Biogeographic Patterns in Alpine New Zealand: Using Nuclear and Chloroplast Loci to Investigate Divergence and Secondary Contact in *Notothlaspi* (Brassicaceae)

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

In

Plant Biology

at Massey University, Manawatū

New Zealand

India Islay McJarrow-Keller

2021

Abstract

Lineage divergence and hybridisation are integral processes in biological evolution. In New Zealand alpine environments, isolation by distance and isolation by environment are two key processes that drive evolutionary change in plant lineages. The Pleistocene climate fluctuations isolated and re-introduced lineages, and complex geological substrates host several edaphic endemic floras. However, to our knowledge, plant diversification has not been studied at a population genetic scale in New Zealand alpine environments. Here, the three species in the alpine genus *Notothlaspi* (Brassicaceae) were selected to investigate divergence across heterogeneous landscapes.

Three molecular marker types were trialled in *Notothlaspi*: thirty new nuclear microsatellites, five universal chloroplast loci, and RAPDseq-generated SNPs. Twelve suitable microsatellite markers were selected, and the trnL-F and psbA-trnH loci produced high quality sequences with inter- and intra-specific polymorphisms. The RAPDseq scheme identified 41 promising polymorphic loci with several SNPs.

The twelve microsatellite markers, and the trnL-F chloroplast locus, were used to investigate the spatial distribution of genetic variation and to infer historical influences on the evolution of *Notothlaspi.* Within *N. australe*, an unexplained biogeographic split within the Kahurangi National park was discovered that deviates from expectations under isolation by distance. This suggests diverged ancestral groupings, such as in glacial refugia, and/or local adaptation to substrate. Within *N. rosulatum,* the trnL-F chloroplast locus distinguished populations in Marlborough with greater resolution than the nuclear microsatellite data, indicating a prevalence of pollen-mediated geneflow. A potential past or present hybrid zone was also identified in the Richmond Forest Park, an area of range overlap between all three species.

Here, we describe the first population genetic investigation in the alpine regions of the northern South Island of New Zealand. The findings make important contributions to future studies on biogeography, and the role of heterogeneous environments on the diversification of the New Zealand flora.

Acknowledgments

First and foremost, I would like to express my gratitude to my supervisors, Dr Vaughan Symonds, and Assoc. Prof. Alan Palmer, for their expertise, time, and guidance in completing this thesis. I very thankful for the opportunity to work with you both on such an interesting (in my opinion…) project.

Secondly, I would have been unable to complete fieldwork without the help of a great many people. Phil Garnock-Jones, Simon Walls, Rowan Hindmarsh-Walls, Simon Litchwark, and others helped me find some key sampling sites. I am also indebted to Liam Wardle, Phil Garnock-Jones, Sophie Newmarch, Olivia Wardle, John Henry, Simon Litchwark, Tessa Mitchell-Anyon, Liam Bignell, and, of course, Thomas Mackay-Smith for coming along on long, wild, and sometimes gruelling, *Notothlaspi* hunts. I am also grateful to Matthew Irwin for the loan of the GPS and other gear.

A massive thank you to Prashant Joshi, and the rest of the LoST lab members. We all shared in frustrations, but you all made time in the lab a pleasure rather than a chore. Off the lab bench, the bioinformatic help of Patrick Biggs and John Henry saved me from throwing my computer across the room, as well as months of work.

To the members of the Boffin lounge, past and present. Cheers for all the laughs, digs, and endless cups of tea. Because of you lot, I looked forward to coming into the office during the day, evening, and even during weekends.

Finally, thank you to my family for all the backing and encouragement. Mum and Dad, thank you for being great role models, and for the endless support. Krista and Klem, so I don't take myself too seriously. Also, to Nana and Grandad for being excellent hosts during fieldwork expeditions, and to Grosi, who is sorely missed.

List of Abbreviations

µL	microlitres
%P	percent polymorphic loci
A	number of alleles
AMOVA	analysis of molecular variance
BIC	Bayesian information criterion
BLAST	basic local alignment search tool
bp	base pairs
CASS	cheap as size standard
cDNA	DNA from the chloroplast genome
CTAB	cetyltrimethylammonium bromide
DAPC	discriminant analysis of principal components
ddH2O	double distilled water
DNA	deoxyribonucleic Acid
dNTP	nucleoside triphosphate
e.g.	for example
F_{IS}	Wright's fixation coefficient
F_{ST}	measure of population differentiation
GBS	genotype by sequencing
H_E	expected heterozygosity
H_O	observed heterozygosity
IBD	isolation by distance

IBE	isolation by environment
ID	identification code
Km	Kilometres
m	metres
ML	maximum likelihood
mtDNA	DNA from the mitochondrial genome
mya	million years ago
n	sample size
NCBI	national center for biotechnology information
nDNA	DNA from the nuclear genome
Ne	number of effective alleles
pA	private Alleles
PCoA	principal coordinates analysis
PCR	polymerase chain reaction
RADseq	restriction site-associated DNA sequencing
RAPDseq	random amplified polymorphic DNA sequencing
RIMs	reproductive isolating mechanisms
rpm	revolutions per minute
Rxy	correlation coefficient of Mantel test
s	seconds
SNP	single nucleotide polymorphism
SRA	sequence read archive file
SSR	simple sequence repeat

STE	sucrose, tris, EDTA
TBE	tris borate EDTA
U	units
VCF	variant call format
Ya	years ago

Table of Contents

Table of Contents Continued...

Table of Contents Continued...

List of Figures

List of Tables

Chapter 1

Introduction

How do plant lineages change over time?

1.1 Divergence

An understanding of lineage divergence is fundamental to our grasp of evolution. It occurs constantly at several scales, creating and maintaining lineages. Divergence is defined here as the accumulation of genetic variation between lineages on different evolutionary trajectories. Interestingly, divergence arises during both geographic separation (allopatry), and while maintaining some physical contact (sympatry, heteropatry, or parapatry) (EHRLICH AND RAVEN 1969). Divergence occurs through a variety of mechanisms in a variety of circumstances, but a key component is a lack of gene flow between groups. Gene flow occurs through the interbreeding of individuals, so the pattern of seed and pollen dispersal is an important factor for sessile plants. At the microevolutionary scale, changes in allele frequencies alter populations over time (HARDY 1908; WEINBERG 1908; VIA 2009). When there is no interbreeding, changes in allele frequencies accumulate, and are not shared between the isolated lineages. Eventually, these changes can form macroevolutionary signatures that define the formation of new groups, such as ecotypes or species.

1.1.1 Stochastic Mechanisms of Divergence

Historically, random genetic drift was thought, by some, to be the main mechanism of divergence between separated populations (KIMURA 1983), where allele frequencies alter over generations due to chance, generally moving towards uniformity in the population (KIMURA AND OHTA 1969). Under particular circumstances, certain alleles are randomly inherited at a higher frequency than others, so allele frequencies flux between generations. This can increase the chance of allele loss or fixation. Small populations are more vulnerable to allele loss or fixation from genetic drift, but the process occurs at neutral (not experiencing selection) loci even in larger populations, increasing divergence between isolated gene pools where changes in allele frequency are not shared.

Over time, the importance of linked selection, or 'genetic draft' has become recognised, even as having a stronger effect on patterns of genetic variation than genetic drift (GILLESPIE 2000; NEHER 2013). Linkage disequilibrium is the "non-random association of alleles at different loci" (FLINT-GARCIA *et al.* 2003), which can have numerous sources. One example is 'linkage', the physical proximity of two loci on a chromosome, which reduces the chances that two alleles will be separated by recombination (FLINT-GARCIA *et al.* 2003). Linkage disequilibrium also occurs at physically distant loci, such as between loci in the chloroplast and nuclear genomes that are important for cytonuclear interactions (FIELDS *et al.* 2014). Thus, no locus evolves in complete isolation, instead, complicated patterns of linkage can skew allele frequencies even at 'neutral' loci: for example, a selective sweep bringing an adaptive (useful in the present environment) allele into higher frequency in a population may also bring linked alleles to a higher frequency.

Genetic draft is less affected by effective population size than genetic drift (GILLESPIE 2000). Therefore, if stabilising selection (selection favouring average phenotypes) is common, variation at linked sites may be lost, but if balancing selection (maintenance of multiple phenotypes) is occurring, variation may increase (PREZEWORSKI *et al.* 2005; LEFFLER *et al.* 2012). During times of rapid evolutionary change, the spread of beneficial mutations and the elimination of recurrent deleterious mutations will reduce the effective population size, particularly in genomic regions with low cross-over rates. This, in turn, increases genetic drift in these locations, increasing divergence between populations at these genomic sites. The findings from CHEN *et al.* (2020) and others led them to propose that in small populations, genetic drift was driving patterns of diversity, while in large populations genetic draft better explained them. These two mechanisms are considered among the most important factors in neutral allele frequency change in populations during reproductive isolation.

1.1.2 Selective Mechanisms of Divergence

Differences in selection pressure on populations can likewise drive divergence. Natural selection acting differentially on populations will drive allele frequency differences, favouring useful allelic combinations in their environmental conditions (AGUIRRE-LIGUORI *et al.* 2019). Local adaptation occurs when a population responds to selective pressures in their specific environment, so sorting of standing genetic variation (pre-existing variation within the gene pool) occurs to adapt to local circumstances. VIA (2009) suggests that ecological divergence (divergence due to differential adaptation to ecological factors) occurs in two stages: initial selection against effective migration between lineages, followed by separate selection pressures and responses to selection. Loci under divergent selection can have a wide sphere of linkage, and therefore affect

many 'neutral' loci, as mentioned above (NOSIL *et al.* 2009). It is these loci that define phylogenetic patterns that may eventually be reflected across other loci.

When local adaptation occurs, gene flow between populations in different environments is often unfavourable because hybrid offspring with diluted adaptive genes are selected against (LENORMAND 2002). Evidence for local adaptation was first experimentally demonstrated by Göte Turesson in the 1920s (TURESSON 1922) in his elevational common garden studies. Following this, numerous common garden (growing different plants in equivalent environmental conditions) and reciprocal transplant studies have reinforced the importance of local adaptation, particularly in elevational and edaphic experiments (CLAUSEN *et al.* 1940; SAMBATTI AND RICE 2006; BYARS *et al.* 2007), which have also shown local adaptation to be stronger in large populations (LEIMU AND FISCHER 2008). Local adaptation in sympatry can lead to the development of reproductive isolating mechanisms (RIMS), which are thought to be based on few, pleiotropic (one gene influencing multiple seemingly unrelated traits) loci (YEAMAN AND OTTO 2011; FERRIS *et al.* 2017). For example, allele frequency differences between *Leptosiphon parviflorus* Benth. (Polemoniaceae) populations on sandstone and serpentine soil types has been shown to be maintained in the face of gene flow through environmentally and genetically mediated flowering time differences (DITTMAR AND SCHEMSKE 2017).

1.1.3 Polyploid Divergence

One mode of sympatric divergence is polyploidy, sometimes termed 'instant speciation'. Polyploidy is the heritable duplication of the genome (WOOD *et al.* 2009). This type of rapid divergence can occur through several mechanisms, such as the formation of unreduced gametes, polyspermy, or somatic doubling in meristematic regions or in a zygote (RAMSEY AND SCHEMSKE 1998; TAMAYO-ORDÓÑEZ *et al.* 2016).

Polyploidy is a surprisingly common phenomenon in particular lineages, such as angiosperms and ferns (WOOD *et al.* 2009). Historic polyploidy events are thought to underly several major radiations, such as the emergence of seed plants and angiosperms (SOLTIS *et al.* 2009). Coincidentally, many of these events have occurred around major environmental changes, during stressful conditions where e niche availability probably expanded dramatically (MASON AND PIRES 2015; KARUNARATHNE *et al.* 2018). This is perhaps because polyploidy can increase the evolutionary potential of groups: the duplicated genomic complement allows for neo-, sub-, and non-functionalisation of genes (CHENG *et al.* 2018). In neo-functionalisation, one gene copy retains its original function, while the other can take on a new function. Alternatively, each gene copy can retain different aspects of the original function (sub-functionalisation), or one gene loses function (non-functionalisation). Neo- and sub-functionalization can provide polyploids with opportunities for rapid adaptation and divergence from their progenitors, along with the generation of new genetic variation. In new environments, polyploidy may give lineages the upper hand. A recent climatic extreme was the Pleistocene glacial cycles, with their dramatic shifts in temperature, and the disruption of species ranges due to ice. It has been suggested that polyploids may have been better able to exploit the highly heterogeneous and changing environment (KAGALE *et al.* 2014; VAN DE PEER *et al.* 2017). Which is precisely what may have occurred in New Zealand (JOLY *et al.* 2009a; BEHLING *et al.* 2020; MEUDT *et al.* 2021).

1.1.4 Opportunities for Divergence in New Zealand

The New Zealand alpine regions are a hotbed of species radiations and recent divergences, where high morphological diversity is found between species, but relatively low genetic divergence (WAGSTAFF *et al.* 1999; LOCKHART *et al.* 2001; WAGSTAFF AND BREITWIESER 2004; JOLY *et al.* 2009a; HEENAN AND MCGLONE 2013; MEUDT *et al.* 2015).

New Zealand's geological history has provided many opportunities for allopatric divergence within lineages, such as glaciations, mountain building, oceanic inundation, and volcanic eruptions (WINKWORTH *et al.* 2002). Key recent events thought to be of significance to the patterns of diversification on the South Island occurred during the Pleistocene, where intense cycles of climate change and mountain building occurred (HEENAN AND MCGLONE 2013). Meanwhile, on the North Island, large volcanic eruptions also have had a major impact on species distributions (HORROCKS AND OGDEN 1998; SHEPHERD *et al.* 2007; ELLIS *et al.* 2015). During cold periods, plants and animals tend to migrate to lower altitudes and latitudes, following conditions they are adapted to (TELWALA *et al.* 2013). By the same token, during warm periods they move up in latitude and altitude. Therefore, the cycling of warm and cold periods in New Zealand and around the world isolated populations from others of the same species, and then brought them again into secondary contact (BECK *et al.* 2008; MARRA AND THACKRAY 2010). At the peak of the glaciations, much of the southern alps were covered in ice, while the lowlands were dominated by cold tolerant grasslands (NEWNHAM *et al.* 2013). Pollen records indicate that many forest species, in particular beech, survived in local refugia that would have kept populations isolated from one another (NEWNHAM *et al.* 2013). This is an excellent environment for diversification of forms (WICHMAN *et al.* 2002), as small effective populations are far more susceptible to fixation of genes with random genetic drift. Populations also have greater opportunity for local adaptation in isolation from gene flow with populations in other environments. In the absence of gene flow, genetic changes, along with new mutations, are not shared between the isolated populations: leading to divergence over time. For alpine species living along the long chain of the Southern Alps, cold periods may have isolated populations to the North and the South, while during warm periods contact may have been re-introduced as they re-enter the

alpine zones (WALLIS *et al.* 2016). These isolation and re-introduction patterns allow ample opportunity for hybridisation and introgression (see below), particularly between recently diverged groups, as much of the alpine flora and New Zealand is suggested to be (MCGLONE *et al.* 2001).

1.2 Barriers to Divergence

Gene flow has the general effect of slowing or preventing divergence; it has been suggested that even a single migrant per generation can counteract divergence (FRANKLIN 1980; MILLS AND ALLENDORF 1996). Species, a fundamental unit of human understanding of evolution, were historically viewed as reproductively isolated entities (MAYR 1942). However, hybridisation between species is rampant, particularly in the plant kingdom, so many now view species boundaries as semi-permeable membranes that can be pierced by gene flow (HARRISON AND LARSON 2014). Hybridisation and introgression both involve the re-introduction of gene-flow between diverged lineages. Generally, this occurs within a genus, but there is some evidence for cross-generic hybridisation, such as in New Zealand Gnaphalieae (Compositae) (MCKENZIE *et al.* 2004; SMISSEN AND BREITWIESER 2008). Hybridisation is defined by HARRISON AND LARSON (2014) as 'the interbreeding of individuals from two distinct populations or groups of populations', while introgression is defined as 'the incorporation (usually via hybridization and backcrossing) of alleles from one entity (species) into the gene pool of a second, divergent entity (species)'.

1.2.1 Evolutionary Outcomes of Hybridisation

Hybridisation has various evolutionary outcomes: including the merging of diverged lineages, homoploid hybrid speciation (DONOVAN *et al.* 2010), polyploid speciation

(PERRIE *et al.* 2010), gene transfer between groups (SMISSEN *et al.* 2004), or of reinforcing genetic isolation (MELO *et al.* 2014). In order to survive as a successful homoploid hybrid species, they must outcompete the parental species in their niche, or move into a new one (GROSS AND RIESEBERG 2005). More commonly, hybridisation results in polyploid speciation, introgression, or the reinforcement of reproductive isolation (SOLTIS AND SOLTIS 2009; SCHUMER *et al.* 2014; GARNER *et al.* 2018).

1.2.2 Introgressive Allele Transfer

Introgression allows for adaptive (and non-adaptive) allele transfer between diverged populations, although not all parts of the genome are equally prone to transmission. Some combinations of genes are not favourable to hybrid survival, so they can be preferentially excluded from introgression (TURNER *et al.* 2005; HARRISON AND LARSON 2016). These incompatible alleles and their sphere of linkage remain pure and are termed 'islands' of speciation (TURNER *et al.* 2005). It is areas outside of this sphere of linkage that may be introgressed, and thus introgression is measured in relation to these 'pure' zones. For those alleles that may be introgressed, they are an important source of variation transfer, allowing the passing of pre-adapted advantageous alleles through the 'semi-permeable membranes' separating species. When these alleles are maintained in a new genetic background by natural selection, it is known as adaptive introgression (WHITNEY *et al.* 2006; KIM *et al.* 2008; SUAREZ-GONZALEZ *et al.* 2018).

1.2.3 Detection of Hybridisation

Hybridisation is often identified through intermediate morphology, however introgression can be more difficult to detect. Commonly, plastid genomes (such as chloroplast or mitochondrial genomes) are exchanged through introgression, so in many cases phylogenetic trees based on organellar markers show geographic patterning

(THÓRSSON *et al.* 2001; ACOSTA AND PREMOLI 2010; SCHUSTER *et al.* 2018; LIU *et al.* 2020). Exchange of plastid DNA leads to discordance between gene trees, which is important evidence for introgression, however it can also be attributed to incomplete lineage sorting (ILS) (JOLY *et al.* 2009b). ILS occurs when alleles are not fully segregated into alternative lineages during divergence, therefore, different genes can indicate different phylogenetic relationships: just as with introgression. In the case of introgression, markers often show patterns reflecting the geographic overlap of species during the period of introgression, whereas ILS allele distribution is typically more widespread (SEGATTO *et al.* 2014; GAO *et al.* 2015). Otherwise, using coalescent based frameworks (i.e. approximate Bayesian computation, and isolation with migration model-based programs), and past and present distribution models of species ranges can give some indication of allelic origin (ZHOU *et al.* 2017). JOLY *et al.* (2009b) used a New Zealand alpine genus, *Ranunculus* L. (Ranunculaceae) to trial a new method of differentiating between hybridisation and ILS. The idea was that from certain hybridisation events, the minimum genetic distance at a non-recombining locus will be smaller than for an ILS scenario. Specifically, if the differences at the loci of interest were formed following the historic speciation event between the species, then the minimum differences will be less than for ILS. Using this method, they pointed to hybridisation as a probable explanation for non-monophyly in their study group.

Hybridisation has long been recognised as a feature of the New Zealand flora (COCKAYNE AND ALLAN 1934; RATTENBURY 1962), however, many examples have yet to be ratified by genetic analyses (MCKENZIE 2001). An approach using markers from the different plastid genomes, along with nuclear DNA is important for determining hybridisation, for the reasons mentioned above. The next section will discuss a population

genetic approach to studying patterns of relationship between populations and species, along with a number of different molecular markers from plastid and nuclear DNA.

1.3 Population genetics

Population genetics is often utilised to assess genetic patterns and is capable of detecting recent divergence and introgression as it investigates fine-scale changes in allele frequencies using mathematics and genetic principles (KIMURA AND OHTA 1971). VIA (2009), highlights the importance of studying divergence at the population level rather than just at the species level, so as to observe divergence in action rather than in retrospect. One example they gave was how regions of divergence hitchhiking involved in initial differentiation cannot be detected in retrospect because there are no tangible changes to the chromosomes. In this context, populations are defined as a group of interbreeding individuals, the fundamental unit of evolution. It is at this scale, that evolutionary change is initiated. Differences in allele complement and frequency between populations indicate a lack of gene flow and divergence, showing the distribution of genetic variation within/between species, and across landscapes, along with within-population dynamics (KIMURA AND OHTA 1971; MANDEL *et al.* 2011).

1.3.1 Genetic Marker System

Population geneticists use molecular markers as a small window into the genome (SCHLÖTTERER 2004; NIU *et al.* 2019), however, relative to the whole genome, these windows are very small. So, finding a suitable marker system to represent the wider genome is an ongoing challenge. Plant cells host three genomes, mitochondrial (mtDNA), chloroplastic (cDNA), and nuclear (nDNA), each with different characteristics and patterns of inheritance that are useful for different analyses. The plastid genomes

(mtDNA and cDNA) are typically (with notable exceptions: HANSEN *et al.* (2007); NAKAMURA (2010)) uniparentally inherited, while nDNA is biparentally inherited. This can allow for inferences on the relative contributions of seed and pollen dispersal in a population, and plastid DNA often shows patterns of geographic structure better than nDNA (BAI *et al.* 2014; ROGALSKI *et al.* 2015). They also lack variation among copies, and are effectively haploid, so also can be useful markers for examining polyploids, which tend to be more difficult in population genetics analyses (see below). When a large population is fragmented into multiple small, isolated populations, such as during a glacial period, both the plastid and nDNA will diverge. Upon secondary contact, the nDNA can recombine, and mix, while the plastid DNA may remain divergent, and more representative of geographic location than lineage because pollen flies further than seed, and the plastids are often (but not always, e.g. NEALE *et al.* (1989); HANSEN *et al.* (2007)) maternally inherited, so are passed on through seed (DESPRÉS 2019).

Different loci also change at different rates. Mutations in more conserved regions and their zone of linkage generally are retained, therefore these loci evolve more slowly. This makes them useful for studying deeper phylogenetic scales, as there is less risk of reverse mutations that obscure patterns of relationships through homoplasy (similar characteristics not attributed to shared ancestry) (GOREMYKIN *et al.* 2015). At the population level however, these loci are not likely to show any change, so markers with higher rates of change are typically necessary. Rates of change are highly dependent on selection pressures and loci under stabilising selection will alter more slowly compared to genes involved in adaptation to a new niche. Therefore, many population genetic studies use putatively neutral markers, as it is assumed these will be more representative of whole genome background mutation rates (KIRK AND FREELAND 2011; DEWOODY *et al.* 2021). Although, genetic draft can affect this 'neutrality'.

