the NNI gene pool) and the arrival of mānuka to New Zealand was the D-CGF-CE model (Fig. 5B; Fig. 6; Table 5b). The parameter estimates for this model suggest there has been continuous, asymmetrical gene flow between the two countries, with more gene flow occurring from Australia to New Zealand than vice versa (Australia to NNI = 1.03E-05, NNI to Australia = 2.67E-06). Results of this model also suggest that the split between New Zealand and Australian populations occurred ~581,569 generations ago. The parameter estimates for the Australian gene pool are similar under this model to the parameter estimates determined in the above individual models, with contraction of the Australian gene pool estimated to have begun ~1,235 generations ago, with a decrease in N_e by ~99.0%. However, parameter estimates for the Northland gene pool are comparably different from the individual model estimates, and this gene pool is estimated to have begun expanding ~17,507 generations ago (cf. ~243,602 generations ago), with N_e increasing only by ~605.86% (c.f. the individual NNI expansion model of ~2739.8%).

Discussion

A genomics study supporting indigenous branding

This first population genomic study of *L. scoparium* indicates clear geographic structuring of mānuka across New Zealand and a strong genetic differentiation between New Zealand and Australian populations. This research was co-developed in response to demand for this information by indigenous agribusinesses, with an interest in developing strong regional branding associated with their cultural connections to the land. Such evidence will assist Māori by addressing economic, cultural and ecological aspects. In terms of economic benefits for Māori, our research will underpin the veracity of novel mānuka-based products and assist Māori to achieve the potential of this natural resource, leveraging off global demand for provenance information – particularly for food. Determining the provenance of the plants will provide indigenous innovators producing honey with marketing opportunities unique to Māori. This branding will identify these products as being premium and authentic, and create a strong differentiation from other types of honeys, including from Australia. Finally, understanding mānuka genetic diversity will contribute to better management of this resource, improving the resilience of the honey industry.

Pool-seq

Pooled whole genome re-sequencing (pool-seq) was applied in this study as an alternative to individual genotyping. We chose this strategy as opposed to reduced representation methods such as genotyping by sequencing (GBS) as it gives access to the full set of variants across the genome. In our experience, GBS can yield variable read depth across genomic regions, with many errors and missing data (Bhatia, Wing, & Singh, 2013). Whilst pool-seq does not give true genotypes and does not allow haplotyping, we believe our dataset is superior to a GBS dataset as we obtained a large proportion of the reference genome with a consistent read depth. Additionally, pool-seq is an economical choice compared with barcoding the complete sampling set, as only a limited number of costly sequencing indices are required. This allowed us to sequence a larger set of individual trees across more regions

(2,325 trees across 76 sites). One hundred times coverage and >25 individuals per pool were recommended to attain reliable results (Fracassetti et al., 2015; Lynch et al., 2014), and we were able to achieve this for the majority of populations. This enabled us to estimate allelic frequencies within populations. Our results are consistent with morphological (Ronghua et al., 1984), phenological (Ronghua et al., 1984) and chemotypic (Douglas et al., 2004; Perry et al., 1997) studies that have suggested there is geographical structuring of mānuka across New Zealand, and builds on previous genetic work by Buys et al. (2019).

Population genetic structure

Genetics are commonly mentioned as a factor influencing the observed variation in the chemotypic compounds, morphology and phenology across mānuka's range. However, no previous studies have investigated the population genetic structure of *L. scoparium* across New Zealand, or between New Zealand and Australia, or provided insight into the demographic history of this species. Our population genetic investigation indicates that there are five genetically and geographically distinct gene pools of *L. scoparium* in New Zealand, and that these populations are genetically distinct from Australian *L. scoparium* populations. DAPC, *F*st and SplitsTree analyses all support the genetic differentiation of these gene pools. However, the TreeMix analyses suggest that despite this genetic differentiation, gene flow has occurred between these gene pools – including between New Zealand and Australian populations. Genetic differentiation (*F*st) indicates that New Zealand populations have been separated from Australian populations for a long period of time; however again, the TreeMix analyses suggest that multiple migration events have occurred from Australia to New Zealand (most likely to NNI populations). There is no evidence to support significant migration occurring from New Zealand back to Australia.

Buys et al. (2019) described three genetically and geographically distinct clades of mānuka (northern North Island, central and southern North Island, South Island), with some weakly supported sub-structuring within these clades. With more sampling sites and a larger genetic dataset, we were able to elucidate two additional clades – dividing the Buys et al. (2019) South Island clade into two (NESI and SWSI), and establishing East Cape populations as a separate gene pool (ECNI). Buys et al. (2019) also discussed the potential reason for poor resolution within the branches of the three New Zealand mānuka clades they described, suggesting it was the result of conflicting signal within their dataset. We can now confirm that the conflicts in signal they encountered are likely the result of gene flow occurring between populations and gene pools within New Zealand and between Australian and New Zealand populations.

Between the New Zealand populations, pairwise *F*st was relatively low (cf. difference between New Zealand and Australian populations), particularly between neighbouring populations where gene flow is more readily facilitated. This is demonstrated by ECNI and CSNI gene pools having the lowest average pairwise *F*st, and NESI and SWSI also exhibiting relatively low average pairwise *F*st. Interestingly, the highest pairwise *F*st were not between the most geographically distant gene pools, however, but between ECNI and NESI. The seeds of mānuka are dispersed via wind (Esler & Astridge, 1974), and it is possible that wind may more easily carry seeds from the South Island to the northern North Island, than across the North Island to the East Cape.

Evolutionary history

The genus *Leptospermum* is thought to have evolved in Australia during the Miocene (5 – 23Mya), with the split between *L. scoparium* and its sister species *L. trinervium* estimated to have occurred 15.9 Mya (95% confidence interval = 6.7 – 26.8 Myr) in Australia; arrival of an ancestral *L. scoparium*

lineage to New Zealand would have occurred subsequently (Thompson, 1989; Thornhill et al., 2015). Results of our demographic models are congruent with this. If we assume a generation time of 15 – 20 years for mānuka, our results suggest that the New Zealand lineage of *L. scoparium* split from the Australian lineage ~9-12 Mya. This timing also fits with the arrival and divergence of other woody angiosperm species in New Zealand, following the end of the Miocene Thermal Optima ~15 Mya (Heenan & McGlone, 2019; Prebble et al., 2017). Therefore, the population structure we have observed in New Zealand mānuka (into five gene pools) likely reflects phylogeographic patterns associated with climatic and glacial cycling of the Plio-Pleistocene (12,000 Ka – 5 Mya) (McCulloch et al., 2010; Patterson et al., 2014), and/or New Zealand's turbulent geological history of mountain building and volcanic activity (Kamp, Green, & White, 1989; Trewick & Bland, 2012).

For two gene pools (NNI, SWSI), the individual demographic models indicate that N_e began increasing ~4-7 Mya, suggesting these gene pools may have benefitted from the cooler climate of the Plio-Pleistocene, expanding their ranges in the north of the North Island and in the south western South Island, respectively. These two regions are considered to have been relatively geologically and climatically stable over the last ~5 Myr (M. McGlone, 1985). In particular, the southern South Island during this time would have consisted predominantly of tussock grasslands and shrub-lands, which may have provided suitable open habitat for mānuka to establish and then to expand its range (Alloway et al., 2007; M. S. McGlone, Newnham, & Moar, 2010; Trewick & Bland, 2012). If we take the estimated date of expansion of NNI from the Australian and New Zealand divergence model, it suggests that NNI did not begin expanding until more recently (~260,000-350,140 years ago). This timing is similar to that estimated for the CSNI gene pool, with expansion estimated to have begun ~350,000 – 475,000 years ago. Fires occurred regularly in acid mires during the Plio-Pleistocene in the central and northern North Island, whilst volcanic activity in the central North Island would have also caused fires and significant disturbance to landscapes due to ash fall (Battersby, Wilmshurst, Curran,

487 McGlone, & Perry, 2017; Kitzberger et al., 2016; Trewick & Bland, 2012; Wyse, Wilmshurst, Burns, &
488 Perry, 2018). *Leptospermum scoparium* is a seral species, known to establish in disturbed landscapes,
489 particularly after fire, an ancestral adaptation that evolved to enable it to survive forest fires in
490 Australia (Battersby et al., 2017). Serotinous *L. scoparium* populations are more common in the North
491 Island (correlating with the history of natural fires in New Zealand) and this may be one reason for
492 the success and expansion of NNI and CSNI gene pools in these two regions over the last ~350,000
493 years (Battersby et al., 2017).

494

495 For the two remaining New Zealand gene pools (ECNI, NESI), demographic modelling suggests N_e has
496 been decreasing over the last ~8,250 – 19,000 years. This timing coincides with the end of the Last
497 Glacial Maximum (LGM) (when global ice sheets were at their maximum integrated volume) (Clark et
498 al., 2009) and warming of the climate, suggesting that as glaciers, tussock grassland and alpine shrub-
499 land receded, so too did the distribution of mānuka in the north of the South Island and in the East
500 Cape in the North Island. For the northern South Island in particular, there would have been
501 considerable habitat loss at this time as sea levels rose, and the land bridge connecting the North and
502 South Islands was inundated (Alloway et al., 2007; Trewick & Bland, 2012), potentially explaining the
503 contraction of the NESI gene pool. Results of the demographic modelling also suggest that N_e of the
504 Australian gene pool has decreased significantly (~98.98% decrease) over the last ~17,000 – 23,000
505 years, however, our sampling of this gene pool is very limited (considering the distribution of *L.*
506 *scoparium* in Australia). Sampling across the complete distribution of *L. scoparium* in Australia is
507 required to draw any further conclusions about the evolutionary history of this gene pool.