The first molecular marker system commonly employed to investigate genetic diversity were allozymes (KOENIG AND GEPTS 1989; SINGH *et al.* 1991; HAMRICK AND GODT 1997). These are protein variants with size variation are visualised through gel electrophoresis. Differences between lineages could be visualised, however, they only show expressed phenotypic variation, which may not represent the genome as a whole (SCHLÖTTERER 2004). With the advent of sequencing and genotyping technology, new techniques were developed to view and quantify DNA sequences themselves. The polymorphisms targed include SNPs (Single Nucleotide Polymorphisms) and microsatellites (Simple Sequence Repeats, SSRs) (VAN INGHELANDT *et al.* 2010; GROVER AND SHARMA 2016).

SNPs are the substitution of a single nucleotide for another, a mutation that does not cause a frameshift. Numerous marker systems, such as RAPDseq (randomly amplified polymorphic DNA sequence) (MCLAY 2011; PILKINGTON 2019), RADseq (BAIRD *et al.* 2008) (restriction site associated DNA sequencing), and GBS (Genotyping-by-sequencing) (KISHOR *et al.* 2020), have been developed to identify and genotype SNPs. These systems each use different methods to reduce the complexity of the genome before subsequent sequencing and filtering of the reads to identify SNPs. RAPDseq uses random primers that amplify multiple loci in the genome to reduce complexity in an initial PCR step. PCR products are individually tagged in a subsequent PCR step and pooled for amplicon sequencing. In comparison, GBS and RADseq use restriction enzymes to isolate sites before subsequent PCR amplification and sequencing (GURGUL *et al.* 2019). While both RADseq and GBS are widely used, RAPDseq requires fewer steps, less input DNA, and is broadly applicable to a number of plant species, particularly those containing secondary compounds that inhibit enzymatic activity (MCLAY 2018).

Microsatellite loci are consecutive repeats of one to six base pairs found throughout most eukaryote genomes (TÓTH *et al.* 2000). They have a very high mutation rate through the process of slip strand mispairing during replication, resulting in length polymorphisms with a high resolution, mendelian inheritance, and co-dominance (LEVINSON AND GUTMAN 1987). They are therefore widely used in population genetic studies. However, they suffer from size homoplasy, and cannot be used across higher taxonomic scales (ESTOUP *et al.* 2002). The flanking regions of SSRs are also quite variable, as can be attested to by null alleles resulting from mutations in the priming regions (ISHIBASHI *et al.* 1996; ZARDOYA *et al.* 1996; BATLEY *et al.* 2003; ABBOTT *et al.* 2010).

1.3.2 Statistical and Analytical Methods

For practical purposes, many within-population analyses use a hypothetical population at Hardy-Weinberg equilibrium as the null model (WRIGHT 1949). This population is in stasis, with assortative mating and no mutation, migration, natural selection, and genetic drift. Biological inferences can be made from any deviation from the allele frequencies predicted here. F-statistics are commonly used metrics proposed by WRIGHT (1949). F_{IS} is a within-population metric that examines the frequency of heterozygotes observed relative to the expectation under a model of random mating. A homozygote excess can be indicative of a number of things, including inbreeding, null alleles, positive assortative mating, selection for homozygotes, or population substructure (ROBERTSON AND HILL 1984; STOECKEL *et al.* 2006; FERRAZZINI *et al.* 2007). A heterozygote excess could indicate: low selfing (self-pollination) in small populations; selection for heterozygosity; negative assortative mating, when individuals mate with dissimilar individuals; or asexual reproduction (BALLOUX 2004; STOECKEL *et al.* 2006; ARROYO 2013). F_{ST} is Wright's between-population metric, used to determine the degree

of genetic divergence between populations by comparing the heterozygosity within a sub-population to the data as a whole. However, since the time of Wright, several statistical methods were developed to analyse population genetic data. F_{ST} is but one genetic distance metric of many (NEI 1972; PREVOSTI *et al.* 1975; REYNOLDS *et al.* 1983; GOLDSTEIN *et al.* 1995). Each is suitable for different data types, and many have underlying biological assumptions.

To visualise genetic structure and patterns of relationship, numerous population genetic methods have been developed and derived from other disciplines. The list of programs and methods is extensive, so only four are briefly included here. STRUCTURE (PRITCHARD *et al.* 2000) is one of the most commonly used methods, a Bayesian clustering approach is used to assign individuals to a pre-defined number of populations (*K* value) based on genotype data. Several applications have emerged that increase the utility of the program, by determining which replicates and *K* values are best supported by the data, and giving greater power to alter the visual parameters of the outputted bar plot (ROSENBERG 2004; JAKOBSSON AND ROSENBERG 2007; KOPELMAN *et al.* 2015). Another popular visualisation of genetic structure is a Neighbour-Net (BRYANT AND MOULTON 2004). Unlike a single bifurcating tree, these figures display conflicting signals in the data using a net of relationships which may be more realistic for complex biological systems such as in those with hybridisation, which are complicated to display with a branching tree (SMISSEN AND HEENAN 2007; PAN *et al.* 2019). Some analyses are utilised from other disciplines, for example a Principal Coordinates Analysis (PCoA), or multidimensional scaling, plot considers dissimilarity between data points and finds the dimension through the matrix of points where variation is maximal (REICH *et al.* 2008; GAUCH *et al.* 2019). This method can show groupings between data points, and patterns of relationship between different individuals and populations. A newer method,

Discriminant analysis of Principal Components (DAPC) (JOMBART *et al.* 2010) also uses a principal component method, in addition to discriminant analysis. This method finds clusters in the data, before visualising between-cluster variation (FATOKUN *et al.* 2018; PEREIRA-DIAS *et al.* 2019). Within-cluster genetic variation is minimised, so between-group variation can be better visualised. DAPC, PCoA, Neighbour-Net, and STRUCTURE are four of the more common analyses, however as mentioned above, many more are available. Informed decision making is required when deciding which analysis to use, particularly when analysing complicated systems like polyploids.

1.3.3 Polyploid Population Genetics

Many traditional population genetic models are based on functionally diploid species portraying typical mendelian inheritance patterns (SELKOE AND TOONEN 2006; MEIRMANS *et al.* 2018). However, polyploids are common in many plant lineages (e.g. angiosperms and ferns (WOOD *et al.* 2009)), and even 'diploid' model species, i.e. *Arabidopsis* (Brassicaceae), have been found to be paleopolyploids (BOWERS *et al.* 2003; JIAO *et al.* 2011). Paleopolyploids are derived from an ancestral polyploid event, usually with several modifications to the duplicated gene copies. Allopolyploids are formed when polyploidy coincides with hybridisation of diverged genomes. Within an allopolyploid, corresponding genes or chromosomes from the same lineage are termed homologues, while those from different lineages are termed homoeologue. Homologues tend to pair together, so disomic inheritance is observed in allopolyploids formed from highly diverged progenitors (COMAI 2000; LLOYD AND BOMBLIES 2016). In autopolyploids, all copies of a given chromosome can align during meiosis. This occurs when pairs of chromosomes are similar enough for polysomic pairings. In segmental allopolyploids, both disomic and polysomic inheritance is observed on different chromosome pairings (BHARATHI *et al.* 2010; MASON AND WENDEL 2020). In many cases, more recent

polyploids show multivalent pairings, but older polyploids settle into a more stable, disomic inheritance pattern similar to diploids (LE COMBER *et al.* 2010).

The issues with using polyploids in population genetic studies have been extensively documented (DUFRESNE *et al.* 2014; BOURKE *et al.* 2018), and are exacerbated in non-model organisms where often ploidy levels and segregation patterns are uncertain. As mentioned above, many metrics in population genetics rely on estimations of heterozygosity to determine allele-frequency-based inferences and levels of inbreeding. Polyploids can be partial heterozygotes, full heterozygotes, or full homozygotes. For example, a tetraploid has five potential allele combinations rather than the diploid three. Resolving partial heterozygotes by determining dosage information remains a major challenge in polyploid population genetics (DUFRESNE *et al.* 2014), and the increased copy number in polyploids can lead to bias estimation of heterozygosity using diploid models. This increased copy number can also obscure if marker information represents a single, or multiple duplicated loci. The inheritance pattern of study species is particularly important for population genetic studies, for whom many models assume specific inheritance patterns. Most statistical methods have been designed for diploids, so uncertainty is introduced in polyploid datasets. Allopolyploids often display diploid-like inheritance patterns; however, the challenge lies with detecting which genes or markers arise from which homoeologous pairings. Segmental allopolyploids and autopolyploids can have complicated pairing patterns, creating difficulty for genetic analysis when these are not known (BOURKE *et al.* 2018). However, while the nuclear genome doubles, plastid markers remain unduplicated, and are thus useful marker options for polyploid population genetics.

1.4 *Notothlaspi*

Worldwide, population genetics studies investigate intraspecific divergence (SCHIERENBECK 2017; COPPI *et al.* 2020; YAMAMOTO *et al.* 2020), however there is a marked lack of these studies in New Zealand, particularly in alpine regions. Alpine regions in New Zealand make excellent environments for studying recent divergence, as mentioned above, due to the recent origin of much of the flora and highly heterogeneous landscapes and geological foundations. Knowledge of geographic patterns of genetic variation and divergence intraspecifically also would greatly strengthen New Zealand's world renown conservation programs, which could use this information to redirect resources to hotspots of diversity. To investigate microevolutionary spatial patterns in the New Zealand alpine, the *Notothlaspi* (Brassicaceae) genus was selected. The distribution of the genus crosses highly heterogeneous geological soil types and is found in recently uplifted and glaciated mountain ranges, so it may be an excellent group for investigating divergence in the New Zealand alpine environment.

The *Notothlaspi* genus was first described by Sir Joseph Hooker in 1864, and later re-examined by HEENAN (2019). It is the only genus in the tribe Notothlaspideae (MITCHELL AND HEENAN 2000), and is comprised of three species, *N. rosulatum, N. australe,* and *N. viretum*. The genus is endemic to the South Island of New Zealand, where it occupies harsh, but less competitive environments. All three species are alpine herbs that prefer rocky habitats, have fleshy leaves, and small white flowers with four petals (see Table 1.1 for characteristics). Morphologically, *N. australe* is highly variable (see Figure 1.1), although much of this variability is likely to be linked to environmental conditions such as substrate type and exposure to solar radiation. From personal observation, individuals tucked in shady rock crevasses tend to have more rosettes and greener pigmentation than individuals from the same populations that grew in more

exposed locations. *Notothlaspi rosulatum* are variable in size, but less diverse in habit (Figure 1.1). The reproductive biology of the three species is unknown, however the flowers of *N. australe* and *N. rosulatum* both produce strong scents. *Notothlaspi rosulatum* produces a sweet smell, while *N. australe* flowers are more cloying. The production of scent implies they are (or have in the past) trying to attract a pollinator to promote outcrossing. The stamens are within reach of the stigma, however, this does not discount other potential mechanisms of self-incompatibility. While no pollination surveys have been carried out, Coleoptera (beetles) and Lepidoptera (butterflies or moths) visitation has been observed on *N. rosulatum* in the field (Figure 1.2), and Hymenoptera (ants) on *N. australe.*

The three species within the genus are alpine specialists, it is likely they were only able to move into the alpine zone from 0.95Ma (HEENAN AND MCGLONE 2013). Prior to this, the Southern Alps were not high enough to form a persistent alpine habitat above the tree line during climate fluctuations. A number of alpine specialists found in similar fellfield and scree slope locations to *Notothlaspi* have been split into species in different areas within the *Notothlaspi* range, such as the New Zealand *Gentianellas* (Gentianaceae) (GLENNY 2004). This is particularly true across different substrates. *Notothlaspi australe* populations grow on soils with vastly different physical, chemical, and biological properties, formed from parent materials such as calcareous, quartzo-feldspathic, granitic, volcanic, and ultramafic rock.

Local adaptation in heterogeneous landscapes has been shown worldwide to be a driver of diversification (WALTER *et al.* 2016; MUSKER *et al.* 2021), so *Notothlaspi* may be an excellent group to study divergence across environmental niches intraspecifically, as well as investigating interspecific dynamics in potential hybrid zones.

Table 1.1: Characteristics of *Notothlaspi spp.*. Taken from personal observation and HEENAN (2019).

Characteristic	*N. rosulatum*	*N. australe*	*N. viretum*
Habitat	Scree slopes	Fellfields and rock ledges	
Inflorescence	Terminal, upright, racemose	Axillary	Axillary
Number of rosettes	Typically, one	One to numerous	One to numerous
Life cycle	Annual/biennial	Perennial	Perennial
Leaf shape	Rosulate, >17mm wide, numerous dentate-serrate teeth, caducous hairs (upper surface), lower surface glabrous, purple to brown to green	Ovate to broadly ovate, 1-6 teeth, glabrous or sparsely hairy, green to brown	Linear to linear-narrowly lanceolate, 0.7-2.4mm wide, entire or occasionally 1-pinnatafid, glabrous, green
Substrate	Greywacke	Wide ranging substrates, including ultramafic, siliciclastic, calcareous, volcanic, and granitic	Ultramafic harzburgite

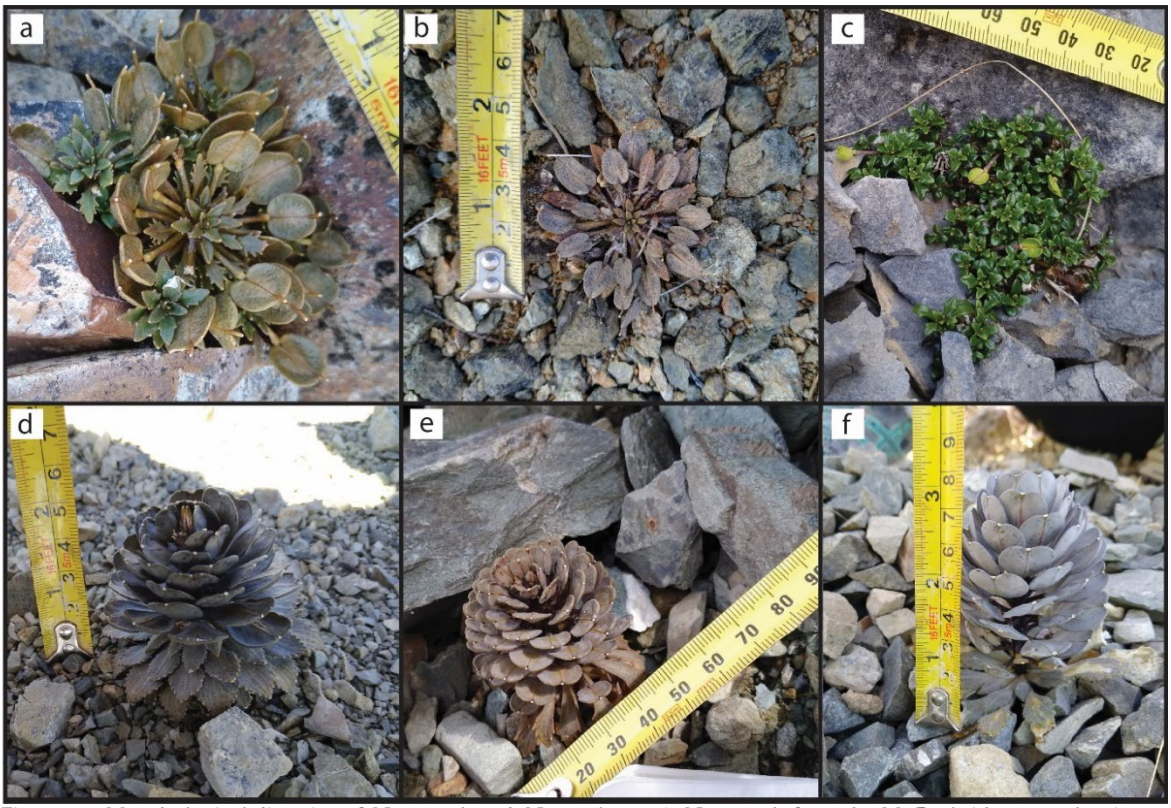


Figure 1.1: Morphological diversity of *N. australe* and *N. rosulatum* a) *N. australe* from the Mt Peel ridge on volcanic conglomerate (Kahurangi National park, NZ), photographed during January. b) *N. australe* on ultramafic fellfield near Dun Mountain (Richmond Ra., NZ) photographed in early February. c) *N. australe* on marble at the Mt Owen Massif, (Kahurangi National Park, NZ) photographed in late January. d*) N. rosulatum* on the Amuri ski field, NZ, late January. e) *N. rosulatum* on Mt Richmond (Richmond Ra., NZ), late January. f) *N. rosulatum* on the banks of Lake Tennyson (Molesworth Recreational Reserve, NZ), late January.

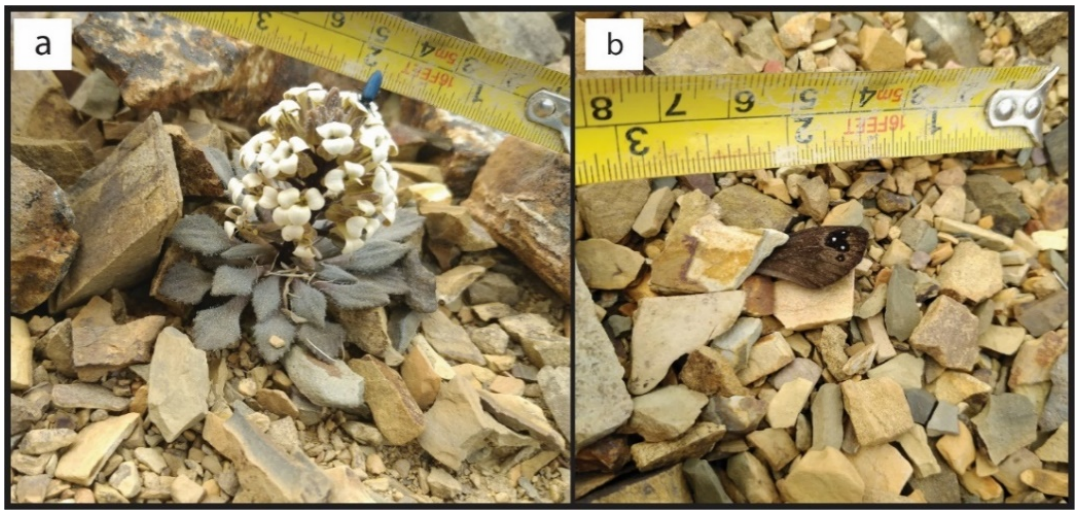


Figure 1.2: Insect visitation on *N. rosulatum* flowers. a) Coleopterra of unconfirmed identity visiting *N. rosulatum*, *Oedemeridae* or *Cerambycidae.* b) The wing of an unidentified *Lepidoptera* (perhaps *Percnodaimon merula*). They were observed and recorded visiting multiple flowers on the same individual, and hopping between individuals on the Mt Robert Ridge, Nelson Lakes, NZ.

Figure 1.3 shows the distribution of the *Notothlaspi* species. *Notothlaspi australe* and *N. viretum* are sympatric, occupying similar habitats in the harsh ultramafic environment of the Red Hills. This is the only place *N. viretum* has been found (HEENAN 2019). *Notothlaspi australe* and *N. rosulatum* are broadly separated by the alpine fault, however they are both found in the Richmond ranges and in the Nelson Lakes area, where there may be potential for gene flow. They are also separated by chromosome counts; *N. rosulatum* has 2n = 36-38 and N. australe has a 2n = 90-100 (DAWSON AND BEUZENBERG 2000), however these counts are each from a single location.

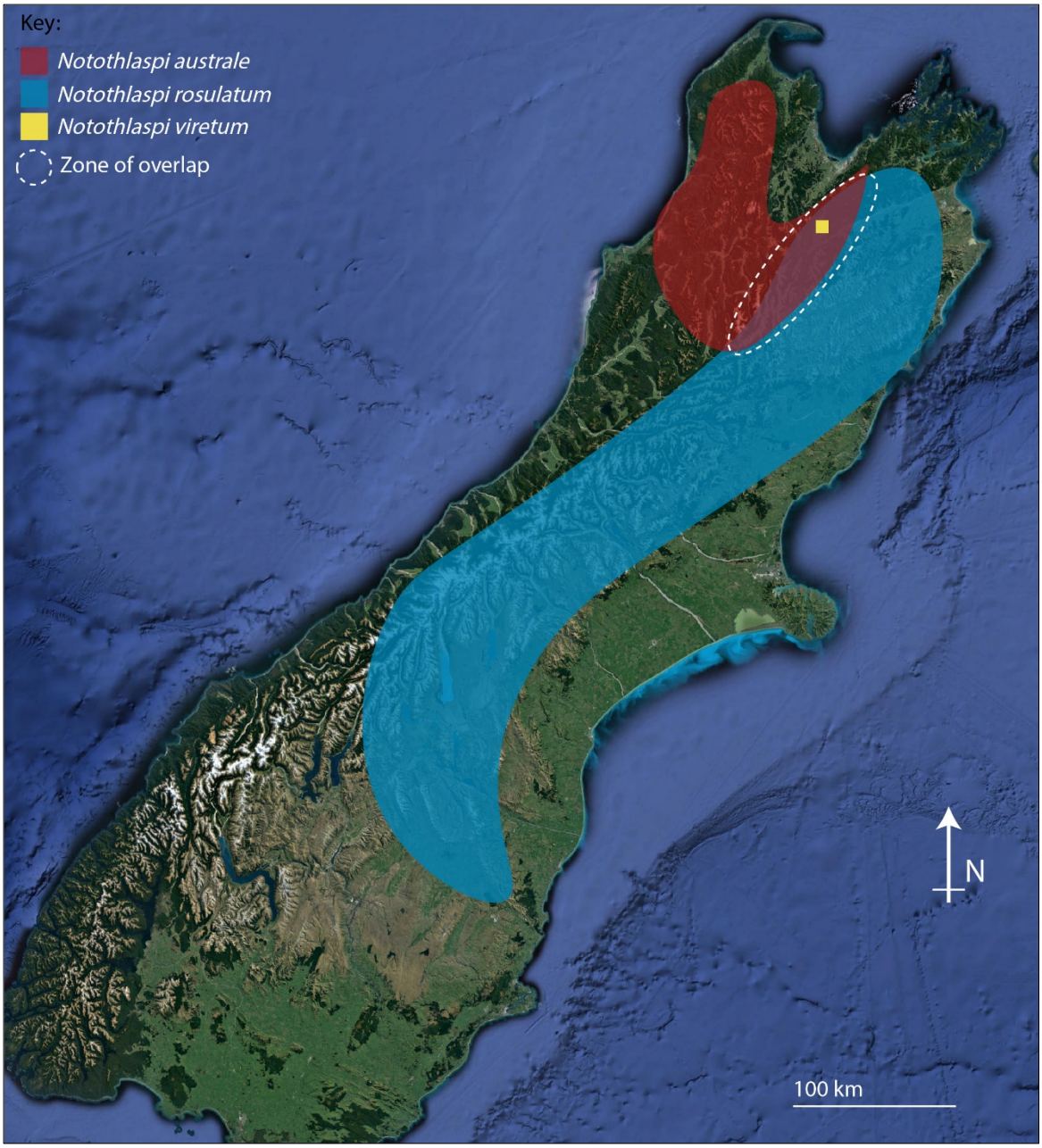


Figure 1.3: Distribution of *Notothlaspi* species, based on herbaria records (Australasian virtual herbarium (https://avh.chah.org.au/)) and iNaturalist (https://www.inaturalist.org/home). *Notothlaspi australe* (red) is primarily found on the north west of the island, while *N. rosulatum* (blue) is found in mountainous areas along the east. Their distributions overlap in the Richmond ranges and the Nelson Lakes areas (within dotted white circle). *N. viretum* is in the yellow square, found sympatric with *N. australe*, but not *N. rosulatum*.

1.5 Research Objectives

Due to its heterogeneous environmental distribution, potential hybrid zone, and varying ploidal levels, the *Notothlaspi* group could be particularly interesting for studying evolution in New Zealand alpine environments. A population genetic study would provide information on the extent of gene flow, showing if plants on different mountain tops are isolated (sky islands) or if pollinators and seed can readily bridge valleys separating populations. It would also provide some indication of divergence due to geographic isolation or local adaptation, and/or hybridisation between species. Therefore, the objectives of this study are to

1. Develop intra-specific and intra-generic markers suitable for population genetics investigations in Notothlaspi,
2. Determine the distribution of genetic variation across the alpine landscape within and among *Notothlaspi* species,
3. Infer historical influences on the adaptive and evolutionary history of *Notothlaspi.*

1.6 References

Abbott, C. L., S. R. Gilmore, C. T. Lewis, J. T. Chapados, R. D. Peters *et al.*, 2010 Development of a SNP genetic marker system based on variation in microsatellite flanking regions of *Phytophthora infestans*. Canadian Journal of Plant Pathology 32: 440-457.

Acosta, M. C., and A. C. Premoli, 2010 Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus,* Nothofagaceae). Molecular Phylogenetics and Evolution 54: 235-242.

Aguirre - Liguori, J. A., B. S. Gaut, J. P. Jaramillo - Correa, M. I. Tenaillon, S. Montes - Hernández *et al.*, 2019 Divergence with gene flow is driven by local adaptation to temperature and soil phosphorus concentration in *teosinte* subspecies (*Zea mays parviglumis* and *Zea mays mexicana*). Molecular Ecology 28: 2814-2830.

Arroyo, J., 2013 Heterozygote excess through life history stages in *Cestrum miradorense* Francey (Solanaceae), an endemic shrub in a fragmented cloud forest habitat. Plant Biology 15: 176-185.

Bai, W.-N., W.-T. Wang and D.-Y. Zhang, 2014 Contrasts between the phylogeographic patterns of chloroplast and nuclear DNA highlight a role for pollen-mediated gene flow in preventing population divergence in an East Asian temperate tree. Molecular Phylogenetics and Evolution 81: 37-48.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver *et al.*, 2008 Rapid SNP Discovery and genetic mapping using sequenced RAD markers. PLOS ONE 3: e3376.

Balloux, F., 2004 Heterozygote excess in small populations and the heterozygote - excess effective population size. Evolution 58: 1891-1900.

Batley, J., R. Mogg, D. Edwards, H. O'Sullivan and K. J. Edwards, 2003 A high-throughput SNuPE assay for genotyping SNPs in the flanking regions of *Zea mays* sequence tagged simple sequence repeats. Molecular Breeding 11: 111-120.

Beck, J. B., H. Schmuths and B. A. Schaal, 2008 Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. Molecular Ecology 17: 902-915.

Behling, A. H., L. D. Shepherd and M. P. Cox, 2020 The importance and prevalence of allopolyploidy in Aotearoa New Zealand. Journal of the Royal Society of New Zealand 50: 189-210.

Bharathi, L. K., Vinod, A. D. Munshi, T. K. Behera, C. Shanti *et al.*, 2010 Cyto-morphological evidence for segmental allopolyploid origin of Teasle gourd (*Momordica subangulata* subsp. *renigera*). Euphytica 176: 79-85.

Bourke, P. M., R. E. Voorrips, R. G. Visser and C. Maliepaard, 2018 Tools for genetic studies in experimental populations of polyploids. Frontiers in Plant Science 9: 513.

Bowers, J. E., B. A. Chapman, J. Rong and A. H. Paterson, 2003 Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422: 433-438.

Bryant, D., and V. Moulton, 2004 Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Molecular Biology and Evolution 21: 255-265.

Byars, S. G., W. Papst and A. A. Hoffmann, 2007 Local adaptation and cogradient selection in the alpine plant, *Poa hiemata*, along a narrow altitudinal gradient. Evolution 61: 2925-2941.

Chen, J., S. Glémin and M. Lascoux, 2020 From drift to draft: How much do beneficial mutations actually contribute to predictions of Ohta's slightly deleterious model of molecular evolution? Genetics 214: 1005-1018.

Cheng, F., J. Wu, X. Cai, J. Liang, M. Freeling *et al.*, 2018 Gene retention, fractionation and subgenome differences in polyploid plants. Nature Plants 4: 258-268.

Clausen, J., D. D. Keck and W. M. Hiesey, 1940 *Experimental studies on the nature of species. I. Effect of varied environments on western North American plants*. Washington DC, Gibson Brothers Inc.

Cockayne, L., and H. Allan, 1934 An annotated list of groups of wild hybrids in the New Zealand flora. Annals of Botany 48: 1-55.

Comai, L., 2000 Genetic and epigenetic interactions in allopolyploid plants. Plant Molecular Biology 43: 387-399.

Coppi, A., A. J. Baker, I. Bettarini, I. Colzi, G. Echevarria *et al.*, 2020 Population genetics of *Odontarrhena* (Brassicaceae) from Albania: The effects of anthropic habitat disturbance, soil, and altitude on a Ni-hyperaccumulator plant group from a major serpentine hotspot. Plants 9: 1686.

Dawson, M. I., and E. J. Beuzenberg, 2000 Contributions to a chromosome atlas of the New Zealand flora—36. Miscellaneous families. New Zealand Journal of Botany 38: 1-23.

Després, L., 2019 One, two or more species? Mitonuclear discordance and species delimitation. Molecular Ecology 28: 3845– 3847.

DeWoody, J. A., A. M. Harder, S. Mathur and J. R. J. M. E. Willoughby, 2021 The long-standing significance of genetic diversity in conservation.

Dittmar, E. L., and D. W. Schemske, 2017 The edaphic environment mediates flowering-time differentiation between adjacent populations of *Leptosiphon parviflorus*. Journal of Heredity 109: 90-99.

Donovan, L. A., D. R. Rosenthal, M. Sanchez - Velenosi, L. H. Rieseberg and F. Ludwig, 2010 Are hybrid species more fit than ancestral parent species in the current hybrid species habitats? Journal of Evolutionary Biology 23: 805-816.

Dufresne, F., M. Stift, R. Vergilino and B. K. Mable, 2014 Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. Molecular Ecology 23: 40-69.

Ehrlich, P. R., and P. H. Raven, 1969 Differentiation of populations. Science 165: 1228-1232.

Ellis, E. A., D. C. Marshall, K. B. Hill, C. L. Owen, P. J. Kamp *et al.*, 2015 Phylogeography of six codistributed New Zealand cicadas and their relationship to multiple biogeographical boundaries suggest a re-evaluation of the Taupo Line. Journal of Biogeography 42: 1761-1775.

Estoup, A., P. Jarne and J.-M. Cornuet, 2002 Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Molecular Ecology 11: 1591-1604.

Fatokun, C., G. Girma, M. Abberton, M. Gedil, N. Unachukwu *et al.*, 2018 Genetic diversity and population structure of a mini-core subset from the world cowpea (*Vigna unguiculata* (L.) Walp.) germplasm collection. Scientific Reports 8: 16035.

Ferrazzini, D., I. Monteleone and P. Belletti, 2007 Genetic variability and divergence among Italian populations of common ash (*Fraxinus excelsior* L.). Annals of Forest Science 64: 159-168.

Ferris, K. G., L. L. Barnett, B. K. Blackman and J. H. Willis, 2017 The genetic architecture of local adaptation and reproductive isolation in sympatry within the *Mimulus guttatus* species complex. Molecular ecology 26: 208-224.

Fields, P. D., D. E. McCauley, E. V. McAssey and D. R. Taylor, 2014 Patterns of cyto-nuclear linkage disequilibrium in *Silene latifolia*: genomic heterogeneity and temporal stability. Heredity 112: 99-104.

Flint-Garcia, S. A., J. M. Thornsberry and E. S. Buckler, 2003 Structure of linkage disequilibrium in Plants. Annual Review of Plant Biology 54: 357-374.

Franklin, I. R., 1980 Evolutionary change in small populations. Conservation biology: an evolutionary-ecological perspective 395.

Gao, Y., B. Ai, H. Kong, M. Kang and H. Huang, 2015 Geographical pattern of isolation and diversification in karst habitat islands: a case study in the *Primulina eburnea* complex. Journal of Biogeography 42: 2131-2144.

Garner, A. G., B. E. Goulet, M. C. Farnitano, Y. F. Molina-Henao and R. Hopkins, 2018 Genomic signatures of reinforcement. Genes 9: 191.

Gauch, H. G., Jr., S. Qian, H.-P. Piepho, L. Zhou and R. Chen, 2019 Consequences of PCA graphs, SNP codings, and PCA variants for elucidating population structure. PLOS ONE 14: e0218306.

Gillespie, J. H., 2000 Genetic drift in an infinite population: The pseudohitchhiking model. Genetics 155: 909-919.

Glenny, D., 2004 A revision of the genus *Gentianella* in New Zealand. New Zealand Journal of Botany 42: 361-530.

Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza and M. W. Feldman, 1995 An evaluation of genetic distances for use with microsatellite loci. Genetics 139: 463-471.

Goremykin, V. V., S. V. Nikiforova, D. Cavalieri, M. Pindo and P. Lockhart, 2015 The root of flowering plants and total evidence. Systematic Biology 64: 879-891.

Gross, B. L., and L. H. Rieseberg, 2005 The ecological genetics of homoploid hybrid speciation. Journal of Heredity 96: 241-252.

Grover, A., and P. C. Sharma, 2016 Development and use of molecular markers: past and present. Critical Reviews in Biotechnology 36: 290-302.

Gurgul, A., A. Miksza-Cybulska, T. Szmatoła, E. Semik-Gurgul, I. Jasielczuk *et al.*, 2019 Evaluation of genotyping by sequencing for population genetics of sibling and hybridizing birds: an example using Syrian and Great Spotted Woodpeckers. Journal of Ornithology 160: 287-294.

Hamrick, J. L., and M. J. W. Godt, 1997 Allozyme diversity in cultivated crops. Crop Science 37: 26-30.

Hansen, A. K., L. K. Escobar, L. E. Gilbert and R. K. Jansen, 2007 Paternal, maternal, and biparental inheritance of the chloroplast genome in *Passiflora* (Passifloraceae): implications for phylogenetic studies. American Journal of Botany 94: 42-46.

Hardy, G. H., 1908 Mendelian proportions in a mixed population. Science 28: 49-50.

Harrison, R., and E. Larson, 2014 Hybridization, introgression, and the nature of species boundaries. Journal of Heredity 105: 795-809.

Harrison, R. G., and E. L. Larson, 2016 Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Molecular Ecology 25: 2454-2466.

Heenan, P., 2019 A taxonomic revision of *Notothlaspi* (Brassicaceae), a specialist alpine genus from New Zealand. Phytotaxa 399: 248-260.

Heenan, P., and M. McGlone, 2013 Evolution of New Zealand alpine and open-habitat plant species during the late Cenozoic. New Zealand Journal of Ecology: 105-113.

Horrocks, M., and J. Ogden, 1998 The effects of the Taupo tephra eruption of c. 1718 BP on the vegetation of Mt Hauhungatahi, central North Island, New Zealand. Journal of Biogeography 25: 649-660.

Ishibashi, Y., T. Saitoh, S. Abe and M. C. Yoshida, 1996 Null microsatellite alleles due to nucleotide sequence variation in the grey-sided vole *Clethrionomys rufocanus*. Molecular Ecology 5: 589-590.

Jakobsson, M., and N. A. Rosenberg, 2007 CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801-1806.

Jiao, Y., N. J. Wickett, S. Ayyampalayam, A. S. Chanderbali, L. Landherr *et al.*, 2011 Ancestral polyploidy in seed plants and angiosperms. Nature 473: 97-100.

Joly, S., P. B. Heenan and P. J. Lockhart, 2009a A Pleistocene inter-tribal allopolyploidization event precedes the species radiation of *Pachycladon* (Brassicaceae) in New Zealand. Molecular Phylogenetics and Evolution 51: 365-372.

Joly, S., P. A. McLenachan and P. J. Lockhart, 2009b A statistical approach for distinguishing hybridization and incomplete lineage sorting. The American Naturalist 174: E54-E70.

Jombart, T., S. Devillard and F. Balloux, 2010 Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11: 94.

Kagale, S., S. J. Robinson, J. Nixon, R. Xiao, T. Huebert *et al.*, 2014 Polyploid evolution of the Brassicaceae during the Cenozoic era. The Plant Cell 26: 2777-2791.

Karunarathne, P., M. Schedler, E. J. Martínez, A. I. Honfi, A. Novichkova *et al.*, 2018 Intraspecific ecological niche divergence and reproductive shifts foster cytotype displacement and provide ecological opportunity to polyploids. Annals of Botany 121: 1183-1196.

Kim, M., M.-L. Cui, P. Cubas, A. Gillies, K. Lee *et al.*, 2008 Regulatory genes control a key morphological and ecological trait transferred between species. Science 322: 1116-1119.

Kimura, M., 1983 *The neutral theory of molecular evolution*. Cambridge University Press.

Kimura, M., and T. Ohta, 1969 The average number of generations until fixation of a mutant gene in a finite population. Genetics 61: 763.

Kimura, M., and T. Ohta, 1971 *Theoretical aspects of population genetics*. Princeton University Press.

Kirk, H., and J. R. Freeland, 2011 Applications and implications of neutral versus non-neutral markers in molecular ecology. International Journal of Molecular Sciences 12: 3966-3988.

Kishor, D., W.-H. Song, Y. Noh, G. P. Lee, Y. Park *et al.*, 2020 Development of SNP markers and validation assays in commercial Korean melon cultivars, using Genotyping-by-sequencing and Fluidigm analyses. Scientia Horticulturae 263: 109113.

Koenig, R., and P. Gepts, 1989 Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of genetic diversity. Theoretical and Applied Genetics 78: 809-817.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg and I. Mayrose, 2015 Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15: 1179-1191.

Le Comber, S., M. Ainouche, A. Kovarik and A. Leitch, 2010 Making a functional diploid: from polysomic to disomic inheritance. New Phytologist 186: 113-122.

Leffler, E. M., K. Bullaughey, D. R. Matute, W. K. Meyer, L. Segurel *et al.*, 2012 Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biology 10.

Leimu, R., and M. Fischer, 2008 A meta-analysis of local adaptation in plants. PloS One 3: e4010.

Lenormand, T., 2002 Gene flow and the limits to natural selection. Trends in Ecology & Evolution 17: 183-189.

Levinson, G., and G. A. Gutman, 1987 Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Molecular Biology and Evolution 4: 203-221.

Liu, B.-B., C. S. Campbell, D.-Y. Hong and J. Wen, 2020 Phylogenetic relationships and chloroplast capture in the *Amelanchier-Malacomeles-Peraphyllum* clade (Maleae, Rosaceae): Evidence from chloroplast genome and nuclear ribosomal DNA data using genome skimming. Molecular Phylogenetics and Evolution 147: 106784.

Lloyd, A., and K. Bomblies, 2016 Meiosis in autopolyploid and allopolyploid *Arabidopsis*. Current Opinion in Plant Biology 30: 116-122.

Lockhart, P. J., P. A. McLenachan, D. Havell, D. Glenny, D. Huson *et al.*, 2001 Phylogeny, radiation, and transoceanic dispersal of New Zealand alpine buttercups: molecular evidence under split decomposition. Annals of the Missouri Botanical Garden: 458-477.

Mandel, J. R., J. M. Dechaine, L. F. Marek and J. M. Burke, 2011 Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. Theoretical and Applied Genetics 123: 693-704.

Marra, M. J., and G. D. Thackray, 2010 Glacial forest refugium in Howard Valley, South Island, New Zealand. Journal of Quaternary Science: Published for the Quaternary Research Association 25: 309-319.

Mason, A. S., and J. C. Pires, 2015 Unreduced gametes: meiotic mishap or evolutionary mechanism? Trends in Genetics 31: 5-10.

Mason, A. S., and J. F. Wendel, 2020 Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. Frontiers in Genetics 11: 1014.

Mayr, E., 1942 *Systematics and the origin of species, from the viewpoint of a zoologist*. New York, Columbia University Press.

McGlone, M., R. Duncan and P. Heenan, 2001 Endemism, species selection and the origin and distribution of the vascular plant flora of New Zealand. Journal of Biogeography 28: 199-216.

McKenzie, R. J., 2001 Intergeneric hybridisation in New Zealand Gnaphalieae (Compositae), PhD in *Botany*. University of Canterbury.

McKenzie, R. J., J. M. Ward, J. D. Lovis and I. Breitwieser, 2004 Morphological evidence for natural intergeneric hybridization in the New Zealand Gnaphalieae (Compositae): *Anaphalioides bellidioides* × *Ewartia sinclairii*. Botanical Journal of the Linnean Society 145: 59-75.

McLay, T., 2011 Conservation genetics of the endemic root holoparasite, *Dactylanthus taylorii*,. MSc in *School of Fundamental Sciences*. Massey University, Massey University, Manawatu, New Zealand.

McLay, T., 2018 Hybridisation between species of *Xanthorrhoea* in Yuraygir National Park revealed using a newly developed next-generation sequencing method. Australian Systematic Botany Society Newsletter: 13-16.

Meirmans, P. G., S. Liu and P. H. van Tienderen, 2018 The Analysis of Polyploid Genetic Data. Journal of Heredity 109: 283-296.

Melo, M. C., A. Grealy, B. Brittain, G. M. Walter and D. Ortiz-Barrientos, 2014 Strong extrinsic reproductive isolation between parapatric populations of an Australian groundsel. New Phytologist 203: 323-334.

Meudt, H. M., D. C. Albach, A. J. Tanentzap, J. Igea, S. C. Newmarch *et al.*, 2021 Polyploidy on Islands: Its Emergence and Importance for Diversification. Frontiers in Plant Science 12: 637214-637214.

Meudt, H. M., J. M. Prebble and C. A. Lehnebach, 2015 Native New Zealand forget-me-nots (*Myosotis*, Boraginaceae) comprise a Pleistocene species radiation with very low genetic divergence. Plant Systematics and Evolution 301: 1455-1471.

Mills, L. S., and F. W. Allendorf, 1996 The one-migrant-per-generation rule in conservation and management. Conservation Biology 10: 1509-1518.

Mitchell, A. D., and P. B. Heenan, 2000 Systematic relationships of New Zealand endemic Brassicaceae inferred from nrDNA ITS sequence data. Systematic Botany 25: 98-105.

Musker, S. D., A. G. Ellis, S. A. Schlebusch and G. A. Verboom, 2021 Niche specificity influences gene flow across fine-scale habitat mosaics in Succulent Karoo plants. Molecular Ecology 30: 175-192.

Nakamura, S., 2010 Paternal inheritance of mitochondria in *Chlamydomonas*. Journal of Plant Research 123: 163-170.

Neale, D. B., K. A. Marshall and R. R. Sederoff, 1989 Chloroplast and mitochondrial DNA are paternally inherited in *Sequoia* D. Don Endl. Proceedings of the National Academy of Sciences 86: 9347-9349.

Neher, R. A., 2013 Genetic draft, selective interference, and population genetics of rapid adaptation. Annual review of Ecology, Evolution, and Systematics 44: 195-215.

Nei, M., 1972 Genetic Distance between Populations. The American Naturalist 106: 283-292.

Newnham, R., M. McGlone, N. Moar, J. Wilmshurst and M. Vandergoes, 2013 The vegetation cover of New Zealand at the Last Glacial Maximum. Quaternary Science Reviews 74: 202-214.

Niu, S., Q. Song, H. Koiwa, D. Qiao, D. Zhao *et al.*, 2019 Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (*Camellia sinensis*) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. BMC Plant Biology 19: 328.

Nosil, P., D. J. Funk and D. Ortiz-Barrientos, 2009 Divergent selection and heterogeneous genomic divergence. Molecular Ecology 18: 375-402.

Pan, D., P. Schönswetter, T. Moser, E. Vitek and G. M. Schneeweiss, 2019 Ancestral remnants or peripheral segregates? Phylogenetic relationships of two narrowly endemic *Euphrasia* species (Orobanchaceae) from the eastern European Alps. AoB Plants 11.

Pereira-Dias, L., S. Vilanova, A. Fita, J. Prohens and A. Rodríguez-Burruezo, 2019 Genetic diversity, population structure, and relationships in a collection of pepper (*Capsicum* spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). Horticulture Research 6: 54.

Perrie, L. R., L. D. Shepherd, P. J. De Lange and P. J. Brownsey, 2010 Parallel polyploid speciation: distinct sympatric gene-pools of recurrently derived allo-octoploid *Asplenium* ferns. Molecular Ecology 19: 2916-2932.

Pilkington, K. M., 2019 The genetic architecture of the divaricate growth form: A QTL mapping approach in *Sophora* (Fabaceae), pp186-193 PhD. in *School of Fundamental Sciences*. Massey University, Massey University, Manawatu, New Zealand.

Prevosti, A., J. Ocana and G. Alonso, 1975 Distances between populations of *Drosophila subobscura*, based on chromosome arrangement frequencies. Theoretical and Applied Genetics 45: 231-241.