508

509 It has been hypothesised, based on morphological data, that *L. scoparium* originated in Tasmania and
510 simultaneously dispersed to Australia and New Zealand (Thompson, 1989). Although we only have
511 one population from Tasmania, our results provide no evidence to support the hypothesis that *L.*

scoparium originated in Tasmania and dispersed to Australia and New Zealand simultaneously. Based on our findings, it is likely that *L. scoparium* originated on mainland Australia, and subsequently dispersed to Tasmania and then to New Zealand, with migration events also occurring from mainland Australia to New Zealand. Wider sampling within Australia and Tasmania would help to confirm this hypothesis, however similar conclusions were reached by Buys et al. (2019) and Stephens (2006) in their phylogenetic and chemotype studies, respectively. Similarly to those of Stephens (2006), our findings suggest that there have been multiple introductions of *L. scoparium* from Australian populations to New Zealand. Eastward, long-distance dispersal of flora (Breitwieser, Glenny, Thorne, & Wagstaff, 1999; Knapp et al., 2005; Lockhart et al., 2001; M. McGlone, Duncan, & Heenan, 2001; Winkworth, Wagstaff, Glenny, & Lockhart, 2002) and fauna (Goldberg, Morgan-Richards, & Trewick, 2015; Goldberg, Trewick, & Paterson, 2008; Koot, Morgan-Richards, & Trewick, 2020) from Australia to New Zealand is not uncommon, and is often attributed to the West Wind Drift, the Tasman Current and the Antarctic Circumpolar Current phenomenon of the Southern Hemisphere (formed following the separation of Antarctica from South America and then Australia in the mid Oligocene ~28-35Mya) (Sanmartíin & Ronquist, 2004; Sanmartín, Wanntorp, & Winkworth, 2007; Winkworth et al., 2002).

Conclusion

The findings from this study will support the commercialisation of mānuka honey based on its indigenous and geographic origins. Additionally, data produced in this study can contribute to the conservation management of the mānuka germplasm, ensuring genetic diversity is maintained in an industry where selective planting is occurring. Furthermore, the genetic structure of New Zealand mānuka as five different clusters, all distinct from one another and from Australian populations, could form the basis for a future reclassification and taxonomic definition of this species. Finally, knowledge of genome variants across its natural range can be used in future research to understand key traits

such as tolerance to the tree pathogen myrtle rust (*Austropuccinia psidii*) (Smith et al. 2020) and adaptation to climatic and environmental conditions.

Acknowledgements

We thank the Māori landowners who gave us the privilege to access plant materials for this study; Treigh Akuhata-Christy (Ngāti Porou), Trey Thomson (Ngāti Porou), Morgan Coleman, Paul Peterson, and Hamish Maule, Chris Morse, Rowan Buxton, Alex Fergus and Jessie Prebble from Manaaki Whenua Landcare Research for their help with sample collection. This research was supported by the New Zealand Ministry for Business Innovation and Employment project "Building resilience and provenance into an authentic Māori honey industry (the Honey Landscape)" (contract number 34213). We thank the Honey Landscape Māori project governance group, Aaron Taikato (Prime Holding Ltd), Shar Amner (Atihau-Whanganui), Huti Watson (Ngāti Porou Miere Ltd), Brenda Tahi (Tūhoe Tuawhenua Trust), Victoria Henstock (Ngāi Tahu) and Twyla McDonald (Tai Tokerau Honey Ltd) for their advice; Emily Buck, Cecilia Deng, John McCallum, Susan Thomson, Samantha Baldwin, Linley Jesson, Chris Kirk, Andrew Granger, Claire Hall, Jenny Green, Nicolas Bordes and Isabel Moller from The New Zealand Institute for Plant & Food Research Ltd. for their advice and assistance during the project design, sampling, data analysis and Māori engagement; Matthew Tinning and the Australian Genome Research Facility (AGRF) for the next-generation sequencing; Dr Phillip Wilcox (University of Otago), Holden Hohaia (Manaaki Whenua Landcare Research) and Maui Hudson (University of Waikato) for discussions about Māori engagement and data sovereignty.

References

Adams, C. J., Boult, C. H., Deadman, B. J., Farr, J. M., Grainger, M. N., Manley-Harris, M., & Snow, M. J. (2008). Isolation by HPLC and characterisation of the bioactive fraction of New Zealand manuka (Leptospermum scoparium) honey. *Carbohydrate research, 343*(4), 651-659.

Adams, C. J., Manley-Harris, M., & Molan, P. C. (2009). The origin of methylglyoxal in New Zealand manuka (Leptospermum scoparium) honey. *Carbohydrate research, 344*(8), 1050-1053.

563 Allan, H. H. (1961). Flora of New Zealand. Volume I. *Flora of New Zealand. Volume I.*(8 x 5 1/2).
564 Alloway, B. V., Lowe, D. J., Barrell, D. J., Newnham, R. M., Almond, P. C., Augustinus, P. C., . . . McGlone,
565 M. S. (2007). Towards a climate event stratigraphy for New Zealand over the past 30 000 years
566 (NZ-INTIMATE project). *Journal of Quaternary Science, 22*(1), 9-35.
567 Anderson, E. C., Skaug, H. J., & Barshis, D. J. (2014). Next-generation sequencing for molecular ecology:
568 a caveat regarding pooled samples. *Molecular ecology, 23*(3), 502-512.
569 Andrews, S. (2016). FastQC: a quality control tool for high throughput sequence data. 2010. In.
570 Atrott, J., Haberlau, S., & Henle, T. (2012). Studies on the formation of methylglyoxal from
571 dihydroxyacetone in Manuka (Leptospermum scoparium) honey. *Carbohydrate research, 361*,
572 7-11.
573 Battersby, P. F., Wilmshurst, J. M., Curran, T. J., McGlone, M. S., & Perry, G. L. (2017). Exploring fire
574 adaptation in a land with little fire: serotiny in Leptospermum scoparium (Myrtaceae). *Journal*
575 *of biogeography, 44*(6), 1306-1318.
576 Bean, A. (1992). The genus Leptospermum Forst. et Forst. F.(Myrtaceae) in northern Australia and
577 Malesia. *Austrobaileya*, 643-659.
578 Bean, A. (2004). Three new species of Leptospermum (Myrtaceae) from Queensland and northern
579 New South Wales. *Telopea, 10*(4), 831-838.
580 Beattie, H. (1920). *Nature-lore of the southern Maori*: New Zealand Institute.
581 Bertelsen, H., Gregersen, V., Poulsen, N., Nielsen, R. O., Das, A., Madsen, L. B., . . . Larsen, L. B. (2016).
582 Detection of genetic variation affecting milk coagulation properties in Danish Holstein dairy
583 cattle by analyses of pooled whole-genome sequences from phenotypically extreme samples
584 (pool-seq). *Journal of animal science, 94*(4), 1365-1376.
585 Best, E. (1904). Maori medical lore. Notes on sickness and disease among the Maori people of New
586 Zealand, and their treatment of the sick; together with some account of various beliefs,
587 superstitions and rites pertaining to sickness, and the treatment thereof, as collected from
588 the Tuhoe tribe. Part I. *The Journal of the Polynesian Society, 13*(4 (52), 213-237.
589 Best, E. (1929). *Fishing methods and devices of the Maori*: Ams Pr Inc.
590 Bhatia, D., Wing, R., & Singh, K. (2013). Genotyping by sequencing, its implications and benefits. *Crop*
591 *improv, 40*(2), 101-111.
592 Breitwieser, I., Glenny, D. S., Thorne, A., & Wagstaff, S. J. (1999). Phylogenetic relationships in
593 Australasian Gnaphalieae (Compositae) inferred from ITS sequences. *New Zealand Journal of*
594 *Botany, 37*(3), 399-412.
595 Buys, M. H., Winkworth, R. C., de Lange, P. J., Wilson, P. G., Mitchell, N., Lemmon, A. R., . . . Klápště, J.
596 (2019). The phylogenomics of diversification on an island: applying anchored hybrid
597 enrichment to New Zealand Leptospermum scoparium (Myrtaceae). *Botanical Journal of the*
598 *Linnean Society*.
599 Cheeseman, T. F. (1925). VI. The flora of Rarotonga, the chief Island of the Cook group. *Transactions*
600 *of the Linnean Society of London. 2nd Series. Botany, 6*(6), 261-313.
601 Christe, C., Stölting, K. N., Paris, M., Fraïsse, C., Bierne, N., & Lexer, C. (2017). Adaptive evolution and
602 segregating load contribute to the genomic landscape of divergence in two tree species
603 connected by episodic gene flow. *Molecular ecology, 26*(1), 59-76.
604 Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., . . . McCabe, A. M. (2009).
605 The last glacial maximum. *Science, 325*(5941), 710-714.
606 Colenso, W. (1868). On the Geographical and Economic Botany of the North Island of New Zealand.
607 *Transaction of the New Zealand Institute, 1*(2), 1-58.
608 Corander, J., Majander, K. K., Cheng, L., & Merilä, J. (2013). High degree of cryptic population
609 differentiation in the B altic S ea herring C lupea harengus. *Molecular ecology, 22*(11), 2931-
610 2940.
611 Crowe, A. (2001). A field guide to the native edible plants of New Zealand, 3rd edn. Birckenhead. In:
612 Auckland: Godwit Publishing Limited.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., & Nielsen, R. (2016). SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics, 32*(12), 1895-1897.