Prezeworski, M., G. Coop and J. D. Wall, 2005 The signature of positive selection on standing genetic variation. Evolution 59: 2312-2323.

Pritchard, J. K., M. Stephens and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Ramsey, J., and D. W. Schemske, 1998 Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annual Review of Ecology and Systematics 29: 467-501.

Rattenbury, J. A., 1962 Cyclic hybridization as a survival mechanism in the new zealand forest flora. Evolution 16: 348-363.

Reich, D., A. L. Price and N. Patterson, 2008 Principal component analysis of genetic data. Nature Genetics 40: 491-492.

Reynolds, J., B. S. Weir and C. C. Cockerham, 1983 Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105: 767-779.

Robertson, A., and W. G. Hill, 1984 Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. Genetics 107: 703-718.

Rogalski, M., L. do Nascimento Vieira, H. P. Fraga and M. P. Guerra, 2015 Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. Frontiers in Plant Science 6: 586.

Rosenberg, N. A., 2004 distruct: a program for the graphical display of population structure. Molecular Ecology Notes 4: 137-138.

Sambatti, J. B., and K. J. Rice, 2006 Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). Evolution 60: 696-710.

Schierenbeck, K. A., 2017 Population-level genetic variation and climate change in a biodiversity hotspot. Annals of Botany 119: 215-228.

Schlötterer, C., 2004 The evolution of molecular markers — just a matter of fashion? Nature Reviews Genetics 5: 63-69.

Schumer, M., G. G. Rosenthal and P. Andolfatto, 2014 How common is homoploid hybrid speciation? Evolution 68: 1553-1560.

Schuster, T. M., S. D. Setaro, J. F. Tibbits, E. L. Batty, R. M. Fowler *et al.*, 2018 Chloroplast variation is incongruent with classification of the Australian bloodwood eucalypts (genus *Corymbia*, family Myrtaceae). PloS One 13: e0195034.

Segatto, A. L. A., A. L. R. Cazé, C. Turchetto, U. Klahre, C. Kuhlemeier *et al.*, 2014 Nuclear and plastid markers reveal the persistence of genetic identity: A new

perspective on the evolutionary history of *Petunia exserta*. Molecular Phylogenetics and Evolution 70: 504-512.

Selkoe, K. A., and R. J. Toonen, 2006 Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters 9: 615-629.

Shepherd, L. D., L. R. Perrie and P. Brownsey, 2007 Fire and ice: volcanic and glacial impacts on the phylogeography of the New Zealand forest fern *Asplenium hookerianum*. Molecular Ecology 16: 4536-4549.

Singh, S. P., R. Nodari and P. Gepts, 1991 Genetic diversity in cultivated common bean: I. Allozymes. Crop Science 31: 19-23.

Smissen, R., and I. Breitwieser, 2008 Species relationships and genetic variation in the New Zealand endemic *Leucogenes* (Asteraceae: Gnaphalieae). New Zealand Journal of Botany 46: 65-76.

Smissen, R., I. Breitwieser and J. Ward, 2004 Phylogenetic implications of trans-specific chloroplast DNA sequence polymorphism in New Zealand Gnaphalieae (Asteraceae). Plant Systematics and Evolution 249: 37-53.

Smissen, R. D., and P. B. Heenan, 2007 DNA fingerprinting supports hybridisation as a factor explaining complex natural variation in *Phormium* (Hemerocallidaceae). New Zealand Journal of Botany 45: 419-432.

Soltis, D., V. A. Albert, J. Leebens-Mack, C. D. Bell, A. H. Paterson *et al.*, 2009 Polyploidy and angiosperm diversification. American Journal of Botany 96: 336-348.

Soltis, P. S., and D. E. Soltis, 2009 The role of hybridization in plant speciation. Annual Review of Plant Biology 60: 561-588.

Stoeckel, S., J. Grange, J. F. FERNÁNDEZ - MANJARRES, I. Bilger, N. FRASCARIA - LACOSTE *et al.*, 2006 Heterozygote excess in a self-incompatible and partially clonal forest tree species- *Prunus avium* L. Molecular Ecology 15: 2109-2118.

Suarez-Gonzalez, A., C. Lexer and Q. C. Cronk, 2018 Adaptive introgression: a plant perspective. Biology Letters 14: 20170688.

Tamayo-Ordóñez, M., L. Espinosa-Barrera, Y. Tamayo-Ordóñez, B. Ayil-Gutiérrez and L. Sánchez-Teyer, 2016 Advances and perspectives in the generation of polyploid plant species. Euphytica 209: 1-22.

Telwala, Y., B. W. Brook, K. Manish and M. K. Pandit, 2013 Climate-induced elevational range shifts and increase in plant species richness in a himalayan biodiversity Epicentre. PLOS One 8: e57103.

Thórsson, Æ. T., E. Salmela and K. Anamthawat-Jónsson, 2001 Morphological, cytogenetic, and molecular evidence for introgressive hybridization in birch. Journal of Heredity 92: 404-408.

Tóth, G., Z. Gáspári and J. Jurka, 2000 Microsatellites in different eukaryotic genomes: survey and analysis. Genome Research 10: 967-981.

Turesson, G., 1922 The genotypical response of the plant species to the habitat. Hereditas 3: 211-350.

Turner, T. L., M. W. Hahn and S. V. Nuzhdin, 2005 Genomic islands of speciation in *Anopheles gambia*. PLoS Biology 3: e285.

Van de Peer, Y., E. Mizrachi and K. Marchal, 2017 The evolutionary significance of polyploidy. Nature Reviews Genetics 18: 411-424.

Van Inghelandt, D., A. E. Melchinger, C. Lebreton and B. Stich, 2010 Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theoretical and Applied Genetics 120: 1289-1299.

Via, S., 2009 Natural selection in action during speciation. Proceedings of the National Academy of Sciences 106: 9939-9946.

Wagstaff, S. J., and I. Breitwieser, 2004 Phylogeny and classification of *Brachyglottis* (Senecioneae, Asteraceae): an example of a rapid species radiation in New Zealand. Systematic Botany 29: 1003-1010.

Wagstaff, S. J., P. B. Heenan and M. J. Sanderson, 1999 Classification, origins, and patterns of diversification in New Zealand *Carmichaelinae* (Fabaceae). American Journal of Botany 86: 1346-1356.

Wallis, G. P., J. M. Waters, P. Upton and D. Craw, 2016 Transverse Alpine Speciation Driven by Glaciation. Trends in Ecology & Evolution 31: 916-926.

Walter, G. M., M. J. Wilkinson, M. E. James, T. J. Richards, J. D. Aguirre *et al.*, 2016 Diversification across a heterogeneous landscape. Evolution 70: 1979-1992.

Weinberg, W., 1908 *Über den nachweis der vererbung beim menschen*.

Whitney, K. D., R. A. Randell and L. H. Rieseberg, 2006 Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. The American Naturalist 167: 794-807.

Wichman, S. R., S. D. Wright, E. K. Cameron, D. J. Keeling and R. C. Gardner, 2002 Elevated genetic heterogeneity and Pleistocene climatic instability: inferences from nrDNA in New Zealand *Coprosma* (Rubiaceae). Journal of Biogeography 29: 943-954.

Winkworth, R. C., S. J. Wagstaff, D. Glenny and P. J. Lockhart, 2002 Plant dispersal N.E.W.S from New Zealand. Trends in Ecology & Evolution 17: 514-520.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon *et al.*, 2009 The frequency of polyploid speciation in vascular plants. Proceedings of the National Academy of Sciences 106: 13875-13879.

Wright, S., 1949 The genetical structure of populations. Annals of Eugenics 15: 323-354.

Yamamoto, M., D. Takahashi, K. Horita and H. Setoguchi, 2020 Speciation and subsequent secondary contact in two edaphic endemic primroses driven by Pleistocene climatic oscillation. Heredity 124: 93-107.

Yeaman, S., and S. P. Otto, 2011 Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. Evolution: International Journal of Organic Evolution 65: 2123-2129.

Zardoya, R., D. M. Vollmer, C. Craddock, J. T. Streelman, S. Karl *et al.*, 1996 Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (*Pisces: Perciformes*). Proceedings of the Royal Society of London. Series B: Biological Sciences 263: 1589-1598.

Zhou, Y., L. Duvaux, G. Ren, L. Zhang, O. Savolainen *et al.*, 2017 Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. Heredity 118: 211-220.

Chapter 2

Marker Development

2.1 Abstract

The New Zealand endemic genus *Notothlaspi* is comprised of three species, *N. australe, N. rosulatum,* and *N. viretum.* For future population genetic analysis, three molecular marker systems were developed and/or trialled: microsatellite markers, universal chloroplast markers, and the RAPDseq method. Twelve microsatellite markers were designed from publicly available *N. australe* whole genome sequencing reads. The results showed all twelve markers to be suitable for *N. australe,* but in *N. viretum* most markers amplified multiploid alleles, which reduces their downstream utility. There was also a higher frequency of missing data in *N. rosulatum*, which may reflect ascertainment bias. Regardless, all twelve markers have potential for future population genetic analysis of *N. australe* and *N. rosulatum*: they are polymorphic, easily scorable, and can be multiplexed by six to save time and genotyping costs. Of the five universal chloroplast loci trialled, the trnL-F and psbA-trnH loci are suitable for *Notothlaspi*. Both were polymorphic within and between *N. australe* and *N. rosulatum*, however, homopolymers within each sequence prevented bi-directional sequencing. The RAPDseq scheme produced 41 informative loci, across the twelve *N. australe* and *N. rosulatum* individuals sampled; *N. viretum* was unavailable at the time of sampling. This approach appears to have great potential for use generating population genetic markers. The markers

developed here will be valuable for future interspecific and intraspecific analysis of the *Notothlaspi* genus.

2.2 Introduction

Fundamentally, populations change through time via shifts in allele frequencies. To study evolution, an understanding of processes affecting changes in allele frequencies is important. Some of these processes are non-adaptive (i.e. genetic drift or draft), and some are in response to natural selection (i.e. local adaptation) (see chapter 1); both categories contribute to short- and long-term evolutionary dynamics and the success of a lineage (BUFFALO AND COOP 2020).

The New Zealand flora is an excellent group for studying processes around divergence and hybridisation. Most endemic plant genera have arrived recently (in respect to geological time) via long distance dispersal, and the recent Pleistocene climate changes and mountain uplift are thought to have provided open niches to support evolutionary radiations (WINKWORTH *et al.* 2005; HEENAN AND MCGLONE 2013). Therefore, while several plant genera are morphologically diverse, often congeners are not strongly genetically diverged (SMISSEN AND BREITWIESER 2008; MORGAN-RICHARDS *et al.* 2009; WAGSTAFF *et al.* 2010; HEENAN *et al.* 2017; MILLAR *et al.* 2017).

Notothlaspi Hook. (Brassicaceae) is a New Zealand endemic genus comprised of three species (HEENAN 2019), *N. australe, N. rosulatum,* and *N. viretum.* All three species grow in rocky, alpine environments on the South Island. *Notothlaspi australe* and *N. rosulatum* are locally abundant within their range and are considered 'Not Threatened' under the New Zealand classification system. *Notothlaspi viretum* is classified as 'Threatened, Nationally Critical' and is only found within a 1 hectare area (DE LANGE AND

I. 2018; HEENAN 2019) that overlaps with the distribution of both *N. australe* and *N. rosulatum*. While ploidy is unknown, all species have had at least one chromosome count: 2n= 36-38 for *N. rosulatum*, and 2n = 90-100 for *N. viretum* and *N. australe* (DAWSON AND BEUZENBERG 2000). *N. australe* is of particular interest in this work, as the species grows on a wide range of geological soil types that typically have high rates of endemism, so it may be an interesting case for studying local adaptation and intraspecific divergence. Given the range overlap and potential for hybridization among *Notothlaspi* species, any consideration of genetic variation within *N. australe* will benefit from genetic data from its congeners.

To detect temporal or spatial patterns of change, several types of molecular marker can be employed. Different marker types can reveal different patterns or relationships, so it is vital to select the appropriate combination for a specific hypothesis and biological system (ARIS-BROSOU AND EXCOFFIER 1996; PATWARDHAN *et al.* 2014; GROVER AND SHARMA 2016). Studies can target adaptive or conserved loci which have faster or slower rates of change, or 'neutral' loci that are often assumed to accumulate mutations at a background mutation rate that may better represent the genome as a whole (KIRK AND FREELAND 2011), depending on the questions being addressed. Markers with different mutation rates can be used to study different scales of divergence. For instance, highly variable markers are suitable for a population genetic study, as they may be able to differentiate between individuals of the same population. However, when using the same markers to investigate old divergence events, reverse mutations leading to homoplasy are possible. Different genomes are also subject to different processes, for example, nucleotide substitution rate varies among the nuclear, chloroplast, and mitochondrial genomes (WOLFE *et al.* 1987). To this end, we evaluate three marker systems for potential use in *Notothlaspi*: microsatellite; RAPDseq; and universal chloroplast markers.

Microsatellite markers are simple sequence repeat (SSR) regions. Slipped strand mispairing during DNA replication adds or removes repeat units, creating length polymorphisms (LEVINSON AND GUTMAN 1987). They are typically highly variable, multiallelic, co-dominant, and relatively inexpensive to develop and genotype, making them a popular choice for population studies (GUICHOUX *et al.* 2011; KALIA *et al.* 2011). Like many nuclear markers, using microsatellite loci for polyploids can lead to an over-estimation of heterozygosity, increased issues with null alleles, and difficulty determining allelic dosage (reviewed in DUFRESNE *et al.* (2014)). If there is doubt as to the number of loci a given marker amplifies, results may be scored as dominant with presence/absence distributions (RODZEN AND MAY 2002). Here, using a publicly available HiSeq generated Illumina whole genome sequencing reads, microsatellite markers are generated and trialled on the three *Notothlaspi* species.

Plastid DNA has not been shown to duplicate alongside genomic DNA during whole genome duplication events. Therefore, compared with nuclear markers, there tend to be fewer issues with polyploidy, although there are duplicated genes within the plastid genomes (PIRIE *et al.* 2007; BENNETT *et al.* 2017). Universal chloroplast markers are commonly used in systematics, and in conjunction with nuclear markers, can detect hybridisation and introgression (see chapter 1). Compared to highly mutable markers like nuclear microsatellites, chloroplast markers are known to change more slowly, and are thus not as commonly used intraspecifically, although there are many examples in the literature (LI *et al.* 2021; NIEVES-ORDUÑA *et al.* 2021; XU *et al.* 2021). Due to their variability in other systems, in the present study, the trnL-F, psbA-trnH, ndhF, and rpl16 loci were selected for trialling in *Notothlaspi*. Successful markers will amplify a product across the *Notothlaspi* species and reveal inter-population polymorphisms.

RAPDseq (randomly amplified polymorphic DNA sequences) is a system to reduce the analytical complexity of the genome using random primers to produce amplicons from across the genome for sequencing (McLay 2018). By comparing and aligning sequenced PCR amplicons bioinformatically, SNPs (Single Nucleotide Polymorphisms) can be identified to investigate heterozygosity within individuals and polymorphisms between individuals. Using this method, polymorphic characters can be detected at many sites genome-wide, providing more widespread data than microsatellites, although assumptions of neutrality are harder to make. In this study, a pilot project was run to determine the suitability of this marker system for studying the population genetics of *Notothlaspi*.

Chapter 1 outlines three objectives for this study; this chapter addresses objective one and develops the tools for future investigations into objectives two and three. Objective one is to "Develop intra-specific and intra-generic markers suitable for population genetics investigations in *Notothlaspi*".

2.3 Methods

2.3.1 Taxon Sampling and DNA extraction

Leaves were collected from 824 *N. australe* and *N. rosulatum* individuals from 64 sites in 18 wider locations (Figure 2.1). Based on morphological characteristics, individuals from location L were not assigned to a species. One to five leaves were taken from each plant and young leaves were preferentially selected. Each individual was photographed, and each site has a GPS coordinate (Supplementary Table 1). The leaves were stored in sealed bags with silica gel for desiccation. As *N. viretum* was unable to be collected from the field, leaf samples from 13 herbarium specimens were requested from, and provided

by, the Allan herbarium (Manaaki Whenua – Landcare research): Lincoln (CHR5330233, CHR512603, CHR512597, CHR512595, CHR514589, CHR512596, CHR532706, CHR512602, CHR512601, CHR512598, CHR637799, CHR637800, CHR512600, CHR512599). Fifty individuals from across different sites and species were selected for marker testing (Table 2.1).

Before DNA extraction, leaf tissue was allowed to dry for a minimum of two weeks. Due to difficulties extracting high quality and quantity DNA from *Notothlaspi*, three DNA extraction methods were trialled, a modified STE CTAB method (SHEPHERD AND MCLAY 2011), the DNeasy plant mini kit (Qiagen, Hilden, Germany), and the DNeasy Power Plant Pro kit (Qiagen). Manufacturer's instructions were followed from the DNeasy plant mini kit, except a final elution volume of 100µL, rather than 200µL. The DNeasy power plant pro kit has several options. The method used began with tissue disruption in a bead basher at 5000rpm for 90s. 410µL of bead solution and 40µL PSS solution (provided in the kit) were used. The tubes were then shaken for 10 minutes by taping them onto a vortexer at full speed. Following this, manufacturer's instructions were followed, except 250µL were added at step five, and only a single elution of 100µL at step thirteen. DNA presence and quality were checked on a TBE based agarose gel.

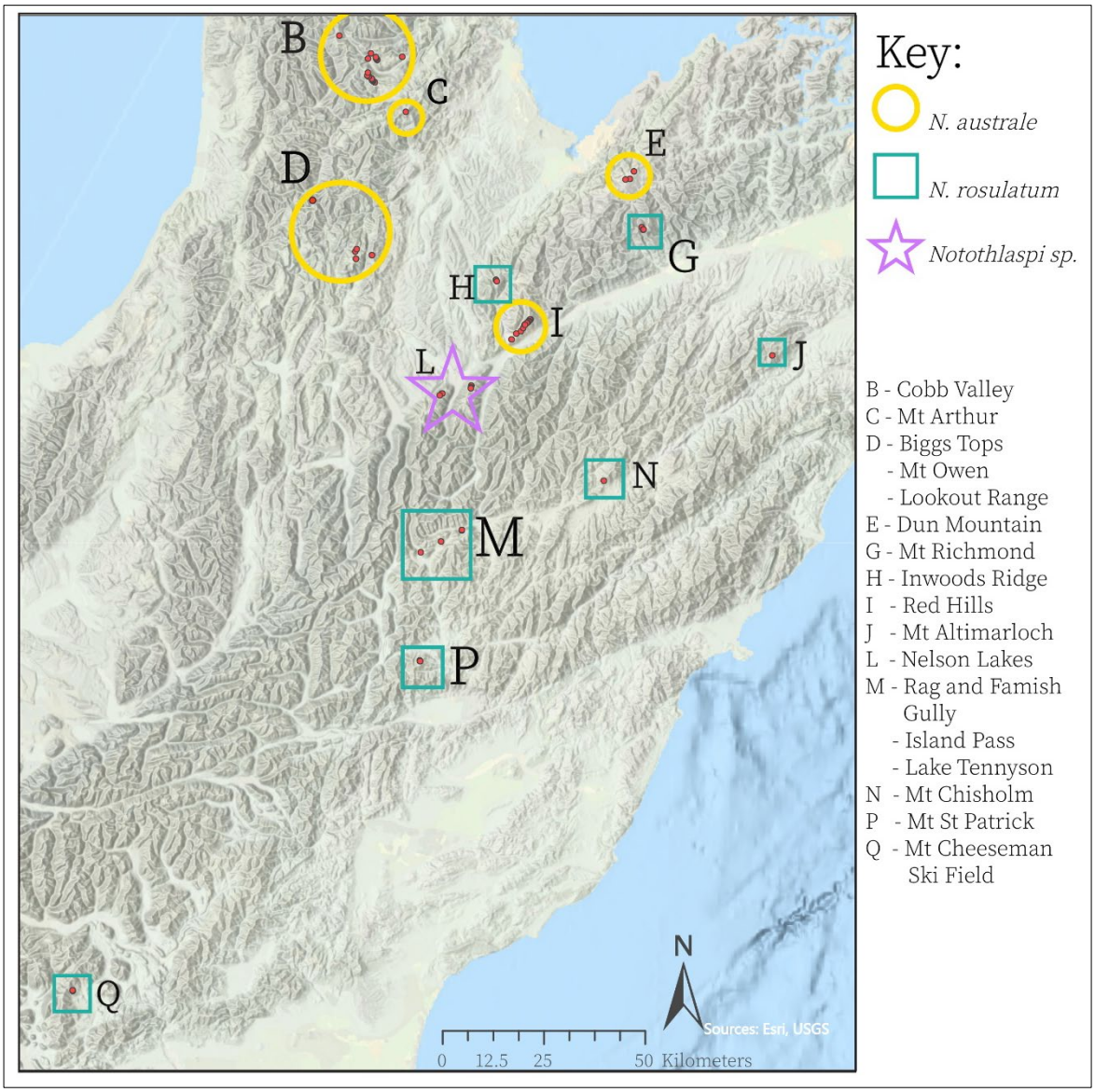


Figure 2.1: Map of sampling sites. The small red circles are GPS points recorded at each collecting site. Large yellow circles around *N. australe* sites are labelled with the location code. Green squares surround *N. rosulatum* sites. Inside the purple star are sites that were unable to be identified to species using morphology. Location names and key are given to the right of the map.