Dennenmoser, S., Vamosi, S. M., Nolte, A. W., & Rogers, S. M. (2017). Adaptive genomic divergence under high gene flow between freshwater and brackish-water ecotypes of prickly sculpin (Cottus asper) revealed by Pool-Seq. *Molecular ecology, 26*(1), 25-42.

Douglas, M. H., van Klink, J. W., Smallfield, B. M., Perry, N. B., Anderson, R. E., Johnstone, P., & Weavers, R. T. (2004). Essential oils from New Zealand manuka: triketone and other chemotypes of Leptospermum scoparium. *Phytochemistry, 65*(9), 1255-1264.

Endler, L., Betancourt, A. J., Nolte, V., & Schlötterer, C. (2016). Reconciling differences in pool-GWAS between populations: a case study of female abdominal pigmentation in Drosophila melanogaster. *Genetics, 202*(2), 843-855.

Esler, A. E., & Astridge, S. J. (1974). Tea tree (Leptospermum) communities of the Waitakere Range, Auckland, New Zealand. *New Zealand Journal of Botany, 12*(4), 485-501.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS genetics, 9*(10), e1003905.

Ferretti, L., Ramos-Onsins, S. E., & Pérez-Enciso, M. (2013). Population genomics from pool sequencing. *Molecular ecology, 22*(22), 5561-5576.

Fischer, M. C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K. K., . . . Widmer, A. (2013). Population genomic footprints of selection and associations with climate in natural populations of Arabidopsis halleri from the Alps. *Molecular ecology, 22*(22), 5594-5607.

Fracassetti, M., Griffin, P. C., & Willi, Y. (2015). Validation of pooled whole-genome re-sequencing in Arabidopsis lyrata. *PloS one, 10*(10).

Futschik, A., & Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics, 186*(1), 207-218.

Gardiner, S., Bassett, H., Noiton, D., Bus, V., Hofstee, M., White, A., . . . Rikkerink, E. (1996). A detailed linkage map around an apple scab resistance gene demonstrates that two disease resistance classes both carry the V f gene. *Theoretical and Applied Genetics, 93*(4), 485-493.

Goldberg, J., Morgan-Richards, M., & Trewick, S. A. (2015). Intercontinental island hopping: Colonization and speciation of the grasshopper genus Phaulacridium (Orthoptera: Acrididae) in Australasia. *Zoologischer Anzeiger-A Journal of Comparative Zoology, 255*, 71-79.

Goldberg, J., Trewick, S. A., & Paterson, A. M. (2008). Evolution of New Zealand's terrestrial fauna: a review of molecular evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1508), 3319-3334.

Grey, G. (1904). *Maori lore: The traditions of the Maori people, with the more important of their legends*: By authority, J. Mackay.

Gudgeon, W. E. (1905). Maori religion. *The Journal of the Polynesian Society, 14*(3 (55), 107-130.

Heenan, P. B., & McGlone, M. S. (2019). Cenozoic formation and colonisation history of the New Zealand vascular flora based on molecular clock dating of the plastid rbcL gene. *New Zealand Journal of Botany, 57*(4), 204-226.

Hiroa, T. R. (1949). The coming of the Maori. *Maori Purposes Fund Board. Christchurch, New Zealand: Whitcombe & Tombs*.

Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from Pool-seq data. *Genetics, 210*(1), 315-330.

Hudson, M., Nanibaa'A, G., Sterling, R., Caron, N. R., Fox, K., Yracheta, J., . . . Brown, A. (2020). Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data. *Nature Reviews Genetics*, 1-8.

Huson, D. H., & Bryant, D. (2005). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution, 23*(2), 254-267.

663 Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers.
664 *Bioinformatics, 24*(11), 1403-1405.
665 Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new
666 method for the analysis of genetically structured populations. *BMC genetics, 11*(1), 94.
667 Kamp, P. J. (1986). Late Cretaceous-Cenozoic tectonic development of the southwest Pacific region.
668 *Tectonophysics, 121*(2-4), 225-251.
669 Kamp, P. J., Green, P. F., & White, S. H. (1989). Fission track analysis reveals character of collisional
670 tectonics in New Zealand. *Tectonics, 8*(2), 169-195.
671 Kitzberger, T., Perry, G., Paritsis, J., Gowda, J., Tepley, A., Holz, A., & Veblen, T. (2016). Fire–vegetation
672 feedbacks and alternative states: common mechanisms of temperate forest vulnerability to
673 fire in southern South America and New Zealand. *New Zealand Journal of Botany, 54*(2), 247-
674 272.
675 Knapp, M., Stöckler, K., Havell, D., Delsuc, F., Sebastiani, F., & Lockhart, P. J. (2005). Relaxed molecular
676 clock provides evidence for long-distance dispersal of Nothofagus (southern beech). *PLoS*
677 *biology, 3*(1).
678 Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format
679 data in R. *Molecular ecology resources, 17*(1), 44-53.
680 Koot, E. M., Morgan-Richards, M., & Trewick, S. A. (2020). An alpine grasshopper radiation older than
681 the mountains, on Kā Tiritiri o te Moana (Southern Alps) of Aotearoa (New Zealand). *Molecular*
682 *Phylogenetics and Evolution*, 106783.
683 Lamb, S., Mortimer, N., Smith, E., & Turner, G. (2016). Focusing of relative plate motion at a
684 continental transform fault: Cenozoic dextral displacement> 700 km on New Zealand's Alpine
685 Fault, reversing> 225 km of Late Cretaceous sinistral motion. *Geochemistry, Geophysics,*
686 *Geosystems, 17*(3), 1197-1213.
687 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods,*
688 *9*(4), 357.
689 Leblois, R., Gautier, M., Rohfritsch, A., Foucaud, J., Burban, C., Galan, M., . . . Gharbi, K. (2018).
690 Deciphering the demographic history of allochronic differentiation in the pine processionary
691 moth Thaumetopoea pityocampa. *Molecular ecology, 27*(1), 264-278.
692 Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and
693 population genetical parameter estimation from sequencing data. *Bioinformatics, 27*(21),
694 2987-2993.
695 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The sequence
696 alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.
697 Lockhart, P. J., McLenachan, P. A., Havell, D., Glenny, D., Huson, D., & Jensen, U. (2001). Phylogeny,
698 radiation, and transoceanic dispersal of New Zealand alpine buttercups: molecular evidence
699 under split decomposition. *Annals of the Missouri Botanical Garden*, 458-477.
700 Lynch, M., Bost, D., Wilson, S., Maruki, T., & Harrison, S. (2014). Population-genetic inference from
701 pooled-sequencing data. *Genome biology and evolution, 6*(5), 1210-1218.
702 Lyne, A. (1993). Leptospermum namadgiensis (Myrtaceae), a new species from the Australian Capital
703 Territory–New South Wales border area. *Telopea, 5*(2), 319-324.
704 Lyne, A., & Crisp, M. (1996). Leptospermum jingera (Myrtaceae–Leptospermoideae): a new species
705 from north-eastern Victoria. *Australian Systematic Botany, 9*(3), 301-306.
706 Mavric, E., Wittmann, S., Barth, G., & Henle, T. (2008). Identification and quantification of
707 methylglyoxal as the dominant antibacterial constituent of Manuka (Leptospermum
708 scoparium) honeys from New Zealand. *Molecular nutrition & food research, 52*(4), 483-489.
709 McCulloch, G. A., Wallis, G. P., & Waters, J. M. (2010). Onset of glaciation drove simultaneous vicariant
710 isolation of alpine insects in New Zealand. *Evolution: International Journal of Organic*
711 *Evolution, 64*(7), 2033-2043.
712 McGlone, M. (1985). Plant biogeography and the late Cenozoic history of New Zealand. *New Zealand*
713 *Journal of Botany, 23*(4), 723-749.

McGlone, M., Duncan, R., & Heenan, P. (2001). Endemism, species selection and the origin and distribution of the vascular plant flora of New Zealand. *Journal of biogeography, 28*(2), 199-216.

McGlone, M. S., Newnham, R. M., & Moar, N. T. (2010). The vegetation cover of New Zealand during the Last Glacial Maximum: Do pollen records under-represent woody vegetation. *Terra Australis, 32*, 49-68.

Mimee, B., Duceppe, M. O., Véronneau, P. Y., Lafond-Lapalme, J., Jean, M., Belzile, F., & Bélair, G. (2015). A new method for studying population genetics of cyst nematodes based on Pool-Seq and genomewide allele frequency analysis. *Molecular ecology resources, 15*(6), 1356-1365.

Ministry for Primary Industries. (2018). *Apiculture: Ministry for Primary Industries 2018 apiculture monitoring programme*. Retrieved from Wellington, New Zealand:

Morgan, E. R., Perry, N. B., & Chagné, D. (2019). Science at the intersection of cultures–Māori, Pākehā and mānuka. *New Zealand Journal of Crop and Horticultural Science, 47*(4), 225-232.

Palmer, J. (1957). The Maori Kotaha. *The Journal of the Polynesian Society, 66*(2), 175-191.

Patterson, M., McKay, R., Naish, T., Escutia, C., Jimenez-Espejo, F., Raymo, M., . . . Expedition, I. (2014). Orbital forcing of the East Antarctic ice sheet during the Pliocene and Early Pleistocene. *Nature Geoscience, 7*(11), 841.