Table 2.1: The individuals and locations used for trialling each marker type. Individual codes describe the location (the letter), site (first number), and individual (last number). Highlighted columns indicate which analysis method the DNA from each individual was used for. Yellow highlighted columns are *N. australe*, purple columns are unknown *Notothlaspi* sp., green is *N. rosulatum,* and orange are *N. viretum.*

Location	Individual code	Microsatellites	Chloroplast loci	RAPDseq
Cobb Valley	B10_01			
	B17_17			
	B17-15			
	B4_01			
Mt Owen	D1_11			
Lookout Range	D4_10			
Cobb Valley	E1_02			
	E1_30			
	E3-15			
Red Hills	I10-11			
	I6_02			
	I7_01			
Nelson Lakes	Lr1_07			
	Lr1_23			
Mt Richmond	G2_03			
	G2_16			
Inwoods Ridge	H1_02			
	H1_06			
	H1-18			
Mt Altimarloch	J1_01			
	J1_02			
Rag and Famish Gully	M1-14			
	M4_02			
	M4_17			
Mt Chisholm	N1_01			

Location	Individual code	Microsatellites	Chloroplast loci	RAPDseq
Mt St Patrick	P1_08			
	P1-11			
Cheeseman Ski Field	Q2_02			
	Q2_16			
Allan Herbarium	CHR512597			
	CHR512603			
Cobb Valley	B2-02			
Red Hills	I7-01			
Inwoods Ridge	H1-18			
Cheeseman Ski Field	Q2-04			
Cobb Valley	B5-06			
	B5-10			
	B18-03			
	B18-04			
Dun Mountain	E3-01			
Red Hills	I8-05			
	Lr1-11			
	Lr1-20			
Inwoods Ridge	H2-12			
	H2-14			
Mt St Patrick	P3-01			
	P3-09			

2.3.2 Microsatellite Markers

For microsatellite marker development, raw reads from an unspecified paired end whole genome sequencing run were used. The data were accessed through NCBI (NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION 1988) in an SRA (Sequencing Read Archive) file, under accession ERX4171254 (https://www.ncbi.nlm.nih.gov/sra/ERX4171254[accn]), uploaded by WALDEN *et al.* (2020) in June 2020. These were quality checked using FastQC (ANDREWS 2010), and

subsequently uploaded to the galaxy web platform (AFGAN *et al.* 2018) for trimming using Trim-Galore (Krueger F. Trim-Galore!, accessible at http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The output was checked using FastQC again, and subsequently assembled using SPAdes (BANKEVICH *et al.* 2012) in the galaxy web platform (AFGAN *et al.* 2018). FastQC reported 13,012,372 reads prior to trimming, and 12,980,838 reads following trimming. The assembly was viewed in Geneious v9.1.8 (KEARSE *et al.* 2012), where di-, tri-, and tetra- nucleotide repeat motifs, with seven to twelve repeat units were searched for. The contigs containing suitable SSRs were BLAST (Basic Local Alignment Search Tool) searched in Geneious to check for contamination or plastid sequence. Primers were designed in Geneious using primer3 v2.3.4. (ROZEN AND SKALETSKY 2000) from >60bp away from the repeat region, with a minimum PCR product size of 150bp, a maximum PCR product size of 400bp, a primer size of 18-22bp, a 5' GC clamp of 1, and primer annealing temperature of 54–59°C. Thirty primer pairs were designed, one was selected per contig and one per locus. We aimed to select primers with different target sizes, to prevent issues with multiplexing for genotyping. On the 5' end of each forward primer, a 19bp M13 sequence was added, and a pigtail sequence was added to the 5' end of each reverse primer for all markers except Na0003, Na0298, Na0011, and Na0010, to ensure consistent non-template adenylation (BROWNSTEIN *et al.* 1996). There were no obvious issues with non-template adenylation for Na0003, Na0298, Na0011, or Na0010.

Notothlaspi genomic DNA for trialling the microsatellite loci was extracted using all three methods mentioned above. Initially, the primers were screened against 7 individuals from a broad geographic distribution and representing all three *Notothlaspi* species (Table 2.1). For each microsatellite amplification, the PCR cocktail, in a 10μL volume, contained ddH2O, 1X Buffer BD (Solis Biodyne, Tartu Estonia), 2.5μM $MgCl_2$,

10μM of each dNTP, 0.2μM forward primer, 4.5μM reverse primer, 4.5μM M13 Dye (FAM, VIC, or NED), 0.5U Firepol *Taq* Polymerase (Solis Biodyne, Tartu Estonia), and 1μL full strength DNA. PCR amplification began with a denaturation step at 95°C for three minutes, followed by 30 cycles of 95°C for 30s, 52°C for 30s, and 72°C for 45s, 72°C for 20 minutes then followed. Each PCR reaction was checked on a 1.5% agarose gel stained with ethidium bromide. PCR products from three different loci, each with a different dye incorporated, were initially pooled, and 1.25μL of this pooled product was added to a 9μL mix of Hi-Di formamide (Applied Biosystems, Carlsbad, California, USA) and CASS size standard (SYMONDS AND LLOYD 2004). Results were viewed and manually scored in GENEMAPPER (v4.0) (Applied Biosystems). Following initial trials, it was evident the PCR products were strong enough to multiplex by six, so pairs of markers without overlapping size ranges were selected for the same M13 dye. Twelve suitable markers (Table 2.2) were identified based on successful amplification across the three *Notothlaspi* species, and the maximum number of alleles amplified within individuals, with preference given to diploid loci for ease of later analysis. Selected markers were trialled with a further 23 individuals (Table 2.1), using the same method described above. Observed heterozygosity and the number of alleles were determined using GenALEx v6.5 (PEAKALL AND SMOUSE 2006; PEAKALL AND SMOUSE 2012).

Table 2.2: Microsatellite primer information. Primer sequences, repeat motif, allele size range, and the M13 fluorescent dye used during amplification.

Locus	Direction	Primer sequences	Repeat Motif	Size range (bp)	M13 Dye
Na0003	F	ACCAGGTTCTGATGCAATAG	AT	175-261	FAM
	R	AATAAGGGACCGTTAATGGG			
Na0010	F	CTACGCCTTTTTGTTGAAGG	AT	348-374	FAM
	R	GGGTATCACAGTTTGGAAC			
Na0011	F	GTCGAAGAACTGAAAGAAGC	AT	206-268	NED
	R	GAACCAGTTTGTATTCCGAC			
Na0078	F	TAATCATGGCCCTGTAAGTG	TGA	261-376	NED
	R	AACTGAGAGAAGGGGATTTC			
Na0298	F	GAAGCAGAAAATACGGTTGG	AC	221-264	VIC
	R	AAGGAATGAGAATAAGGGGG			
Na27207	F	GGGCTGTTTTCGACTAAATC	CTT	262-294	VIC
	R	ATCTATCAACTCGCCAACTC			
Na11999	F	TGAAACTGCTAAACCATTGC	CTC	180-208	FAM
	R	GTGTCCTTTGAATTTTGGGG			
Na1332	F	TACGTTTTTGCTTCTGGTTG	GAG	278-324	VIC
	R	CGAGATACTGAGGGACAAAG			
Na20965	F	GCTTACATGGCAAAGGATTC	GAG	330-362	NED
	R	TTCTGGGCCTATTTAGCTTG			
Na8420	F	TTTTGCAGTAGCTCTACACC	GAAA	282-301	NED
	R	GATCTCCCCTAATCGAACAC			
Na6568	F	TTCAATGGTGTCTCGACTTC	GTG	310-335	FAM
	R	TAAACTGCACGATGGACTAC			
Na0043	F	TTCATGAGGTTAGGGTTTCC	TTC	344-374	VIC
	R	TGAGAAATCCGTCGAAGAAG			

2.3.3 Chloroplast Markers

To determine if universal chloroplast markers were variable enough within *Notothlaspi*, five loci were initially trialled: trnL-F, rpL16, ndhF, rps16, and psbA-trnH

(Table 2.3) (TABERLET *et al.* 1991; OLMSTEAD AND SWEERE 1994; JORDAN *et al.* 1996; SHAW *et al.* 2007). DNA used in this trial was isolated using both the STE CTAB method, and the DNeasy plant mini kit. PCR amplification was performed on two individuals from each of *N. australe* and *N. rosulatum* (Table 2.1) using a reaction volume of 20μL, 10x NEB buffer (New England Biolabs, Ipswich, Massachusetts, United States), 10μM of each dNTP, 20 μM forward primer, 20μM reverse primer, 100% DMSO, 100U NEB *Taq* (New England Biolabs), and 1μL DNA. PCR began with five minutes at 80°C, followed by 30 cycles of 95°C for one minute, 54°C for one minute, and 65°C for four minutes, then, a final step of 65°C for five minutes. All PCR products were assessed on a 1.5% agarose gel. ExoSAP-IT (New England Biolabs) was used to clean up the PCR products, 8.5μL of PCR product were added to 7.75μL ddH2O, 0.5μL of SAP, and 0.25μL of *Exo*-I, before heating to 37°C for 30 minutes, and then 80°C for 15 minutes. The sequencing reaction, using the BigDye™ Sequencing Ready Reaction mix (Thermo Fisher Scientific, Waltham, Massachusetts, United States), was of four minutes at 96°C, before 24 cycles of 96°C for 10s, 57°C for five seconds, and 60°C for three minutes. The samples were then submitted to the Massey Genome Service for clean-up and Sanger sequencing on an ABI3730 DNA Analyzer (Applied Biosystems). The output results were viewed, aligned, and trimmed in geneious v9.1.8. In order to screen for duplicate copies of primer sequences, a search of each primer sequence in a *N. australe* chloroplast genome (NC_049680.1) (O'LEARY *et al.* 2016) was performed in geneious v9.1.8

Table 2.3: Universal Chloroplast primer sequences.

Locus	Primer Name	Sequence (5'-3')
trnL-F	trnL$^{(UAA)}$_TabC	CGAAATCGGTAGACGCTACG
trnL-F	trnF$^{(GAA)}$_TabF	ATTTGAACTGGTGACACGAG
rpL16	R1516	CCCTTCATTCTTCCTCTATGTTG
rpL16	F17	GCTATGCTTAGTGTGTGACTCGTT G
ndhF	1b	TGGGACTTCTTCTTTTTCC
ndhF	R1318	CGAAACATATAAAATGCGGTTAATCC
rps16	5trnK uuuXI	TTAAAAGCCGAGTACTCTACC
rps16	x2F2	AAAGTGGGTTTTTATGATCC
psbA-trnH	psbA-F	GTTATGCATGAACGTAATGCTC
psbA-trnH	trnH$^{(GUG)}$	CGCGCATGGTGGATTCACAATCC

2.3.4 RAPDseq

Ten combinations of RAPD primer pairs (PILKINGTON 2019) were trialled using twelve individuals across *N. australe* and *N. rosulatum* (Table 2.1). The DNeasy plant mini kit (Qiagen) was used to extract DNA for this trial. A mixture of 5x HF fusion buffer (Thermo Fisher Scientific), 2.4µM each dNTP, 0.5µM forward primer, 0.5µM reverse primer, 3% DMSO, and 1x unit of Phusion *Taq* polymerase was added to 1µL DNA in a 10µL volume. Amplification was by 25 cycles of one-minute at 95°C, 90s at 50°C, ramped up at 0.2°C/sec to 72°C for 1-minute, followed by a hold at 72°C for 5 minutes. PCR products were visualised on a TBE-based 1.5% agarose gel stained with ethidium bromide. To check that different concentrations of DNA did not alter gel results, a serial dilution experiment was run for three primer pairs: F19 + R14, F15 + R10b, and F14 + R7 (see PILKINGTON (2019) for sequences) using a single *N. australe* DNA extraction. DNA

was diluted by ½, ¼, and 1/8, and two replicates of each dilution were used for PCR (same method as above).

Two primer pairs were chosen based on consistency and strength of bands across species and populations. The chosen primer pairs, F19 + R14 and F15 + R10b, were run against twelve *N. australe* and *N. rosulatum* DNA extractions (Table 2.1) using the above PCR protocol and conditions. Three microlitres of each PCR product was run on a 1.5% agarose gel and stained in ethidium bromide. The PCR products for each individual were then pooled and cleaned up using Agencourt AMPure XP beads at a 0.8x ratio to remove primer dimers and to size select against products <300bp in length.

The pooled and cleaned PCR products had individual sequence IDs and Illumina sequencing tags attached through a second round of PCR. Within a 25μL volume, 1x HF buffer (Thermo Fisher Scientific), 10μM each dNTP, 3% DMSO, 1x Phusion *Taq* (Thermo Fisher Scientific), and 0.5μM of each primer were mixed with 5μL template PCR product. The PCR programme was 95°C for two minutes, 18 cycles of 95°C for 10s, 70°C for one minute, and 72°C for two minutes, and finishing with 72°C for five minutes. The products were visualised on a 1.5% agarose gel, then cleaned using Agencourt AMPure beads before elution in 25μL of 10μM Tris-HCl (pH8). The PCR products were quantified using a Qubit 2.0 fluorometer (Invitrogen, Carlsbad, California, United States) and equal amounts of each were pooled into a volume of 24μL with a concentration of 6.02ng/μL.

The Massey genome service completed library preparation, sequenced the amplicons in one-quarter of a 250bp paired-end MiSeq run on the Illumina platform, and demultiplexed and quality checked the sequences. They did this by mapping the raw sequences against the PhiX genome in Bowtie2, so any PhiX sequences were removed. Adapters were removed through the fastq-mcf program from the ea-utils suite of tools

(http://code.google.com/p/ea-utils/; version 1.1.2-621). The sequences were then trimmed to their longest contiguous segment with a quality score above 0.01 using dynamictrim from SolexaQA++ (http://solexaqa.sourceforge.net/). SolexaQA++, fastQC, and fastQscreen were used for quality control, checking the raw sequences, processed sequences (before trimming), and trimmed sequences.

ipyrad v0.9.74 (EATON AND OVERCAST 2020) in Python v3.9.5 (Python software Foundation, 2001) was then used to filter and process the data. The *de novo* assembly method was used, and the data type defined as 'pairddrad'. It was decided to filter the raw reads to a maximum number of five 'N's (low quality bases), a phred Q score of 33, a minimum read length of 70, and the (stricter) adapter/primer score of 2. ipyrad initially groups similar sequences from within a sample into a cluster based on a similarity score (90% was used) using VSEARCH (ROGNES *et al.* 2016), and aligns reads within a cluster using MUSCLE (EDGAR 2004) to generate a consensus sequence, then estimates the heterozygosity and sequencing error rate per sample. A filter can be added to set the maximum number of alleles allowed per consensus, due to the unknown ploidy of *Notothlaspi*, this was set to 4. Consensus sequences with more than 5% 'N's were then removed, as were clusters with a sequencing depth of less than 6 or greater than 10,000. The within-individual consensus sequences were then aligned (also using MUSCLE) between all twelve samples. Before outputting results, a final filtering step defined the maximum number of indels per locus as 10, and the maximum percent heterozygous sites per locus as 20%. For the parameter 'minimum number of samples per locus' four different values (5, 6, 7, and 10) were trialled to find the correct balance between missing data and number of returned loci. Summary statistics were provided by the ipyrad output, and a Neighbour-Net created between samples in Splitstree5 (HUSON 1998).

2.4 Results

2.4.1 Microsatellite Markers

Initially, primers for 30 microsatellite markers were designed and trialled. Twelve polymorphic loci were selected due to their variability, simplicity of scoring, and successful amplification across *N. australe* and *N. rosulatum* in wide ranging populations (Table 2.2). *Notothlaspi viretum* individuals consistently produced three or more alleles and were therefore excluded from further analysis. The number of alleles per locus ranged from 7 (Na0011) to 18 (Na0298) (Table 2.4). Na8426 had the lowest observed heterozygosity at 0.26, while Na0043 had the highest value at 0.81. Missing data, excluding results removed from the dataset due to multiple alleles, were exclusively from *N. rosulatum* samples (Table 2.4). The frequency of missing data varied from zero (Na0298, Na0011, Na11999, Na1332, Na0043, Na6568, and Na8426) to six (Na20965) out of the 23 samples. In particular, the Q and G populations did not amplify specific markers: amplification of Q failed for Na003, Na0010, and partially for Na20965; while individuals from population G did not amplify Na27207, and Na20965.

Table 2.4: Microsatellite loci statistics. The sample size (n), number of alleles (Na), observed heterozygosity (H_O), and the missing data for each species at each locus.

Locus	N	Na	Ho	*N. australe* Missing data	*N. rosulatum* missing data
Na0003	20	16	0.500	0	3
Na0298	23	18	0.652	0	0
Na0011	23	7	0.348	0	0
Na27207	21	12	0.571	0	2
Na0078	21	8	0.333	0	2
Na0010	21	9	0.476	0	2
Na11999	23	8	0.565	0	0

Locus	N	Na	Ho	*N.australe* Missing data	*N. rosulatum* missing data
Na1332	22	13	0.636	0	0
Na20965	15	13	0.400	0	6
Na0043	21	16	0.810	0	0
Na6568	22	13	0.591	0	0
Na8426	23	10	0.261	0	0

2.4.2 Chloroplast Markers

PCR products from the ndhF locus did not present bands on an agarose gel, so Sanger sequencing was not attempted. The rps16 locus produced strong bands on a gel, however, sequencing results showed several internal homopolymers that rendered the sequence unusable. rpl16 produced moderately strong bands on a gel, but poor-quality sequencing results. The psbA-trnH showed strong bands on the gel, but had a large homopolymer in the centre of the amplified sequence. The psbA end of the homopolymer revealed 10 polymorphisms from 107bp, in which each individual had a unique haplotype. The trnL-F sequences had very strong bands on the gel, and in the TabF direction, produced 268bp of high-quality sequence with 10 polymorphisms among the four individuals. Like at the psbA locus, each individual had a unique haplotype. The TabC direction had a long homopolymer near the beginning of the sequence, so most of the sequence was unusable. Several of the polymorphisms were close to the homopolymer, so new primers could not be developed to amplify internal parts of the sequence, thus only results from the TabF direction can be used.

2.4.3 RAPDseq

Serial dilutions of template DNA concentrations used for PCR showed that at lower concentrations some fading in visual products occurred, particularly with the primer pair of F14 and R17 (Figure 2.2). However, fading was minimal and did not qualitatively affect banding results. This gives confidence in using DNA extractions of differing concentrations for the RAPDseq protocol. Figure 2.3 shows the PCR results at each stage of the process.

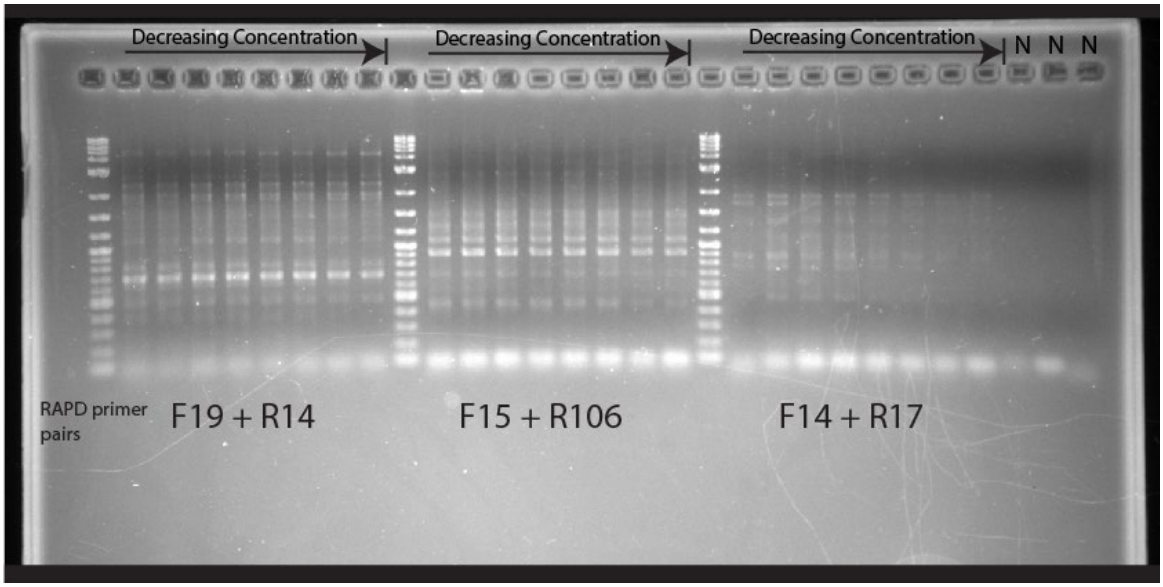
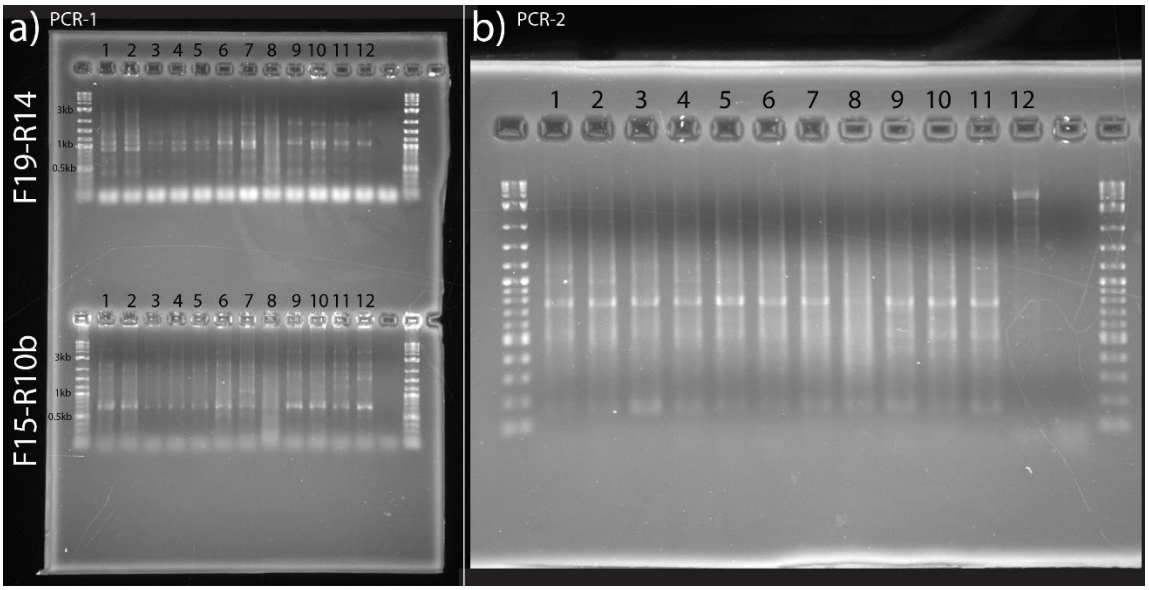


Figure 2.2: Serial dilution experiment. Serial dilutions using full, half, quarter, and eighth strength DNA for PCR were tested against three separate primer pairs with one replicate per PCR reaction. Pairs of replicates are put next to each other on the gel, with the full-strength DNA on the left, and the eighth-strength DNA to the right. The three negatives are in order on the far right. Results are shown on a 1.5% agarose gel with a 1kb+ ladder.



Bioinformatically, there is a trade-off between the number of loci returned by filtering, and the number of individuals that share these loci. We trialled varying values for 'minimum number of samples per locus', finding that more stringent measures (10 samples) returned 13 loci, while using fewer samples per locus (five samples) yielded more loci (76), as expected. Here, the data was filtered by a minimum six samples per locus, which returned 41 variable loci.