Perry, N. B., Brennan, N. J., Van Klink, J. W., Harris, W., Douglas, M. H., McGimpsey, J. A., . . . Anderson, R. E. (1997). Essential oils from New Zealand manuka and kanuka: chemotaxonomy of Leptospermum. *Phytochemistry, 44*(8), 1485-1494.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS genetics, 8*(11), e1002967.

Porter, N. G., & Wilkins, A. L. (1999). Chemical, physical and antimicrobial properties of essential oils of Leptospermum scoparium and Kunzea ericoides. *Phytochemistry, 50*(3), 407-415.

Prebble, J. G., Reichgelt, T., Mildenhall, D. C., Greenwood, D. R., Raine, J. I., Kennedy, E. M., & Seebeck, H. C. (2017). Terrestrial climate evolution in the Southwest Pacific over the past 30 million years. *Earth and Planetary Science Letters, 459*, 136-144.

R Core Team. (2013). R: A language and environment for statistical computing.

Rellstab, C., Zoller, S., Tedder, A., Gugerli, F., & Fischer, M. C. (2013). Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species. *PloS one, 8*(11).

Riley, M. (1994). *Māori healing and herbal: New Zealand ethnobotanical sourcebook*: Viking Sevenseas NZ.

Roda, F., Ambrose, L., Walter, G. M., Liu, H. L., Schaul, A., Lowe, A., . . . Ortiz-Barrientos, D. (2013). Genomic evidence for the parallel evolution of coastal forms in the Senecio lautus complex. *Molecular ecology, 22*(11), 2941-2952.

Ronghua, Y., Mark, A., & Wilson, J. (1984). Aspects of the ecology of the indigenous shrub Leptospermum scoparium (Myrtaceae) in New Zealand. *New Zealand Journal of Botany, 22*(4), 483-507.

Sanmartíin, I., & Ronquist, F. (2004). Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic biology*, 216-243.

Sanmartín, I., Wanntorp, L., & Winkworth, R. C. (2007). West Wind Drift revisited: testing for directional dispersal in the Southern Hemisphere using event-based tree fitting. *Journal of biogeography, 34*(3), 398-416.

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals—mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics, 15*(11), 749.

Stephens, J., Molan, P. C., & Clarkson, B. D. (2005). A review of Leptospermum scoparium (Myrtaceae) in New Zealand. *New Zealand Journal of Botany, 43*(2), 431-449.

Stephens, J. M. C. (2006). *The factors responsible for the varying levels of UMF® in mānuka (Leptospermum scoparium) honey.* The University of Waikato,

Thompson, J. (1989). revision of the genus Leptospermum (Myrtaceae). *Telopea*.

765 Thornhill, A. H., Ho, S. Y., Külheim, C., & Crisp, M. D. (2015). Interpreting the modern distribution of
766 Myrtaceae using a dated molecular phylogeny. *Molecular Phylogenetics and Evolution, 93*, 29-
767 43.
768 Thrimawithana, A. H., Jones, D., Hilario, E., Grierson, E., Ngo, H. M., Liachko, I., . . . Bicknell, R. (2019).
769 A whole genome assembly of Leptospermum scoparium (Myrtaceae) for mānuka research.
770 *New Zealand Journal of Crop and Horticultural Science*, 1-28.
771 Trewick, S., & Bland, K. (2012). Fire and slice: palaeogeography for biogeography at New Zealand's
772 North Island/South Island juncture. *Journal of the Royal Society of New Zealand, 42*(3), 153-
773 183.
774 Tung, P.-Y., Blischak, J. D., Hsiao, C. J., Knowles, D. A., Burnett, J. E., Pritchard, J. K., & Gilad, Y. (2017).
775 Batch effects and the effective design of single-cell gene expression studies. *Scientific reports,*
776 *7*, 39921.
777 Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T., & Nuzhdin, S. V. (2010). Population
778 resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nature*
779 *genetics, 42*(3), 260.
780 Waitangi Tribunal. (2011). Ko Aotearoa tēnei : a report into claims concerning New Zealand law and
781 policy affecting Māori culture and identity: Te taumata tuatahi (Wai 262). *Waitangi Tribunal*
782 *Report, Wellington, New Zealand*.
783 Williams, S., King, J., Revell, M., Manley-Harris, M., Balks, M., Janusch, F., . . . Dawson, M. (2014).
784 Regional, annual, and individual variations in the dihydroxyacetone content of the nectar of
785 manuka (Leptospermum scoparium) in New Zealand. *Journal of agricultural and food*
786 *chemistry, 62*(42), 10332-10340.
787 Winkworth, R. C., Wagstaff, S. J., Glenny, D., & Lockhart, P. J. (2002). Plant dispersal news from New
788 Zealand. *Trends in ecology & evolution, 17*(11), 514-520.
789 Wyse, S. V., Wilmshurst, J. M., Burns, B. R., & Perry, G. L. (2018). New Zealand forest dynamics: a
790 review of past and present vegetation responses to disturbance, and development of
791 conceptual forest models. *New Zealand Journal of Ecology, 42*(2), 87-106.

792

793 **Data Accessibility**

794 Permission from representatives of the Indigenous Peoples (Māori) was obtained for using the plant
795 material used for this study. Further studies using this material, raw sequencing data and final genome
796 assembly will require consent from the Māori iwi (tribe) and landowners who exercises guardianship
797 for this material according to Aotearoa New Zealand's Treaty of Waitangi and the international Nagoya
798 protocol on the rights of indigenous peoples. Access to raw and analyzed data will require permission
799 from representatives of the iwi.

800

801 **Author Contribution**

802 Emily Koot: performed research, analysed data, wrote the paper. Elise Arnst: collected samples, wrote
803 the paper. Melissa Taane: extracted DNA. Kelsey Goldsmith: collected samples, extracted DNA. Peri
804 Tobias: collected samples, extracted DNA. Eleanor Dormont: collected samples. Tate Hancox:

805 collected samples. Kate Delaporte: collected samples, Amali Thrimawithana: analysed data. Kiri
806 Reihana: collected samples. Santiago C. González-Martínez: analysed data, wrote the paper. Victor
807 Goldsmith: coordination of the access to plant material with Maori landowners. Gary Houliston:
808 designed research, wrote the paper. David Chagné: designed research, analysed data, wrote the
809 paper.

810

Tables and Figures

Table 1. Metadata for 76 sampling sites of *Leptospermum scoparium* and *Kunzea robusta* across New Zealand and Australia. Exact locations of sampling sites (e.g. latitudes and longitudes) are kept anonymous to protect valuable indigenous assets as agreed with the Māori landowners who granted access to their land for sampling. Ownership identifies if samples were taken from private, public, council or Department of Conservation (DOC) land. Iwi identifies the Iwi (tribal group) associated with each sampling site in New Zealand. Pool size is the number of samples for which DNA was successfully extracted and pooled for each site. DAPC cluster is the cluster to which the Kmeans clustering analysis determined each population belonged, whilst Gene Pool is the gene pool name (and geographic region) associated with each cluster: NNI = northern North Island, CSNI = central and southern North Island, ECNI = East Cape North Island, NESI = north eastern South Island, SWSI = south western South Island, Australia = Australian (including Tasmania). NZ = New Zealand.

Population	Country	Region	Ownership	Iwi	Pool_size	DAPC Cluster	Gene pool
LC	Australia	Victoria	Public	-	15	6	Australia
LS	Australia	Victoria	Public	-	28	6	Australia
Pool_1	Australia	Victoria	Public	-	25	6	Australia
Pool_2	Australia	Victoria	Public	-	30	6	Australia
TLs	Australia	Tasmania	Public	-	15	6	Australia
VLs	Australia	Victoria	Public	-	15	6	Australia
P001	NZ	Manawatu	Private	Ngāti Kahungunu, Rangitāne, Ngāti Raukawa ki te Tonga, Ngāti Toa Rangatira	27	3	CSNI
P002	NZ	Bay of Plenty	Private	Tūhoe, Ngāti Kahungunu, Te Wairoa iwi and hapu	30	3	CSNI
P006	NZ	Bay of Plenty	Private	Tūhoe, Ngāti Kahungunu	29	3	CSNI
P009	NZ	Hawke's Bay	Private	Tūhoe, Ngāti Kahungunu, Ruapani ki Waikaremoana, Te Wairoa iwi and hapu	26	3	CSNI
P010	NZ	Bay of Plenty	Private	Tūhoe, Ngāti Kahungunu	29	3	CSNI
P012	NZ	Waikato	Private	Maniapoto	30	3	CSNI
P013	NZ	Waikato	Private	Maniapoto, Ngāti Tama	27	3	CSNI
P014	NZ	Waikato	Private	Maniapoto, Raukawa	26	3	CSNI
P015	NZ	Manawatu-Whanganui	Private	Whanganui Iwi / Te Atihaunui a Pāpārangi, Te Korowai o Wainuiārua (Central Whanganui), Whanganui Land Settlement (Lower Whanganui)	29	3	CSNI

P016	NZ	Manawatu-Whanganui	Private	Whanganui Iwi / Te Atihaunui a Pāpārangi, Ngāti Rangi, Te Korowai o Wainuiārua (Central Whanganui)	30	3	CSNI
P017	NZ	Manawatu-Whanganui	Private	Whanganui Iwi / Te Atihaunui a Pāpārangi, Ngāti Rangi, Te Korowai o Wainuiārua (Central Whanganui)	27	3	CSNI
P018	NZ	Bay of Plenty	Private	Tūhoe, Ngāti Kahungunu	28	3	CSNI
P019	NZ	Hawke's Bay	Private	Ngāti Kahungunu, Te Wairoa iwi and hapu	28	3	CSNI
P023	NZ	Hawke's Bay	Private	Tūhoe, Ngāti Kahungunu, Ngāti Pāhauwera, Te Wairoa iwi and hapu	29	3	CSNI
P026	NZ	Gisborne	Private	Ngāi Tāmanuhiri, Te Wairoa iwi and hapu	29	3	CSNI
P027	NZ	Gisborne	Private	Ngāi Tāmanuhiri, Ngāti Kahungunu, Te Wairoa iwi and hapu	30	3	CSNI
P029	NZ	Hawke's Bay	Private	Ngāti Kahungunu, Ngāti Pāhauwera	28	3	CSNI
P033	NZ	Waikato	Private	Ngāti Tūwharetoa	28	3	CSNI
P045	NZ	Manawatu-Whanganui	Private	Ngāti Hauiti	30	3	CSNI
P050	NZ	Waikato	Private	Ngāti Hako, Ngāti Hei, Ngāti Maru (Hauraki), Ngāti Rāhiri Tumutumu, Ngāi Tai ki Tāmaki, Ngāti Tamaterā, Ngāti Whanaunga	26	3	CSNI
P065	NZ	Taranaki	Private	Ngāti Ruanui, Ngā Rauru Kītahi	30	3	CSNI
P069	NZ	Wellington	DOC	Ngāti Kahungunu, Rangitāne, Ngāti Kahungunu ki Wairarapa - Tāmaki Nui ā Rua	30	3	CSNI
P070	NZ	Manawatu-Whanganui	DOC	Ngāti Kahungunu, Rangitāne, Ngāti Kahungunu ki Wairarapa - Tāmaki Nui ā Rua	30	3	CSNI
P071	NZ	Manawatu-Whanganui	DOC	Ngāti Kahungunu, Rangitāne, Ngāti Kahungunu ki Wairarapa - Tāmaki Nui ā Rua	30	3	CSNI

P073	NZ	Bay of Plenty	Private	Te Ure o Uenukukōpako / Ngāti Whakaue	30	3	CSNI
P005	NZ	Gisborne	Private	Ngāti Porou	29	1	ECNI
P025	NZ	Gisborne	Private	Ngāti Porou	29	1	ECNI
P028	NZ	Gisborne	Private	Ngāti Porou	29	1	ECNI
P030	NZ	Gisborne	Private	Ngāti Porou	28	1	ECNI
P031	NZ	Gisborne	Private	Ngāti Porou	30	1	ECNI
P044	NZ	Gisborne	Private	Ngāti Porou	26	1	ECNI
P046	NZ	Bay of Plenty	Private	Te Whānau a Apanui	25	1	ECNI
P008	NZ	Gisborne	Private	Ngāti Porou	26	5	Kānuka
P011	NZ	Bay of Plenty	Private	Ngāti Manawa, Ngāti Whare, Tūhoe, Ngāti Kahungunu	24	5	Kānuka
P020	NZ	Canterbury	Private	Ngāi Tahu	30	7	NESI
P021	NZ	Canterbury	Private	Ngāi Tahu	30	7	NESI
P022	NZ	Canterbury	Private	Ngāi Tahu	30	7	NESI
P034	NZ	Tasman	Private	Ngāti Toa Rangatira, Te Atiawa o Te Waka-a-Māui, Ngāti Apa ki te Rā Tō, Rangitāne o Wairau, Ngāti Kuia, Ngāti Rārua, Ngāti Kōata, Ngāti Tama ki Te Tau Ihu	23	7	NESI
P035	NZ	Tasman	Private	Ngāti Toa Rangatira, Te Atiawa o Te Waka-a-Māui, Ngāti Apa ki te Rā Tō, Rangitāne o Wairau, Ngāti Kuia, Ngāti Rārua, Ngāti Kōata, Ngāti Tama ki Te Tau Ihu	25	7	NESI
P036	NZ	Marlborough	Private	Ngāti Toa Rangatira, Te Atiawa o Te Waka-a-Māui, Rangitāne o Wairau, Ngāti Kuia, Ngāti Rārua, Ngāti Kōata, Ngāti Tama ki Te Tau Ihu	26	7	NESI
P048	NZ	Canterbury	DOC	Ngāi Tahu	30	7	NESI
P052	NZ	Marlborough	DOC	Ngāti Toa Rangatira, Ngāi Tahu	30	7	NESI

P054	NZ	Tasman	Private	Ngāti Toa Rangatira, Te Atiawa o Te Waka-a-Māui, Ngāti Apa ki te Rā Tō, Ngāti Kuia, Ngāti Rārua, Ngāti Kōata, Ngāti Tama ki Te Tau Ihu	30	7	NESI
P055	NZ	Tasman	Private	Ngāti Toa Rangatira, Te Atiawa o Te Waka-a-Māui, Ngāti Apa ki te Rā Tō, Ngāti Kuia, Ngāti Rārua, Ngāti Kōata, Ngāti Tama ki Te Tau Ihu	27	7	NESI
P056	NZ	Tasman	Private	Ngāti Toa Rangatira, Te Atiawa o Te Waka-a-Māui, Ngāti Apa ki te Rā Tō, Ngāti Kuia, Ngāti Rārua, Ngāti Kōata, Ngāti Tama ki Te Tau Ihu	29	7	NESI
P072	NZ	Marlborough	DOC	Ngāti Toa Rangatira, Te Atiawa o Te Waka-a-Māui, Rangitāne o Wairau, Ngāti Kuia, Ngāti Kōata	30	7	NESI
P003	NZ	Northland	Private	Ngāpuhi, Ngāti Whātua, Te Roroa	30	4	NNI
P004	NZ	Northland	Private	Ngāpuhi	28	4	NNI
P007	NZ	Northland	Private	Ngāpuhi, Ngātiwai	29	4	NNI
P024	NZ	Northland	Private	Ngāpuhi, Ngāti Whātua, Te Roroa	21	4	NNI
P037	NZ	Northland	DOC	Ngāti Kahu	29	4	NNI
P038	NZ	Northland	Private	Te Rarawa	25	4	NNI
P039	NZ	Northland	Private	Te Rarawa	26	4	NNI
P040	NZ	Northland	DOC	Ngāi Takoto, Ngāti Kuri, Te Aupōuri	28	4	NNI
P041	NZ	Ahipara	DOC	Te Rarawa	29	4	NNI
P043	NZ	Northland	Private	Ngāpuhi, Ngātiwai	27	4	NNI
P066	NZ	Waikato	DOC	Ngāti Hako, Ngāti Hei, Ngāti Maru (Hauraki), Te Patukirikiri, Ngāti Rāhiri Tumutumu, Ngāi Tai ki Tāmaki, Ngāti Tamaterā, Ngāti Whanaunga	28	4	NNI
P067	NZ	Auckland	Council	Ngāti Whātua, Ngāti Whātua o Ōrākei, Te Kawerau a Maki, Ngāti Tamaoho, Te Ākitai Waiohua	29	4	NNI
P068	NZ	Auckland	Council	Ngāti Whātua, Te Uri o Hau	30	4	NNI

P047	NZ	West Coast	DOC	Ngāi Tahu	27	2	SWSI
P049	NZ	West Coast	DOC	Ngāti Rārua, Ngāi Tahu	30	2	SWSI
P051	NZ	Southland	DOC	Ngāi Tahu	28	2	SWSI
P057	NZ	West Coast	DOC	Ngāti Toa Rangatira, Te Atiawa o Te Waka-a-Māui, Ngāti Apa ki te Rā Tō, Ngāti Rārua, Ngāi Tahu	30	2	SWSI
P058	NZ	Otago	DOC	Ngāi Tahu	30	2	SWSI
P059	NZ	Southland	DOC	Ngāi Tahu	26	2	SWSI
P060	NZ	Otago	DOC	Ngāi Tahu	30	2	SWSI
P061	NZ	Southland	DOC	Ngāi Tahu	29	2	SWSI
P062	NZ	Otago	DOC	Ngāi Tahu	29	2	SWSI
P063	NZ	West Coast	DOC	Ngāti Rārua, Ngāi Tahu	22	2	SWSI
P064	NZ	Canterbury	Private	Ngāi Tahu	29	2	SWSI

819

820 **Table 2. Whole genome variant detection in *Leptospermum scoparium* and *Kunzea robusta* using pool**
821 **sequencing.** Seven different datasets are shown, based on including *Kunzea robusta* (kānuka) or not, including
822 Australian populations or not and with minor allele frequencies (MAF) of 2% and 5% applied. Variant filtering
823 was performed with VCFtools. Only single nucleotide polymorphisms (SNPs) with no missing data and a coverage
824 greater than 100X were included (filtered SNPs).

Number of populations	MAF	Filtered SNPs	Filtered SNP density
76 populations (All)	0%	5,503,881	45.21
76 populations (All)	2%	2,513,694	98.99
76 populations (All)	5%	1,498,405	166.07
74 populations (No Kānuka)	2%	4,049,649	61.45
74 populations (No Kānuka)	5%	2,580,451	96.43
68 populations (NZ only)	2%	3,270,864	76.08
68 populations (NZ only)	5%	2,526,589	98.49

825

Table 3. Average pairwise *F*st for each gene pool with standard deviations in brackets. Matrix of pairwise *F*st values for all populations can be found in Supplementary material 5. NNI = northern North Island, CSNI = central and southern North Island, ECNI = East Cape North Island, NESI = north eastern South Island, SWSI = south western South Island, NZ = New Zealand.