Table 2.5 shows the summary statistics for the results of data filtering in ipyrad. The number of raw reads ranged from 100,635 to 318,370, and these were filtered down to between 65,296 and 156,474 reads. From these results, the estimated percent heterozygosity ranged from 0.46% to 0.8%, and the error rates between 0.0004 and 0.000787. None of the samples had all 41 returned loci. B5-10 had the highest number of loci, 39, while H2-12 had the lowest, 24 loci. H2-12 also had the lowest number of raw reads, at 100635.

The Neighbor-Net in Figure 2.4 shows the inferred relationships between individuals based on the RAPDSeq data. There are contrasting patterns in the data, however, individuals from each species are segregated on either side of a split. On both sides, individuals from the same site are not grouped together, except H2 samples.

Table 2.5: Summary statistics from the ipyrad pipeline.

Samples	Number of Raw Reads	Number of Filtered Reads	Estimated % Heterozygosity	Estimate of error rate	Number of loci in results
B18-03	222920	149433	0.79	0.00068	34
B18-04	218346	143300	0.56	0.000542	26
B15-06	188077	96924	0.74	0.00049	32
B5-10	300011	102840	0.62	0.000787	39
E3-01	218072	143606	0.8	0.000752	32
H2-12	100635	65296	0.65	0.00044	24
H2-14	318370	153994	0.5	0.0004	30
I8-05	204086	109062	0.61	0.000568	25
Lr1-11	574123	375547	0.46	0.000436	26
Lr1-20	252830	136428	0.57	0.000429	29
P3-01	198839	115210	0.67	0.000543	30
P3-09	299137	156474	0.62	0.000609	31

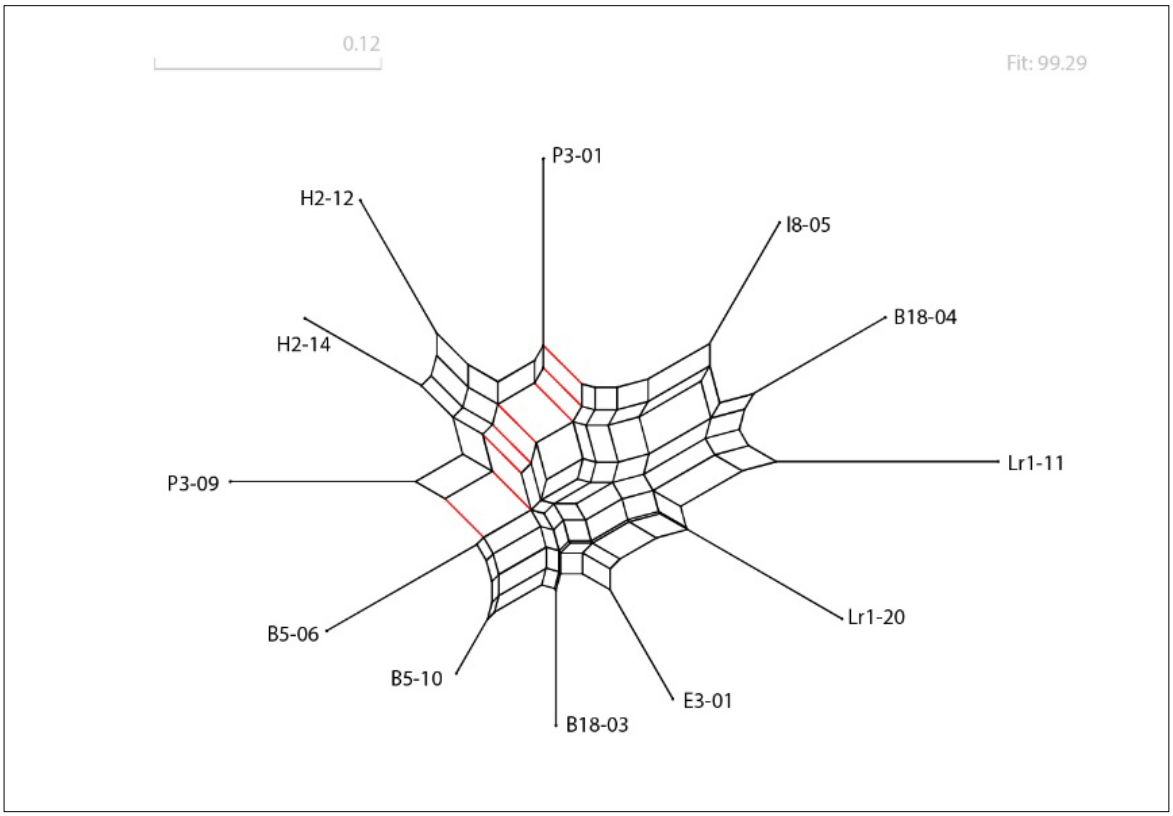


Figure 2.4: Neighbour-Net of RAPDseq results. The split highlighted in red differentiates samples from *N. australe* vs *N. rosulatum*.

2.5 Discussion

In order to undertake population genetic investigations, suitable molecular markers must be selected and trialled. Here, we explore three different marker types for suitability in population genetic studies of the *Notothlaspi* genus.

Microsatellite loci are useful for population genetic studies as they tend to have high mutation rates and can often differentiate among individuals. Twelve polymorphic microsatellite markers were developed here that are suitable for intra-specific and inter-specific analysis of *N. australe* and *N. rosulatum.* Unfortunately, most markers revealed three or more alleles from *N. viretum* individuals, so they are of limited use for comparing with the other two species without using a presence/absence scoring approach. This is

because heterozygosity estimates cannot be made without dosage information in polyploid loci (DUFRESNE *et al.* 2014). Some markers also had higher than expected missing data within *N. rosulatum.* The primers were designed from *N. australe* DNA, so an increase of missing data in specific populations of *N. rosulatum* could be ascertainment bias resulting in null alleles, which are not uncommon in cross-species amplification (TORRES-DÍAZ *et al.* 2021) due to, e.g., mutations in the priming regions of the marker (CALLEN *et al.* 1993). Failed PCR can also be a result of other issues, including slippage during PCR amplification, short allele dominance or poor-quality DNA (TABERLET *et al.* 1996; WATTIER *et al.* 2002). Missing data has several implications for population genetic analysis, particularly leading to lower estimates of heterozygosity, as a genotype with a 'missing' allele would be identified as a homozygote (CALLEN *et al.* 1993). Allelic diversity also has been reported to decrease with cross-species amplification (SELKOE AND TOONEN 2006), as a locus may be more conserved if it remains following divergence.

The trnL-F and psbA-trnH loci each had ten polymorphisms between populations and species, however, both loci could only be read in a single direction due to homopolymers within each sequence. While they may not be verified by cross-directional sequencing, all sequences are of good quality and the four individuals, two from *N. australe* and two from *N. rosulatum,* had unique chloroplast haplotypes. The variability of these markers, particularly between populations of the same species, makes them suitable for intra- and inter-specific studies of the *Notothlaspi* genus. It is useful to use markers from different genomes to detect processes like introgression (TURNER *et al.* 2005). Processes such as chloroplast capture lead to incongruent nuclear and plastid gene trees, which typically would not be detected without the use of multiple markers (OKUYAMA *et al.* 2005; KAWABE *et al.* 2018). It is therefore valuable to have suitable markers from the chloroplast genome for the analysis of *Notothlaspi*.

The RAPDseq trial found several informative loci that may be suitable for further population genetic analysis, however, due to the timing of data collection, this approach could not be tested beyond the trial stage. One challenge with this analysis lies with the trade-off between missing data and the number of loci returned. If fewer samples per locus are allowed, there will be an increase in missing data, however, a greater number of individuals with representative sequences at each locus returns fewer results (VALENCIA *et al.* 2018; WAGNER *et al.* 2020). Another way to increase the number of retained loci is to decrease the minimum depth per locus, however this can lead to underestimated estimates of heterozygosity (EATON AND ISAAC 2019). With more individuals, the power to discriminate will increase, and individuals with poor representation can be filtered out. Therefore, the next steps are to test against numerous individuals and determine how much variation can be identified within and between populations.

PILKINGTON (2019) used ten random RAPDseq primer pairs to generate loci for linkage map development in *Sophora*, however we used just two primer pairs for our trial. More loci would generate a greater variety of loci and polymorphisms, however more space within an Illumina MiSeq sequencing run would be required to return the same level of depth. Using two loci generated ample results to differentiate between individuals from the same site, so, when scaled up, we would expect sufficient variation for a population genetic study on *Notothlaspi.*

While RAPDseq has not been used, to our knowledge, for population genetics, it shows great promise as a relatively inexpensive and effective method for finding and genotyping SNPs within and between populations and species (MCLAY 2018; PILKINGTON 2019).

2.6 Conclusion

All three marker systems have potential individually for the *Notothlaspi* system however, using them in conjunction would mitigate individual issues and allow for a robust population genetic study of *Notothlaspi*. One challenge lies with analysing *N. viretum. Notothlaspi viretum* individuals consistently produced more microsatellite alleles than *N. australe* and *N. rosulatum* individuals, which may be indicative of a higher ploidy level. Chloroplast markers will be suitable in this case, but the nuclear loci may prove an issue for downstream analysis. Multiplex amplicon sequencing methods, such as RAPDseq, are more suited to nuclear analysis of polyploids than microsatellites, although there can be bias introduced from differences in annealing of different parental sequences, or from differences in PCR amplification (DUFRESNE *et al.* 2014). Each of these marker systems are applicable to the study of the population genetics of the *Notothlaspi* system, particularly for *N. australe* and *N. rosulatum,* and therefore the objective to develop intra-specific and intra-generic markers suitable for *Notothlaspi* has been accomplished.

2.7 References

Afgan, E., D. Baker, B. Batut, M. Van Den Beek, D. Bouvier *et al.*, 2018 The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research 46: W537-W544.

Andrews, S., 2010 FastQC: A quality control tool for high throughut sequence data.

Aris-Brosou, S., and L. Excoffier, 1996 The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Molecular Biology and Evolution 13: 494-504.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin *et al.*, 2012 SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19: 455-477.

Bennett, M. S., S. H. Shiu and R. E. Triemer, 2017 A rare case of plastid protein-coding gene duplication in the chloroplast genome of *Euglena archaeoplastidiata* (Euglenophyta). Journal of Phycology 53: 493-502.

Brownstein, M. J., J. D. Carpten and J. R. Smith, 1996 Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. Biotechniques 20: 1004-1010.

Buffalo, V., and G. Coop, 2020 Estimating the genome-wide contribution of selection to temporal allele frequency change. Proceedings of the National Academy of Sciences 117: 20672-20680.

Callen, D. F., A. D. Thompson, Y. Shen, H. A. Phillips, R. I. Richards *et al.*, 1993 Incidence and origin of" null" alleles in the (AC) n microsatellite markers. American Journal of Human Genetics 52: 922.

Dawson, M. I., and E. J. Beuzenberg, 2000 Contributions to a chromosome atlas of the New Zealand flora—36. Miscellaneous families. New Zealand Journal of Botany 38: 1-23.

de Lange, P. J., Rolfe, J.R., Barkla, J.W., Courtney, S.P., Champion, P.D., Perrie, L.R., Beadel, S.M., Ford, K.A., Breitwieser, I., Schonberger,, and H.-W. I., R., Heenan, P.B. & Ladley, K, 2018 Conservation status of New Zealand indigenous vascular plants, 2017, pp. 82 in *New Zealand Threat Classification Series 22*. Department of Conservation, Wellington.

Dufresne, F., M. Stift, R. Vergilino and B. K. Mable, 2014 Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. Molecular Ecology 23: 40-69.

Eaton, D., and O. Isaac, 2019 Ipyrad. Assembly: Parameters, pp. Read the Docs.

Eaton, D. A., and I. Overcast, 2020 ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics 36: 2592-2594.

Edgar, R., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797.

Grover, A., and P. C. Sharma, 2016 Development and use of molecular markers: past and present. Critical Reviews in Biotechnology 36: 290-302.

Guichoux, E., L. Lagache, S. Wagner, P. Chaumeil, P. Léger *et al.*, 2011 Current trends in microsatellite genotyping. Molecular Ecology Resources 11: 591-611.

Heenan, P., 2019 A taxonomic revision of *Notothlaspi* (Brassicaceae), a specialist alpine genus from New Zealand. Phytotaxa 399: 248-260.

Heenan, P. B., and M. S. McGlone, 2013 Evolution of New Zealand alpine and open-habitat plant species during the late Cenozoic. New Zealand Journal of Ecology: 105-113.

Heenan, P. B., T. R. Millar, R. D. Smissen, M. S. McGlone and A. D. Wilton, 2017 Phylogenetic measures of neo-and palaeo-endemism in the indigenous vascular flora of the New Zealand archipelago. Australian Systematic Botany 30: 124-133.

Huson, D. H., 1998 SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics (Oxford, England) 14: 68-73.

Jordan, W. C., M. W. Courtney and J. E. Neigel, 1996 Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (*Lemnaceae*). American Journal of Botany 83: 430-439.

Kalia, R. K., M. K. Rai, S. Kalia, R. Singh and A. K. Dhawan, 2011 Microsatellite markers: an overview of the recent progress in plants. Euphytica 177: 309-334.

Kawabe, A., H. Nukii and H. Y. Furihata, 2018 Exploring the history of chloroplast capture in *Arabis* using whole chloroplast genome sequencing. International Journal of Molecular Sciences 19: 602.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung *et al.*, 2012 Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649.

Kirk, H., and J. R. Freeland, 2011 Applications and implications of neutral versus non-neutral markers in molecular ecology. International Journal of Molecular Sciences 12: 3966-3988.

Levinson, G., and G. A. Gutman, 1987 Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Molecular Biology and Evolution 4: 203-221.

Li, W.-W., L.-Q. Liu, Q.-P. Zhang, W.-Q. Zhou, G.-Q. Fan *et al.*, 2021 Phylogeography of *Prunus armeniaca* L. revealed by chloroplast DNA and nuclear ribosomal sequences. Scientific Reports 11: 13623.

McLay, T., 2018 Hybridisation between species of *Xanthorrhoea* in Yuraygir National Park revealed using a newly developed next-generation sequencing method. Australian Systematic Botany Society Newsletter: 13-16.

Millar, T. R., P. B. Heenan, A. D. Wilton, R. D. Smissen and I. Breitwieser, 2017 Spatial distribution of species, genus and phylogenetic endemism in the vascular flora of New Zealand, and implications for conservation. Australian Systematic Botany 30: 134-147.

Morgan-Richards, M., R. D. Smissen, L. D. Shepherd, G. P. Wallis, J. J. Hayward *et al.*, 2009 A review of genetic analyses of hybridisation in New Zealand. Journal of the Royal Society of New Zealand 39: 15-34.

National Center for Biotechnology Information, 1988 Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information, pp.

Nieves-Orduña, H. E., M. Müller, K. V. Krutovsky and O. Gailing, 2021 Geographic Patterns of Genetic Variation among Cacao (*Theobroma cacao L.*) Populations Based on Chloroplast Markers. Diversity 13: 249.

O'Leary, N. A., M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad *et al.*, 2016 Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research 44: D733-D745.

Okuyama, Y., N. Fujii, M. Wakabayashi, A. Kawakita, M. Ito *et al.*, 2005 Nonuniform Concerted Evolution and Chloroplast Capture: Heterogeneity of Observed Introgression Patterns in Three Molecular Data Partition Phylogenies of *Asian Mitella* (Saxifragaceae). Molecular Biology and Evolution 22: 285-296.

Olmstead, R. G., and J. A. Sweere, 1994 Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. Systematic Biology 43: 467-481.

Patwardhan, A., S. Ray and A. Roy, 2014 Molecular markers in phylogenetic studies-a review. Journal of Phylogenetics and Evolutionary Biology 2014.

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

Peakall, R., and P. E. Smouse, 2006 GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288-295.

Pilkington, K. M., 2019 The genetic architecture of the divaricate growth form: A QTL mapping approach in *Sophora* (Fabaceae), PhD in *School of Fundamental Sciences*. Massey University, Massey University, Manawatu, New Zealand.

Pirie, M. D., M. P. B. Vargas, M. Botermans, F. T. Bakker and L. W. Chatrou, 2007 Ancient paralogy in the cpDNA trnL-F region in Annonaceae: implications for plant molecular systematics. American Journal of Botany 94: 1003-1016.

Rodzen, J. A., and B. May, 2002 Inheritance of microsatellite loci in the white sturgeon (*Acipenser transmontanus*). Genome 45: 1064-1076.

Rognes, T., T. Flouri, B. Nichols, C. Quince and F. Mahé, 2016 VSEARCH: a versatile open source tool for metagenomics. PeerJ 4: e2584.

Rozen, S., and H. J. Skaletsky, 2000 Primer3, pp. Humana Press, Totowa, NJ.

Selkoe, K. A., and R. J. Toonen, 2006 Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters 9: 615-629.

Shaw, J., E. B. Lickey, E. E. Schilling and R. L. Small, 2007 Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. American Journal of Botany 94: 275-288.

Shepherd, L. D., and T. McLay, 2011 Two micro-scale protocols for the isolation of DNA from polysaccharide-rich plant tissue. Journal of Plant Research 124: 311-314.

Smissen, R., and I. Breitwieser, 2008 Species relationships and genetic variation in the New Zealand endemic *Leucogenes* (Asteraceae: Gnaphalieae). New Zealand Journal of Botany 46: 65-76.

Symonds, V. V., and A. M. Lloyd, 2004 A simple and inexpensive method for producing fluorescently labelled size standard. Molecular Ecology Notes 4: 768-771.

Taberlet, P., L. Gielly, G. Pautou and J. Bouvet, 1991 Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17: 1105-1109.

Taberlet, P., S. Griffin, B. Goossens, S. Questiau, V. Manceau *et al.*, 1996 Reliable Genotyping of Samples with Very Low DNA Quantities Using PCR. Nucleic Acids Research 24: 3189-3194.

Torres-Díaz, C., M. A. Valladares, M. A. Molina-Montenegro and F. Mora-Poblete, 2021 Isolation and characterization of microsatellites for the endangered endemic tree *Nothofagus alessandrii* (Nothofagaceae). Molecular Biology Reports.

Turner, T. L., M. W. Hahn and S. V. Nuzhdin, 2005 Genomic islands of speciation in *Anopheles gambia*. PLoS Biology 3: e285.

Valencia, L. M., A. Martins, E. M. Ortiz and A. Di Fiore, 2018 A RAD-sequencing approach to genome-wide marker discovery, genotyping, and phylogenetic inference in a diverse radiation of primates. PloS One 13: e0201254.

Wagner, F., T. Ott, M. Schall, U. Lautenschlager, R. Vogt *et al.*, 2020 Taming the Red Bastards: Hybridisation and species delimitation in the *Rhodanthemum arundanum*-group (Compositae, Anthemideae). Molecular Phylogenetics and Evolution 144: 106702.

Wagstaff, S. J., B. P. Molloy and J. A. Tate, 2010 Evolutionary significance of long-distance dispersal and hybridisation in the New Zealand endemic genus *Hoheria* (Malvaceae). Australian Systematic Botany 23: 112-130.

Walden, N., D. A. German, E. M. Wolf, M. Kiefer, P. Rigault *et al.*, 2020 Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. Nature Communications 11: 1-12.

Wattier, R., C. Engel, P. Saumitou-Laprade and M. Valero, 2002 Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). Molecular Ecology 7: 1569-1573.

Winkworth, R. C., S. J. Wagstaff, D. Glenny and P. J. Lockhart, 2005 Evolution of the New Zealand mountain flora: origins, diversification and dispersal. Organisms Diversity & Evolution 5: 237-247.

Wolfe, K. H., W.-H. Li and P. M. Sharp, 1987 Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences 84: 9054-9058.

Xu, J., F. Zang, Q. Wu, Y. Wang, B. Wang *et al.*, 2021 Analysis of the genetic diversity and molecular phylogeography of the endangered wild rose (*Rosa rugosa*) in China based on chloroplast genes. Global Ecology and Conservation 28: e01653.

Chapter 3

Using Population Genetics to investigate the Spatial Structure of Genetic Variation in *Notothlaspi*

3.1 Abstract

Heterogeneous environments present excellent opportunities to study lineage divergence and secondary contact. The northern South Island of New Zealand has highly complex geology, and mountainous regions exposed to recent uplift and glaciation. To investigate biogeographic patterns across this region, the three species in the genus *Notothlaspi* (Brassicaceae) were selected due to the unusual ability of *N. australe* to grow on several edaphically challenging soil types. Using a population genetics approach, analyses of the chloroplast trnL-F locus and twelve nuclear microsatellite markers were used to investigate the distribution of genetic variation across the alpine landscape of the northern South Island. These analyses uncovered potential chloroplast-capture and nuclear admixture in the Richmond Forest Park, where all three species occur. Nuclear diversity was higher in populations within the Richmond Forest Park than those of the same species outside of it. Analysis of *N. australe* populations in the Kahurangi National Park revealed an unexplained biogeographic split that departs from models of isolation by distance. To the east, most *N. rosulatum* sites in Marlborough had unique chloroplast haplotypes, but sites were nearly indistinguishable based on nuclear data, suggesting greater pollen-mediated gene flow than seed dispersal. This study provides a foundation for future biogeographic studies in northern South Island alpine plants.

3.2 Introduction

The genotypes of plant species, populations, and even individuals are not uniform, displaying different genetic structures through space and time (LOVELESS AND HAMRICK 1984). The structuring of genetic variation is largely determined by geneflow, which governs the degree to which genetic changes in populations remain independent (SLATKIN 1985). These genetic changes are allele frequency shifts that occur through genetic drift (KIMURA 1983), genetic draft (GILLESPIE 2000), and as responses to selection driven by environmental stimuli (SAMBATTI AND RICE 2006). Most plants have limited mobility, so factors impacting the dispersal of seed and pollen determine the degree of gene flow between populations (LOVELESS AND HAMRICK 1984). Greater physical distance between individuals generally reduces the opportunity for interbreeding, thus isolation by distance (IBD) is often an important concept when discussing divergence (WRIGHT 1943). Major environmental events can cause widespread shifts to population distributions, which can lead to divergence between isolated populations. For example, during the last glacial period (115,000 – 11,7000ya), hot and cold cycles drove plant populations to change or move to survive, isolating and re-introducing contact between populations of the same species (SHEPHERD *et al.* 2007; MAGUILLA *et al.* 2017; REN *et al.* 2017; WANG *et al.* 2017; NEVADO *et al.* 2018; TOMASELLO *et al.* 2020). In contrast to IBD, isolation by environment (IBE), a term coined by (WANG AND BRADBURD 2014) based on a history of observations/principles (HALDANE 1954; KAWECKI 1995; KIRKPATRICK *et al.* 1997)posits that specific allele frequencies suitable for local conditions may be diluted by outcrossing with individuals adapted to different conditions. Therefore, local adaptation can lead to selection against migrants or foreign gametes, which allows for differentiation at small spatial scales (AGUIRRE-LIGUORI *et al.* 2019). IBE may therefore be particularly important in highly heterogeneous landscapes,

where complex patterns of genetic variation can be distributed among different micro-habitats (SORK *et al.* 2010; GAO *et al.* 2012; TEMUNOVIĆ *et al.* 2012; MCKOWN *et al.* 2014; CHEN *et al.* 2017). Such heterogeneous environments are important sites in which to investigate the interplay between the environment and the spatial structure of genetic variation, particularly where there are opportunities for secondary contact and hybridisation (SEEHAUSEN *et al.* 2008; ZHANG *et al.* 2020).

Hybridisation can have long lasting effects on the patterns of genetic structure within a population or species. The re-introduction of geneflow between diverged lineages often occurs during range shifts, when populations come into secondary contact (CHUNCO 2014; TAYLOR *et al.* 2015). Introgression following secondary contact allows the transfer of adaptive alleles between populations, increasing the evolutionary potential of the recipient population (WHITNEY *et al.* 2006; KIM *et al.* 2008; WHITNEY *et al.* 2010; SUAREZ-GONZALEZ *et al.* 2018). Gene transfer does not occur evenly across the genome, instead, certain 'divergence' genes and their spheres of linkage remain distinct to each lineage, while alleles outside of these may be transferred (TURNER *et al.* 2005; MARTIN AND JIGGINS 2017). The transfer of plastids between plant groups is not uncommon and is sometimes the only apparent evidence of gene flow between groups (RIESEBERG AND SOLTIS 1991; TSITRONE *et al.* 2003). Due to the frequency of chloroplast capture in natural populations, the distribution of chloroplast haplotypes is often more representative of geography than lineage (THÓRSSON *et al.* 2001; ACOSTA AND PREMOLI 2010; SCHUSTER *et al.* 2018; LIU *et al.* 2020).

Due to its geological and biological history, New Zealand is an excellent place to investigate the influence of environmental factors on the distribution of genetic diversity. Most plant species arrived recently to the archipelago of New Zealand, as shown by the relatively few endemic genera and families, but high species endemism and diverse

morphologies (MILLAR *et al.* 2017; HEENAN AND MCGLONE 2019). The axial mountain ranges of the Southern Alps are young; tectonism in the Pliocene initiated uplift 2 MYA, providing a range of open niches (HEENAN AND MCGLONE 2013) linked to several evolutionary radiations in alpine plants (WAGSTAFF AND GARNOCK-JONES 1998; LOCKHART *et al.* 2001; WAGSTAFF AND BREITWIESER 2004; JOLY *et al.* 2009a; MEUDT *et al.* 2009). Variable underlying substrate has also been linked to increased plant diversification, as a key driver of local adaptation (ALVAREZ *et al.* 2009; AUGUSTO *et al.* 2017; RAHBEK *et al.* 2019; HULSHOF AND SPASOJEVIC 2020). Soil parent material has impacts on the chemical, physical, and biotic soil conditions, each of which are important characteristics for plant growth (JENNY 1941). Mountain soils in the northern and southern ends of the South Island of New Zealand are derived from narrow bands of variable underlying rock types, such as granitic, ultramafic, schistose, volcanic, silico-clastic, and calcareous rocks (GNS Science 2016). Therefore, these regions are heterogeneous edaphic environments for alpine plants. Some of these rock types (particularly ultramafic, calcareous, and some granitic rocks) are found to host specialised edaphic flora, diverged from relatives on other soil types, indicating geology has had a significant impact on plant diversification in New Zealand (LEE *et al.* 1983; GLENNY 2004; RICHARDSON *et al.* 2012; ROGERS *et al.* 2018).

Along with mountain uplift and geology, climate has had a large effect on the distribution of genetic variation and species ranges in New Zealand. Quaternary climate fluctuations induced a series of cold and hot periods in New Zealand (BEU AND EDWARDS 1984; SUGGATE 1990; MARKGRAF *et al.* 1992). During cold periods, large glacial systems covered much of the Southern Alps, sea levels dropped, and much of terrestrial New Zealand was covered in extensive grasslands and bare ground. Warmer periods, some hotter than the present day, brought the slow return of forests to the wider landscape,

and the surviving alpine plants shifted to higher elevations (RYAN *et al.* 2012; HEENAN AND MCGLONE 2013; WILLIAMS *et al.* 2015). These glacial cycles, particularly events since the last glacial maximum, are thought to underlie many contemporary patterns of floral distribution and genetic structure found today. A classic example of these patterns is the New Zealand Beech genera, *Fuscospora* and *Lophozonia* (Nothofagaceae) (RAWLENCE *et al.* 2021). These are dominant genera in remaining New Zealand forests, but are absent from the 'Manawatu beech gap' and 'central South Island beech gap'. The central South Island gap has been attributed to the severe glacial environment in the area; a region which also separates several other closely related taxa (WARDLE 1963; MCCULLOCH *et al.* 2010; WESTON AND ROBERTSON 2015; RAWLENCE *et al.* 2021).

Patterns of variation in relation to environmental factors are most commonly studied at the species level in New Zealand; few studies investigate genetic structure at the intraspecific level (but see (YOUNG *et al.* 2001; HOLZAPFEL *et al.* 2002; ARMSTRONG AND DE LANGE 2005; SMISSEN *et al.* 2006; SHEPHERD *et al.* 2007; SHEPHERD AND PERRIE 2011; BEEVER *et al.* 2013; PILKINGTON 2014)). To our knowledge, no populations genetics studies have included New Zealand alpine plant species. Studying genetic structure at the population level, rather than the species level, provides a means of investigating phenomena such as divergence, adaptation, and hybridisation as they occur, rather than in retrospect (VIA 2009). To better understand the impacts of mountain building and glacial cycles in heterogeneous geological environments, more intra-specific studies are required.

To investigate the interplay between genetic divergence and secondary contact in the highly heterogeneous geological environment of the northern South Island of New Zealand, the genus *Notothlaspi* (Brassicaceae) was selected. *Notothlaspi australe*, *N. rosulatum,* and *N. viretum* are the only species within the New Zealand endemic tribe

Notothlaspideae (MITCHELL AND HEENAN 2000; HEENAN 2019). This tribe is grouped in lineage two of the recent Brassicaceae phylogeny (WALDEN *et al.* 2020). *Notothlaspi* is an attractive system as *N. australe* populations are found on several challenging soil types, including soils derived from marble, greywacke, granite, volcanics, schist, and interestingly, ultramafic soils. Ultramafic soils are toxic to many plants due to high levels of heavy metals (RAHBEK *et al.* 2019). As there is opportunity for hybridization between the three species, investigations into genetic diversity in *N. australe* need to consider close relatives as well. Therefore, a study of genetic variation and structure was initiated to include all species of *Notothlaspi.*

All three species of *Notothlaspi* occupy rocky habitats in the South Island of New Zealand, with overlapping distributions in the Richmond Forest Park (Figure 3.1). *Notothlaspi australe* grows in the north west of the South Island on fellfields, rocky outcrops, bluffs, and shingly sites, often in shady areas and on south facing slopes (HEENAN 2019) (Figure 3.1). This species is perennial, with a cushion-like habit and multiple loose rosettes. *Notothlaspi rosulatum* mainly grows to the east of the alpine fault, although it is also found in the Richmond Forest Park (Figure 3.1), where there is range overlap with *N. australe* and *N. viretum*. *Notothlaspi rosulatum* is a monocarpic, biennial scree slope specialist, with cryptic morphology and a very patchy distribution (HEENAN 2019). Typically, it occupies unstable greywacke–argillite scree slopes; however, it has been found on other silico-clastic derived soils such as other sandstones, mudstones, and schist. *Notothlaspi viretum* was recently described by HEENAN (2019), and is restricted to near the summit of Maungakura/Red Hill in Marlborough, sympatric with *N. australe*. Under the New Zealand threat classification system, it is 'Threatened, Nationally Critical'. Mechanisms of pollen and seed dispersal within the genus are unknown, but *N. australe* and *N. rosulatum* both have small, white, fragrant flowers and

the strong sweet smell in these flowers may attract pollinators (WRIGHT AND SCHIESTL 2009). Cytological variation has been observed in the genus (DAWSON AND BEUZENBERG 2000). Chromosome counts made of *N. australe* and *N. viretum* plants, taken from the Marlborough Red Hills, both had chromosome counts of 2n = 90–100 (DAWSON AND BEUZENBERG 2000). In contrast, a single *N. rosulatum* individual, sampled from near Lake Tennyson, had substantially fewer chromosomes (2n = 36–38). The large difference in chromosome number implies a polyploidisation event in the history of the genus which may pose challenges for any nuclear genetic analyses (DUFRESNE *et al.* 2014).

This study investigates the geographic distribution of genetic variation in *Notothlaspi*, and the interplay between genetic divergence and secondary contact. Nuclear microsatellite markers and DNA sequence data of the chloroplast trnL-F locus are utilised to achieve the following two objectives: determine the distribution of genetic variation across the alpine landscape within and among *Notothlaspi* species; and infer historical influences on the adaptive and evolutionary history of *Notothlaspi.*

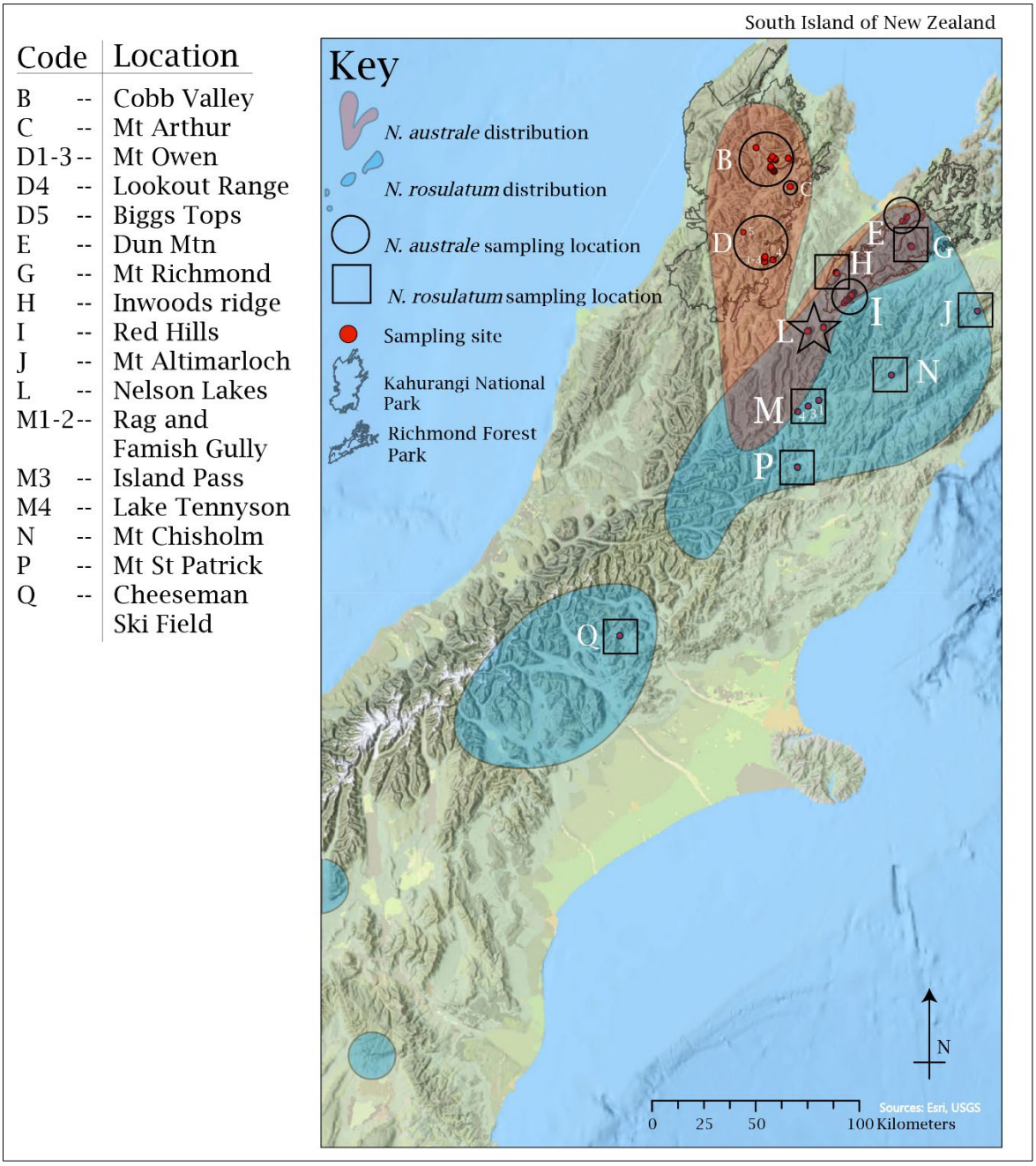


Figure 3.1 Distribution of *Notothlaspi* and the sampling sites included in the present study.

3.3 Methods

3.3.1 Sample Collection and DNA Extraction

Between 2019 and 2021, 567 *N. australe* and 256 *N. rosulatum* individuals were sampled from 64 sites and 18 locations (Figure 3.1) under permission from Nelson City Council and the Department of Conservation (Permit 82446-RES). To expand sampling for the genus, leaf material from 13 *N. viretum* herbarium specimens was provided by Manaaki Whenua's Allan Herbarium (CHR). For clarity, broader sampling areas, such as a valley or mountain, are described as 'locations', while specific sampling plots within locations are referred to as 'sites'. Each site was treated in subsequent analyses as an ecological population. At time of collection individuals were assigned to a species based on morphological features and geographic area. Due to intermediate morphologies and geographic area, plants from location L could not be assigned to *N. australe* or *N. rosulatum* with confidence.

Individual IDs were designed to convey location and site information, for example B5-03 is from location B (the Cobb Valley), site five, and was the third individual sampled. Commonly, fewer than 30 individuals were found within each site, so all sighted individuals were sampled except in location I. Due to the large number of individuals at I, samples were taken along a transect, with 10-20 individuals sampled from each of 12 sites on the transect. For later analyses, sites from different ridgelines in location B (Cobb Valley) were grouped together; B(n) and B(s) are combined sites from the northern and southern valley ridges respectively, and B(w) from a western site. From each individual, a scaled photograph and one to five leaves were taken. Leaves were immediately stored in labelled filter paper in silica gel for desiccation. Total genomic DNA was extracted

following the STE CTAB method (SHEPHERD AND MCLAY 2011), the DNeasy Plant Mini Kit (Qiagen), or the DNeasy Power Plant Pro kit (Qiagen), as described in Chapter 2.

At each site a geographic coordinate was recorded using a Juno T41 (Trimble, Sunnyvale, California, United States) GPS (Supplementary Table 1), along with notes of associated flora, a geological description, and a soil description where practicable. Soil and geological descriptions were made for most *N. australe* sites, as well as for *N. rosulatum* sites from the Richmond Forest Park. For each site where a soil description was made, this description was taken from a hole, dug as deep as possible, within one metre of a *Notothlaspi* plant. These were made for most *N. australe* sites, and for *N. rosulatum* sites in the Richmond Forest Park. Soil physical characteristics were described. These characteristics included: parent material; root abundance and depth; horizons; structure; consistence; colour (using the Munsell colour chart); texture; and percentage of stones in the profile. At sites with more than 30 individuals, a herbarium specimen was collected and deposited in the Dame Ella Campbell herbarium.

3.3.2 Sequencing Chloroplast Loci

The trnL-F locus previously revealed intraspecific variation in *Notothlaspi* (Chapter 2), but the level of within-site variation was untested. To test for within-site polymorphisms, equal amounts of DNA from each individual of a given site were pooled, prior to PCR (Polymerase Chain Reaction) and Sanger sequencing. Each pooled sample combined DNA from two to 20 individuals (Table 3.1). Typically, pooling of DNA was undertaken for individuals occurring at the same site, however, there were some exceptions (Table 3.1). To verify pooled results, individual sequencing was undertaken for most sampled sites (Table 3.1). Overall, sequencing was carried out for 51 separate

individuals, and 21 pools, from 12 locations (Table 3.1). PCR cocktail, PCR protocol, reaction clean up, and Sanger sequencing were all completed as described in chapter 2.

Table 3.1: Number of individuals in pools and individually sequenced per site for chloroplast sequencing.

Species	location	Sites	n (pooled)	n (individually sequenced)
N. australe	B(n)	B2-B6	10	2
	B(s)	B8-B16	11	2
	B(w)	B17	10	1
	D	D1- D3	12	
		D4	5	
		D5	5	
	E	E1		4
		E2	10	2
		E1 and E3	10	2
	I	I1-I4	10	2
		I5-I12	9	2
Unknown	L	L1		1
		L2		1
		L3		1
		L4		1
		L5		1
		L6		2
N. rosulatum	G	G1	4	1
		G2-4	6	3
	H	H1	10	2
		H2	20	2
	J	J1	2	2
	M	M1-M2	10	2
		M3-M4	10	2
	N	N1	4	4
	P	P1-P4	10	
	Q	Q1-1 - Q2-08	10	2
		Q2-09 - Q2-19	11	
N. viretum	Allan Herbarium			13

All PCR products were Sanger sequenced at the Massey Genome Service, using the Sequencing Capillary Separation Service on an ABI3730 DNA Analyzer. Results were analysed in Geneious v9 (KEARSE *et al.* 2012), as described in chapter 2, where ambiguous sites were removed. A deletion of 10 base pairs was found in all four individuals from

location N, these were treated as a single polymorphism, as the nature of the deletions is unknown. A median joining haplotype network was created by exporting a nexus file into popart v1.7 (LEIGH AND BRYANT 2015). For phylogenetic analysis, all usable pooled and individual sequences were aligned in Geneious, along with an outgroup trnL-F sequence; *Chamira circaeoides* (Brassicaceae) was selected as the outgroup based on the recent Brassicaceae phylogeny by (WALDEN *et al.* 2020). The aligned sequences were exported in MEGA format into MEGA-X v10.2.4 (KUMAR *et al.* 2018). Based on BIC (Bayesian information criterion) score, the Tamu ra-3-parameter nucleotide substitution model (TAMURA 1992) was used to construct a maximum likelihood (ML) tree. Support for the ML tree was computed by carrying out 1000 bootstrap replicates on the multiple sequence alignment and subsequently mapping the bootstrap values onto the tree. For formatting, the resulting tree was exported into FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) as a Newick file.

To determine if within-site polymorphisms could be detected using the pooled approach, the chromatograms of pooled genetically distinct individuals were examined. DNA from individuals B17-01 and M1-03 were pooled 1:1, 1:4, 1:10, and 1:20. The trnL-F locus was then amplified using the same PCR protocol and conditions from Chapter 2. Sequence outputs were scrutinised in GENEIOUS v9 to assess if the between-species polymorphisms were visible in the pooled results.

3.3.3 Microsatellite loci:

3.3.3.1 Sampling

From the 823 collected samples, 265 individuals from 16 sites were selected for microsatellite analysis (Supplementary table 1), using 20 individuals per site where possible. These sites were selected to maximise the geographic range for each species, they include nine *N. australe* sites, one L site, and six *N. rosulatum* sites. *N. viretum*

samples have previously (Chapter 2) amplified mostly multiploid loci for microsatellite markers, which complicates their utility for population genetic analysis (DUFRESNE *et al.* 2014). The *N. viretum* samples were therefore omitted from the microsatellite analyses. For most locations, a single site was selected for inclusion in the microsatellite analyses. The exceptions to this were locations B and D, with the purpose of investigating microgeographic patterns based on diverse geological soil types underlying different populations. In B, two transects were taken along the northern (Bn, 2.4km long) and southern (Bs, 6km long) ridgelines surrounding the valley, and samples were also taken from a single site in the far west of the valley (Bw, 10mx2m). From location D, sites D1 and D4 were selected to represent individuals on different geological soil types, and D5 to represent a wider spatial distribution.

3.3.3.2 Genotyping and Data Processing

Twelve microsatellite loci (described in Chapter 2) were used to analyse individuals using the PCR and genotyping methods described in Chapter 2. Six loci were multiplexed at a time to reduce genotyping costs. To this end, loci with non-overlapping size ranges shared one of three fluorescent dyes. PCR products from different markers required different dilution factors: Na0003 and Na0010 were diluted 1:50; Na0298, Na0078, Na27207, Na0011, and Na11999 were diluted 1:25; Na6568 and Na1332 were diluted 1:10; Na0043 and Na8420 were undiluted; and 2x the amount of Na20965 was added to the pool. Alleles were manually scored in GENEMAPPER v4.0 (Applied Biosystems) and formatted for GenAlEx v6.5 (PEAKALL AND SMOUSE 2006; PEAKALL AND SMOUSE 2012). For cases where more than two alleles were amplified from an individual for a given marker, the genotype for the locus was coded as missing, and individuals with greater than 75% missing data were removed from the dataset. To identify potential allele scoring errors, null alleles, and short allele dominance, MICRO-CHECKER v2.3.3 (VAN

OOSTERHOUT *et al.* 2004) was employed. To verify the rate of null alleles at each locus and within sites, both the ML-NULL-Freq (KALINOWSKI AND TAPER 2006)—as suggested by DĄBROWSKI *et al.* (2015)—and the freeNa (CHAPUIS AND ESTOUP 2006) programmes were used.

3.3.3.3 Detecting Genetic Boundaries within Notothlaspi

The Bayesian computation software STRUCTURE v2.3.3 (PRITCHARD *et al.* 2000) was used to identify genetic clusters. For each analysis, 100,000 iterations were discarded as burnin, and one million Markov chain Monte Carlo iterations retained. The admixture model was used, allele frequencies were set to correlated, and K=1–13 were analysed with 11 replicates each. STRUCTURE output results were imported into the web version of STRUCTURE HARVESTER v0.6.94 (EARL 2012), where the Evanno method was employed to produce mean likelihood values of *K*. CLUMPAK (KOPELMAN *et al.* 2015) was used to find consensus among the replicates at each K value using the Markov clustering algorithm of CLUMPP (JAKOBSSON AND ROSENBERG 2007), and to format the output results using DISTRUCT (ROSENBERG 2004).

The full genotype dataset was imported into R Studio (RSTUDIO-TEAM 2020) and converted into a Provesti distance matrix using the poppr package v2.9.2 (PREVOSTI *et al.* 1975; KAMVAR *et al.* 2019). Provesti distance was selected as there are no underlying biological assumptions. The Provesti distance matrix was exported as a Nexus file into SplitsTree5 (HUSON AND BRYANT 2006) to create a Neighbor-Net. To generate a pairwise distance (F_{ST}) matrix, freeNA (CHAPUIS AND ESTOUP 2006) was used, as the ENA method corrects for missing data bias. The input file for freeNA was modified to correct for multiploid loci coded as missing data, as instructed by the software manual.