	Australia	**Tasmania**	**NNI**	**CSNI**	**ECNI**	**NESI**	**SWSI**	**NZ**
Australia	0.187 (0.093)							
Tasmania	0.302 (0.006)	-						
NNI	0.363 (0.021)	0.388 (0.011)	0.049 (0.034)					
CSNI	0.356 (0.018)	0.382 (0.007)	0.143 (0.040)	0. 048 (0.024)				
ECNI	0.377 (0.017)	0.399 (0.006)	0.172 (0.036)	0.079 (0.023)	0.024 (0.016)			
NESI	0.357 (0.019)	0.384 (0.009)	0.184 (0.028)	0.165 (0.021)	0.196 (0.019)	0.073 (0.024)		
SWSI	0.346 (0.019)	0.373 (0.010)	0.178 (0.030)	0.146 (0.026)	0.175 (0.027)	0.106 (0.024)	0.069 (0.030)	
NZ	0.357 (0.020)	0.383 (0.011)	-	-	-	-	-	0.128 (0.057)

Table 4. Summary statistics averaged by gene pool of population genetic parameters calculated for using NPStats. S = segregating sites, π = nucleotide diversity, SD = standard deviation. Individual summary statistics for each populations can be found in Supplementary material 6.

Region	S(SD)	Watterson's θ (SD)	π (SD)	Tajima D (SD)
Australia	118.83 (92.80)	0.0123 (0.0022)	0.0108 (0.0012)	-0.514 (0.3626)
NNI	44.5 (9.33)	0.0087 (0.0012)	0.0069 (0.0005)	-0.905 (0.1424)
CSNI	44.89 (7.47)	0.0089 (0.0013)	0.0074 (0.0008)	-0.811 (0.1515)
ECNI	45.28 (13.86)	0.0087 (0.0017)	0.0072 (0.0013)	-0.765 (0.1472)
NESI	43.36 (4.00)	0.0076 (0.0003)	0.0069 (0.0002)	-0.522 (0.0793)
SWSI	43.17 (5.68)	0.0083 (0.0009)	0.0072 (0.0014)	-0.648 (0.1640)
Kānuka	28.00 (11.00)	0.0104 (0.0046)	0.0081 (0.0035)	-0.849 (0.625)

Table 5a. Maximum likelihood and AIC statistics for the individual demographic models tested. The model considered to be the best fit is in bold. These models were then run 100x (See Supplementary material 7). DeltaL = the difference between the maximum estimated likelihood and the maximum observed likelihood. AIC = Akaike's Information Criterion. Note: in some instances the model with the lowest AIC was not always selected as the best model, as the parameter estimates were not sensible (see Results section). NNI = northern North Island, CSNI = central and southern North Island, ECNI = East Cape North Island, NESI = north eastern South Island, SWSI = south western South Island.

Gene pool(s)	Model	No. of parameters	Maximum estimated likelihood	deltaL	AIC
NNI	Stable	2	-4624575	316784	21296959
	Bottleneck	6	-4498474	68461	20716252
	Expansion	**4**	**-4498275**	**190484**	**20715329**
	Contraction	4	-4620187	312395	21276754
CSNI	Stable	2	-5222475	492068	24050390
	Bottleneck	6	-6627286	458419	30519790
	Expansion	**4**	**-5122746**	**392339**	**23591125**
	Contraction	4	-5177958	447551	23845384
ECNI	Stable	2	-4880280	350141	22474523
	Bottleneck	**6**	**-4817860**	**287721**	**22187079**
	Expansion	4	-4845841	315702	22315933
	Contraction	**4**	**-4823543**	**293404**	**22213246**
NESI	Stable	2	-3497510	70215	16106634
	Bottleneck	**6**	**-3483627**	**56332**	**16042707**
	Expansion	4	-3490089	62794	16072461
	Contraction	**4**	**-3484327**	**57032**	**16045928**
SWSI	Stable	2	-4602161	172147	21193737
	Bottleneck	6	-6068501	117142	27946491
	Expansion	**4**	**-4497117**	**67103**	**20709997**
	Contraction	4	-4598972	168958	21179058
Australia	Stable	2	-5501975	427386	25337533

	Bottleneck	**6**	**-5305720**	**231132**	**24433757**
	Expansion	4	-5503242	428654	25343375
	Contraction	**4**	**-5308952**	**234363**	**24448635**

842

843

844 **Table 5b. Maximum likelihood and AIC statistics for the demographic models used to investigate the**
845 **divergence of Australian and New Zealand *Leptospermum scoparium* gene pools.** The model considered to be
846 the best fit is in bold. These models were then run 100x (See Supplementary material 7). DeltaL = the difference
847 between the maximum estimated likelihood and the maximum observed likelihood. AIC = Akaike's Information
848 Criterion.

Model description	Model ID	No. of parameters	Maximum estimated likelihood	deltaL	AIC
Divergence followed by isolation	D-I	5	-12083216	923326	55645277
Divergence with continuous gene flow	D-CGF	7	-11874074	714183	54682144
Divergence with ancestral gene flow	D-AGF-I	8	-11878674	718783	54703330
Divergence with recent gene flow	D-RGF	8	-11859600	699709	54615490
Divergence with continuous gene flow, Australia contracting& New Zealand expanding	**D-CGF-CE**	**11**	**-11622817**	**462926**	**53525072**

849

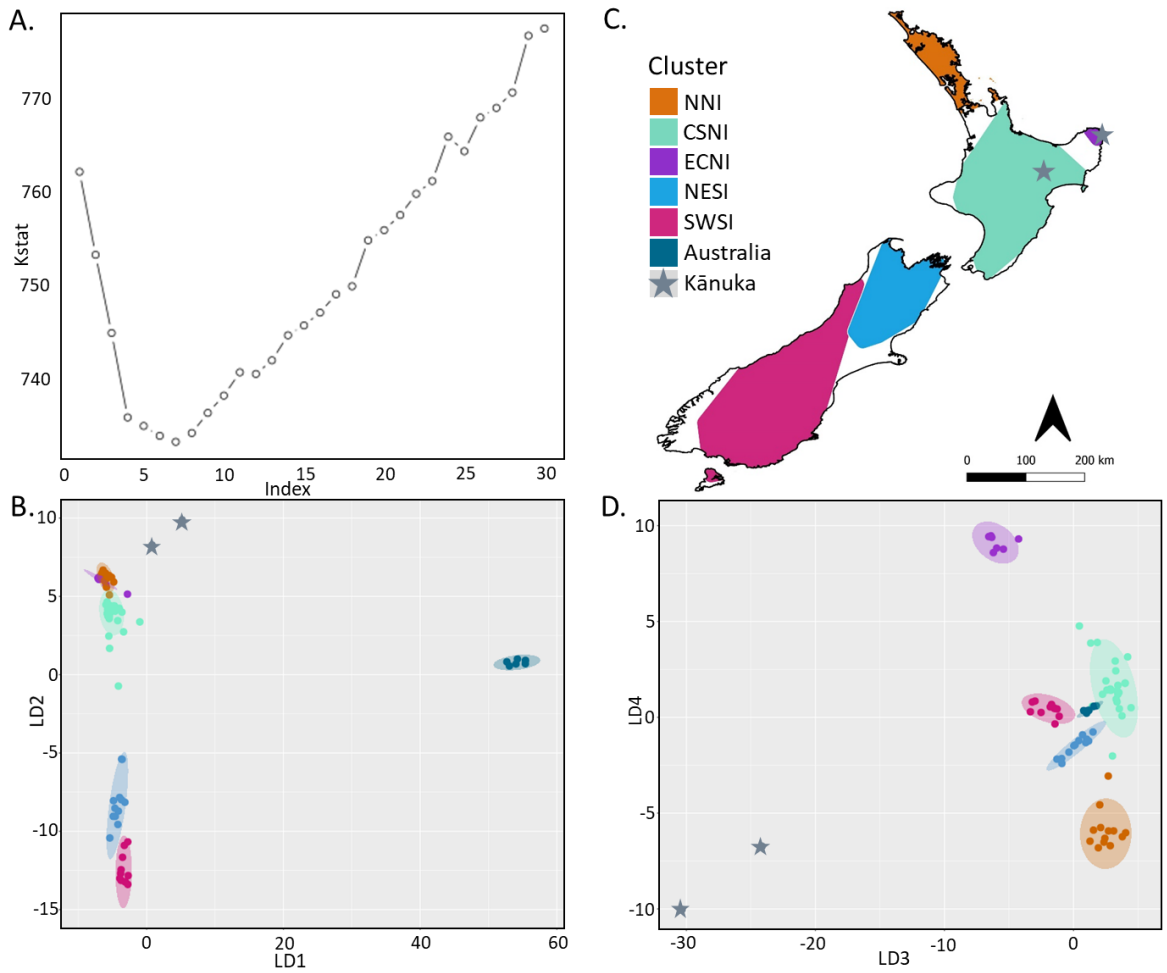
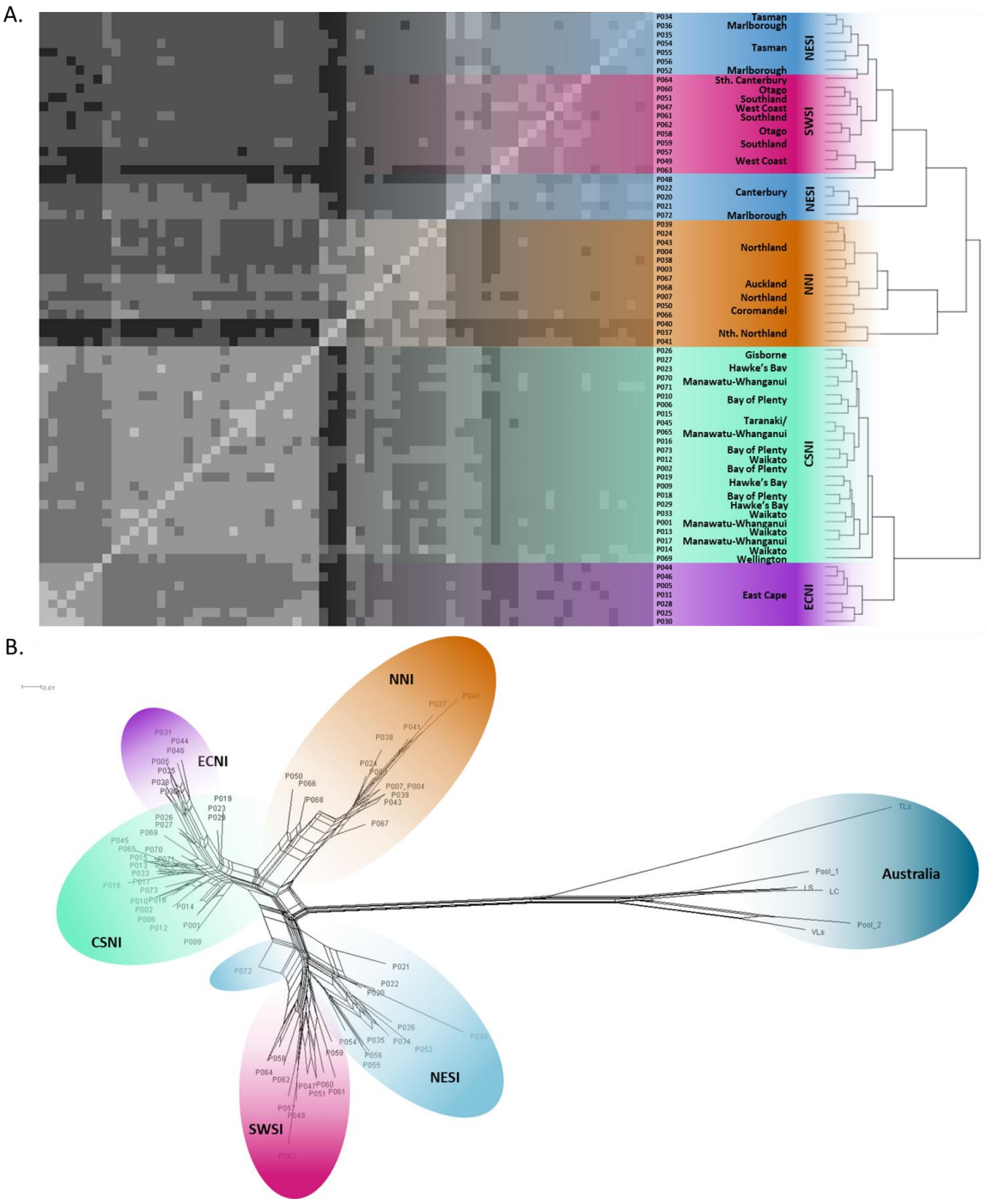


Figure 1. Population structure of *Leptospermum scoparium* in New Zealand and Australia based on whole genome pooled sequencing. A. Broken elbow plot of Bayesian Information Criterion (BIC) scores from Kmeans clustering analysis used to determine the optimal number of clusters (K) within the dataset. The optimal number of clusters is seven, including the kānuka (outgroup), Australia and five New Zealand clusters. B. & D. Discriminant Analysis of Principal components (DAPC) (K = 7) of *Leptospermum scoparium* (dots) and *Kunzea robusta* (stars; kānuka). B. displays the first two dimensions (linear discriminants) of the DAPC analysis (LD1 and LD2), accounting for 69.55% and 13.55% of the variation, respectively. D. displays LD3 and LD4, accounting for 7.54% and 4.56% of the data, respectively. C. Geographical distribution of DAPC clusters in New Zealand coloured by cluster. Colour codes and symbols are the same for B, C, and D. NNI = northern North Island, CSNI = central and southern North Island, ECNI = East Cape North Island, NESI = north-eastern South island, SWSI = south-western South Island.

Figure 2. Posterior-membership probability. Stacked bar plots of posterior-membership probability for each population for five different K values. Kan. = Kānuka, NNI = northern North Island, CSNI = central and southern North Island, ECNI = East Cape North Island, NESI = north-eastern South island, SWSI = south-western South Island. Populations are colour coded according to clustering in Figure 1, with the addition of yellow for Assigned cluster eight and pink for Assigned cluster nine

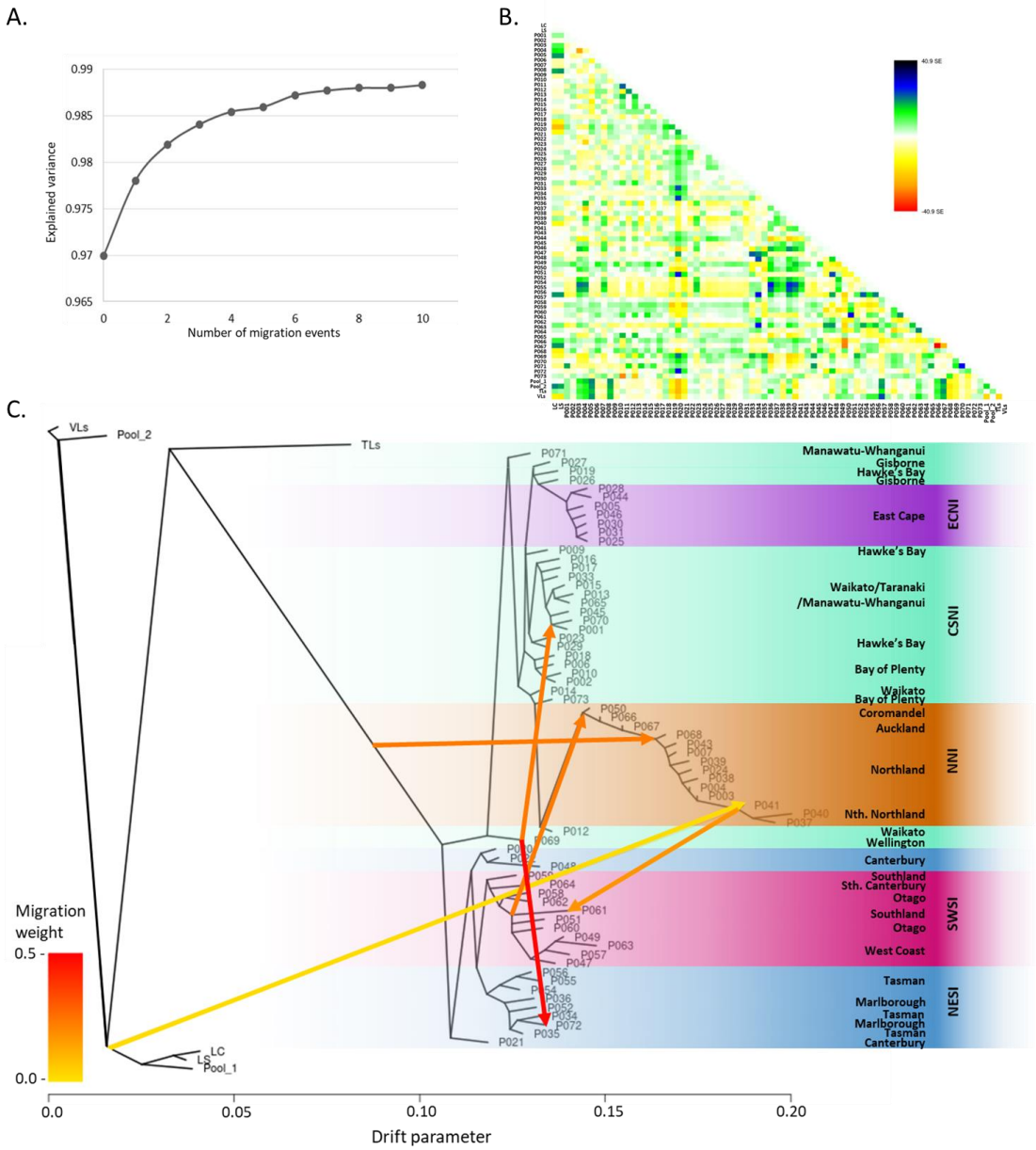
867



868

869 **Figure 3. Pairwise *F*st plot and neighbour-joining network.** A. Pairwise *F*st distance heat map of new Zealand
870 populations with dendrogram based on hierarchical clustering analysis of *F*st distances. B. Phylogenetic
871 neighbour-joining network of *Leptospermum scoparium* in New Zealand and Australia indicating population
872 differentiation using *F*st. The tree was obtained using SplitsTree. NNI = northern North Island, CSNI = central and
873 southern North Island, ECNI = East Cape North Island, NESI = north-eastern South island, SWSI = south-western
874 South Island. Populations are colour coded according to Figure 1 (with the exception of P050 which is clustering
875 with NNI rather than CSNI).

876



878

879 **Figure 4. TreeMix analysis of past migration events among New Zealand and Australia *Leptospermum***
880 ***scoparium*.** A. Estimation of the number of migration events that best explains the data. The explained variance
881 reaches a plateau around 98.6% at five to six migration events. B. Heat map of residuals from TreeMix analysis
882 with six migration events. C. Phylogenetic representation of TreeMix analysis with six migration events. Colour
883 of arrows indicates the significance of each migration event (yellow = less significant, red = more significant).
884 NNI = northern North Island, CSNI = central and southern North Island, ECNI = East Cape North Island, NESI =
885 north-eastern South island, SWSI = south-western South Island. Populations are colour coded according to Figure
886 1 (with the exception of P050 which is clustering with NNI rather than CSNI).

887

888

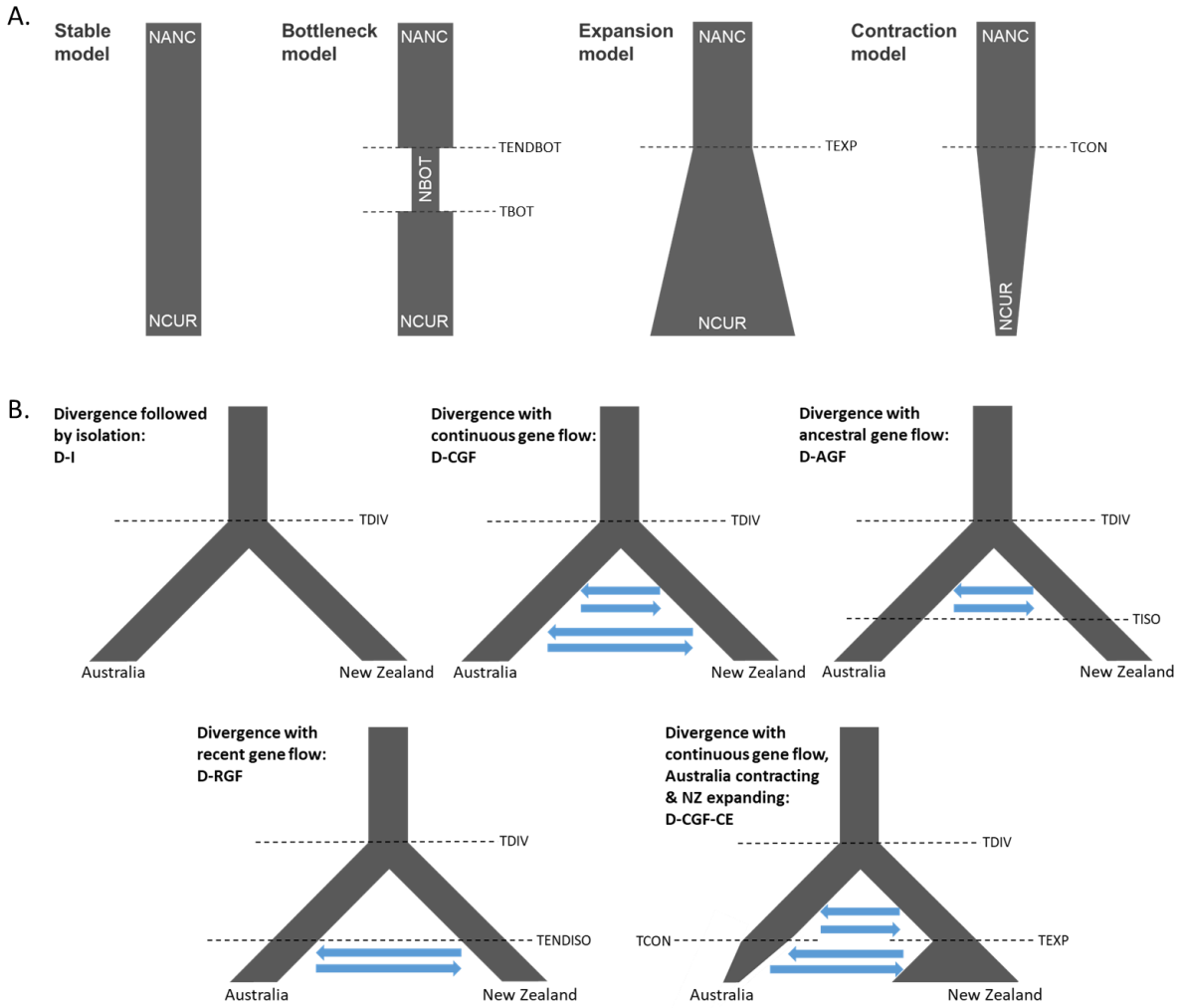


Figure 5. Diagram of the demographic models. A. models tested on individual gene pools and B. the demographic models used to investigate the divergence of Australian and New Zealand *Leptospermum scoparium* gene pools. NANC = Effective population size of ancestral population, NCUR = Effective population size of current population, NBOT = Effective population size of population during bottleneck, TBOT = time of bottleneck, TENDBOT = time of end of bottleneck, TEXP = time of expansion, TCON = time of contraction, TDIV = time of division, TISO = time of isolation, TENDISO = time of end of isolation. Dotted lines indicate when historic events occurred. Blue arrows indicate gene flow and direction of gene flow. Best model (D-CGF-CE) is highlighted inside box.

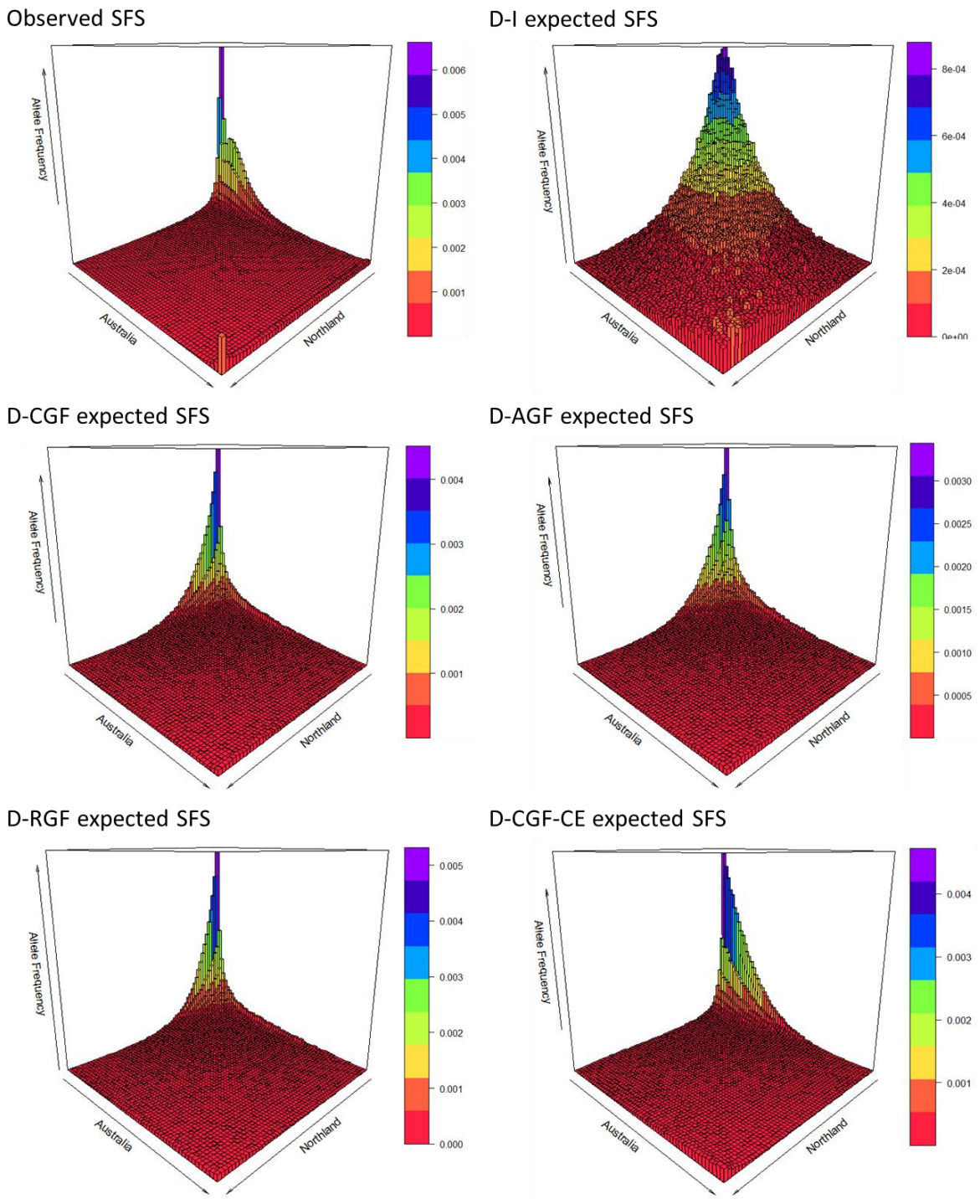


Figure 6. 3D heat maps displaying 2D SFS of Australian and New Zealand *Leptospermum scoparium* gene pools. Top left plot is the observed 2D SFS of the Australian and New Zealand (represented by NNI) gene pools used to run the demographic models, and against which the expected SFS output of all models was compared.