3.3.3.4 Determining the Distribution of Genetic Variation within *N. australe* and *N. rosulatum*

The original nuclear genotype dataset was partitioned into two based on initial assignment of sites to species. The 19 individuals from location (L), with unknown species identity, were included in both datasets. The *N. australe* dataset contained 168 individuals and the *N. rosulatum* dataset contained 102 individuals. Each dataset utilised the information from all 12 microsatellite loci.

Both Neighbor-Net and STRUCTURE analyses were performed on the subdivided datasets, as described above. To observe how variation is subsetted, an AMOVA (Analysis of Molecular Variance) was performed in GenAlEx, using 999 permutations. A Principal Coordinates Analysis (PCoA) and Discriminant Analysis of Principal Components (DAPC) were used to show the partitioning of individuals without (PCoA) and with (DAPC) prior assumed knowledge of site groupings. The PCoA plot was created by first making a Euclidean distance matrix. The corresponding eigenvalues were calculated in R, and the ggplot2 package (WICKHAM 2011) was used to make plots for each dataset. DAPC is a multivariate clustering method that initially finds clusters using successive K-means, before comparing the clusters (JOMBART *et al.* 2010). This approach minimises within-group variation and instead focuses on between-group variation. Prior to the DAPC analysis, the 'find.clusters' function in ADEGENET (JOMBART 2008) was used to determine appropriate groupings of individuals. For the *N. australe* dataset, the lowest BIC number lies at seven clusters, indicating the best 'goodness of fit'. This grouping clumped individuals from the same site together and sorted all individuals from location B in the same group. The lowest BIC number for the *N. rosulatum* dataset was five, incorporating individuals from sites J, M, N, and P in the same cluster, and sorting individuals from L, G, H, and Q into distinct clusters according to site. This information was used for the

DAPC analysis, where 60 Principal Components (PCs), and 100 discriminant functions were retained for *N. australe*, and 15 PCs with 100 discriminant functions retained for *N. rosulatum*. To test if isolation by distance explains genetic patterns, a codominant Mantel test was performed in GenAlEx with 999 permutations. The geographic distance vs genetic distance plot coordinates were imported into R to make a presentable graphic.

GenAlEx v6.5 was used to determine: observed and effective number of alleles; number of private alleles; observed and expected heterozygosity; percent polymorphic loci; and Wright's fixation index. Percent amplification was calculated manually.

3.4 Results

3.4.1 Sample Collection

During collections, all individuals were easily identifiable by morphology except individuals from location L. *Notothlaspi australe* have a perennial growth habit, often forming cushions with multiple vegetative rosettes per individual, with axillary, solitary flowers. *Notothlaspi rosulatum* individuals, in contrast, are biennial and monocarpic, with a pyramidal inflorescence during the reproductive phases. At the time of sampling, individuals at location L were tentatively assigned to *N. rosulatum* due to leaf shape, size, and the crown of flowers (Figure 3.2). However, close examination of photographs taken at the time revealed that some individuals had vegetative rosettes above the floral meristems, suggesting a perennial life-cycle akin to *N. australe.* Several individuals also had multiple rosettes. Previously, plants collected from this location have been described as both *N. rosulatum* and *N. australe,* seen in herbarium records at the Allan and Te Papa herbaria, as well as the public iNaturalist database (Available from https://www.inaturalist.org/observations?place_id=any&subview=map&taxon_id=36660

4). As these individuals could not be identified with confidence, they were included in analyses of both *N. australe* and *N. rosulatum* but excluded from averages of within-

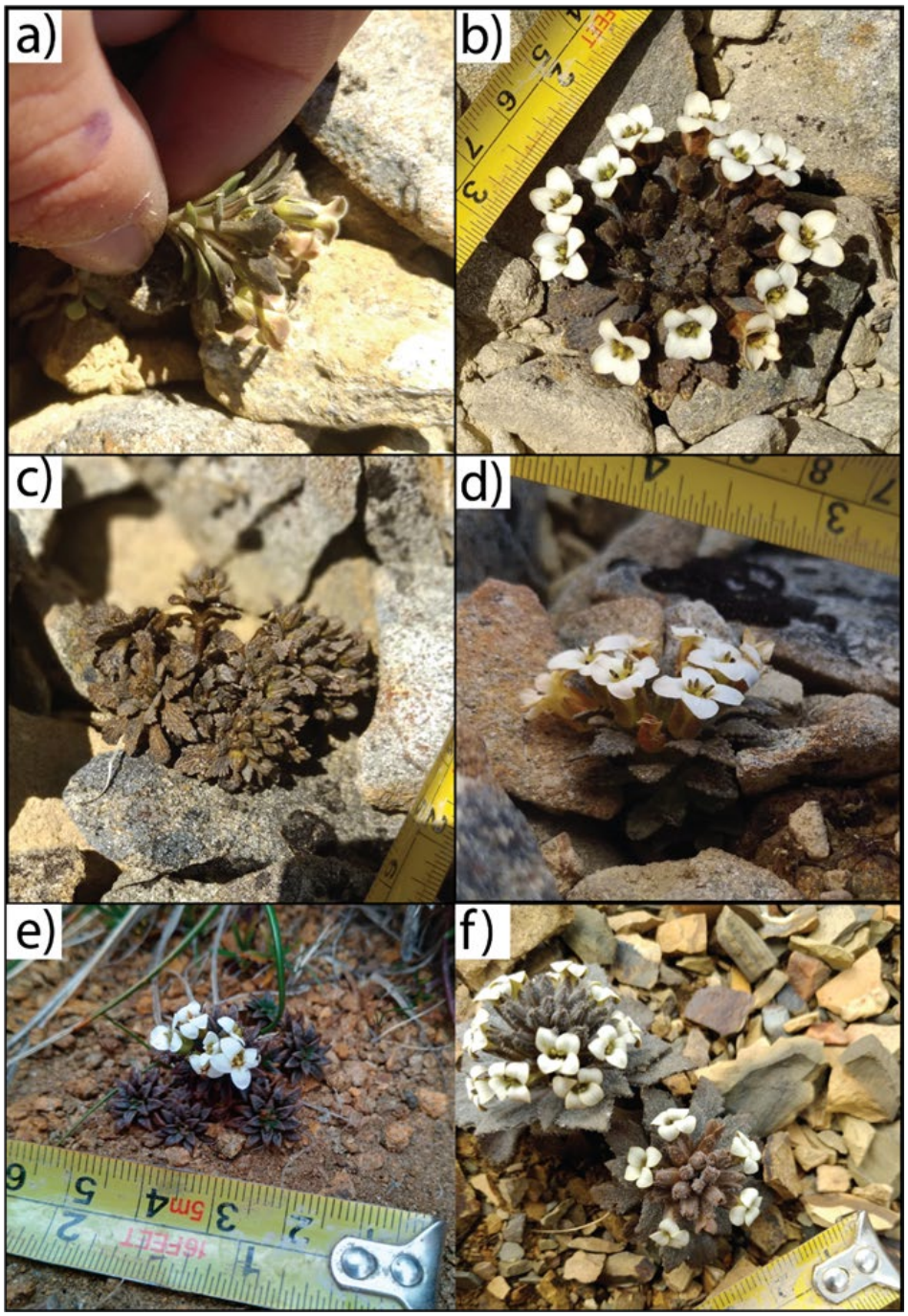


Figure 3.2: Photographs showing morphology of *Notothlaspi* from site L (Nelson Lakes, New Zealand), alongside *N. australe* and *N. rosulatum* individuals. a) side-on image of a plant from site L. b) Aerial image showing a vegetative rosette above flowers on a site L individual. c) Multi-stemmed site L plant with flower buds. d) Side-on view of a plant from L. e) *N. australe* individual from location I at a similar reproductive stage to individuals in panels a-d. f) Two *N. rosulatum* individuals from location H.

species diversity statistics. The geological soil parent material for each location and/or site is found in Table 3.2. *N. australe* sites from the Cobb valley are on a series of volcanic conglomerate, slate, and limestone that is mirrored on northern and southern ridges of the valley. In the southern Kahurangi National Park, sites D1-D3 are on marble-derived soils. Site D4 lies 4.1km away from D1 on coarse granite sand soils. The Richmond Forest Park *N. australe* sites lie on ultramafic rock. *N. rosulatum* locations are usually found on soils derived from successions of greywacke and argillite common in the north eastern South Island. Locations G and H are exceptions, G is found on schistose parent material, while H1 is on a coarse sandstone and H2 on a banded siltstone.

Table 3.2: Geological substrate at different sites.

Species	Location	Description	Geological substrate
N. australe	B	Northern ridge (Bn)	Volcanic conglomerate
			Slate
		Southern ridge (Bs)	Volcanic conglomerate
			Siliclastic mudstone/ slate
			Limestone
		Western site (Bw)	Schist
	D	Mt Owen (D1)	Marble
		Lookout Range (D4)	Granite sand
		Biggs Tops (D5)	Sandstone
	E	Dun Mountain	Dunite
	I	Red Hills	Pyroxenite and dunite
Unknown	L	Mt Robert	Sandstone
N. Rosulatum	G	Mt Richmond	Semi-schist
	H	Inwoods Ridge (H1)	Coarse sandstone
		Inwoods Ridge (H2)	Banded siltstone

3.4.2 Sequencing Chloroplast Loci

In addition to sequences from individual samples, the DNA from up to twenty individuals was pooled to find within-site polymorphisms. To test the detection limits of this approach, four pools with different ratios of two genetically distinct individuals were

amplified and sequenced. Figure 3.3 shows Sanger sequence chromatograms of the pooled samples alongside a 'pure' sequence from the respective source sites. Only in the 1:1 ratio chromatogram were the single nucleotide polymorphisms separating the two sites detectable. In all other ratios, M1-03 is indetectable. While this result did not provide confidence in the ability to detect within-site polymorphisms beyond a 1:1 ratio, individual sequences concurred with the pooled DNA from a given site in all cases except the pooling of E1 and E3. A pool of individual DNAs from sites E1 and E3 was sequenced alongside single-individual sequences from each site (Table 3.1). Individual sequences from E1 and E3 are separated by a single mutation, but the pooled PCR products of E1 and E3 individuals had the same sequence results as E3.

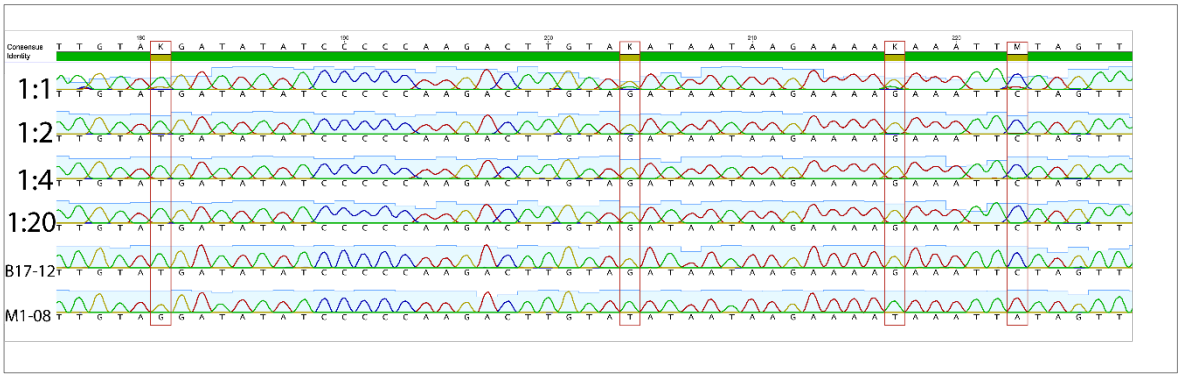


Figure 3.3: Chromatograms in GENEIOUS v9 showing Sanger sequencing results in the TabF direction. Row 1 has a 1:1 ratio of individual B and individual M, row 2 has a 1:4 ratio, row 3 has a 1:10 ratio, and row 4 has a 1:20 ratio. Rows 5 and 6 are individual s from sites B17 and M1. Vertical red boxes highlight single nucleotide polymorphisms between individuals from sites B and M.

The trnL-F locus was used to assess chloroplast sequence variation among *Notothlaspi* sites. The results uncovered two major genetic groupings separated by four mutations, as shown in both the haplotype network (Figure 3.4a) and rooted phylogenetic tree (Figure 3.5). Cluster I is limited to the north of the *Notothlaspi* distribution, and contains all sampled *N. australe* individuals, the *N. viretum* herbarium samples, and individuals from two *N. rosulatum* sites (G and J). The most common haplotype, Ia, is shared by the *N. rosulatum* site G and all *N. australe* sites except site E3 and location I. Haplotype Ib (site E3) is differentiated from Ia by a single mutation.

Haplotype Ic is shared by all sampled individuals in the Red Hills, including both *N. australe* and *N. viretum,* except one *N. viretum* sample (If). The two individuals sampled from site J share haplotype Id, which diverges from the rest of cluster I at the same point as haplotype Ic (Figure 3.4a and Figure 3.5). Cluster II has exclusively *N. rosulatum* individuals. Three sites share haplotype IIb, making it the most common in cluster II. IIb, has a long north to south distribution, over 100km from location H to location P. Two haplotypes separated from IIb by a single mutation belong to M4 (IIc) and N1 (IId), both in the central part of the sampled *N. rosulatum* locations. Individuals from location Q have the IIa haplotype, which is differentiated from the other sampled individuals by three mutations.

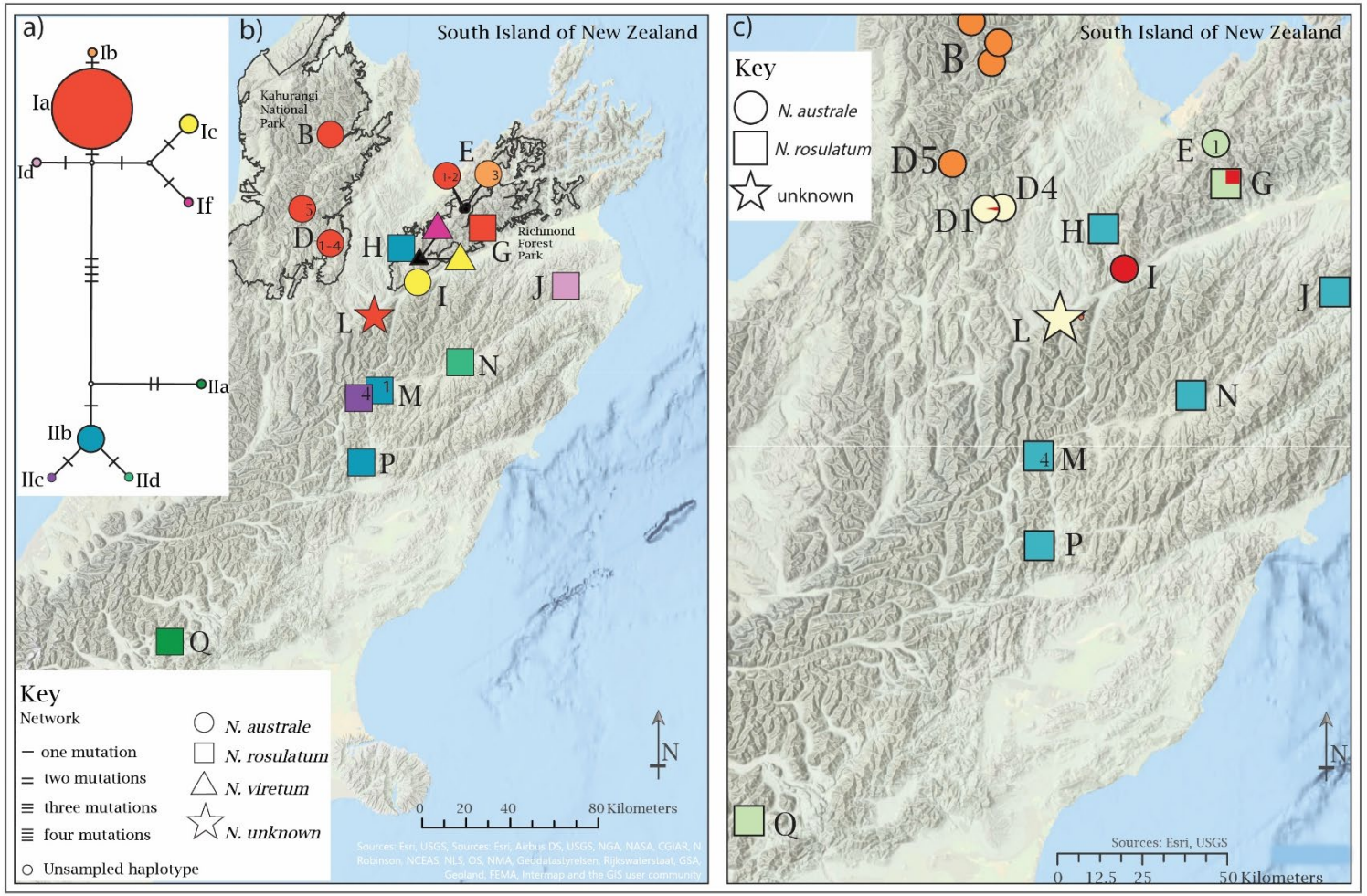


Figure 3.4: Haplotype network, map of chloroplast results, and map of nuclear results. Circles indicate *N. australe* locations, and squares are *N. rosulatum* locations. Small numbers within these shapes indicate which site within the location each shape represents. a) Haplotype network based on chloroplast results. b) Map of chloroplast haplotypes across the landscape. c) A map of nuclear variation distributed across the landscape based on STRUCTURE results at K=5.

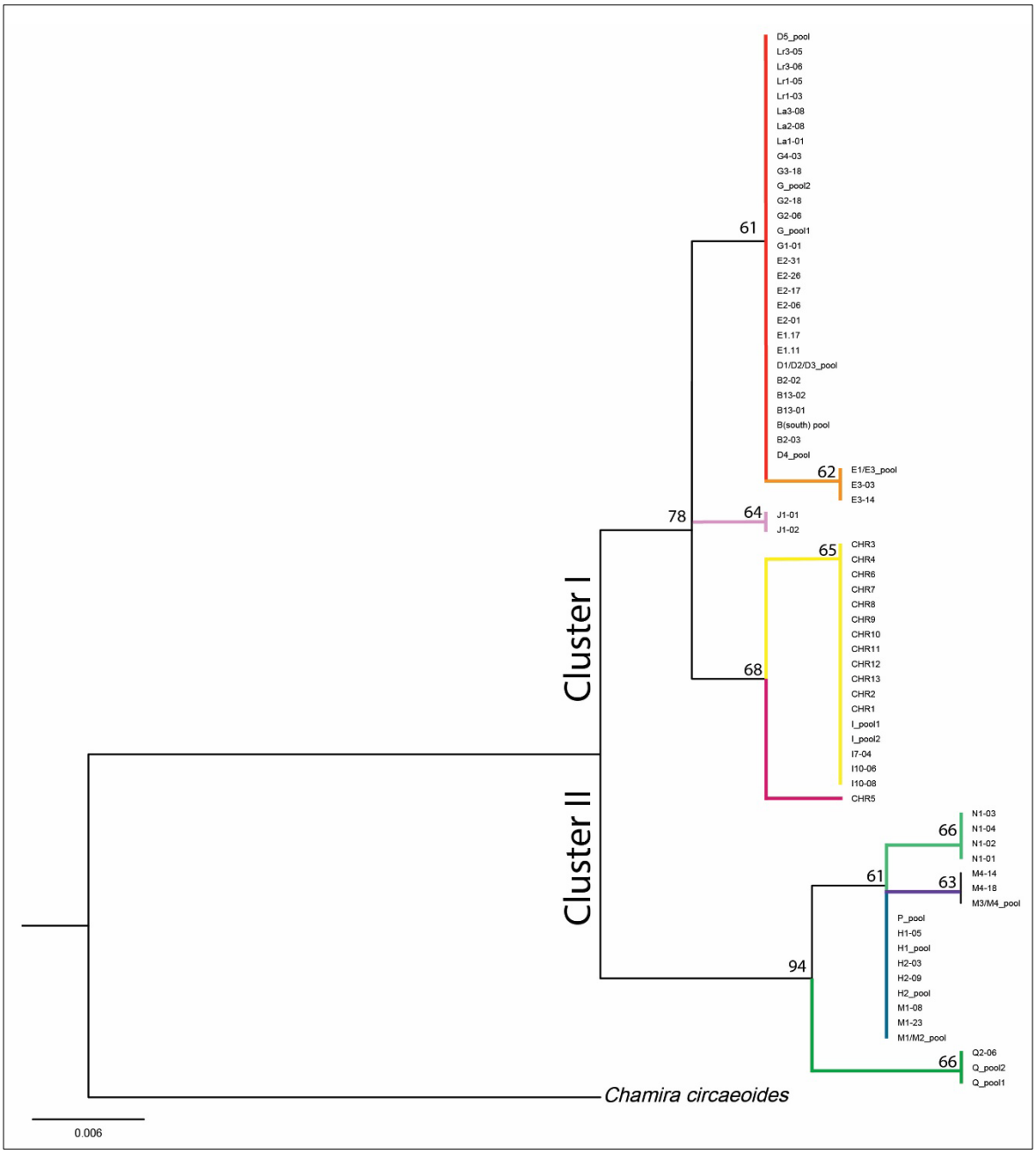


Figure 3.5: Maximum Likelihood tree based on chloroplast data. *Chamira circaeoides* is the outgroup. Line colours are the same as seen in Figure 3.4a.

3.4.3 Detecting Genetic Boundaries within *Notothlaspi*

A STRUCTURE analysis was used to identify genetic groupings in the full microsatellite dataset (Figure 3.6). Using STRUCTURE HARVESTER, K=11 was shown to fit the data the best (delta K=428.4). This high delta K value at K=11 may be due to a drop

in likelihood (mean) between *K*=11 and *K*=12. *K*=12 had a delta *K* value of 0.15, and *K*=13 had one of 0.19, so these graphs were not shown in the figure. The results at K=5 were plotted on a map (Figure 3.4c) for visualisation. For the most part, the clustering at *K*=2 grouped individuals from each species together, except for a few individuals from location I, and all individuals from location E, which had strong signatures of admixture. Some shared group membership between E1 and *N. rosulatum* individuals, particularly sites in G and Q, was also seen at *K*=4, *K*=5, *K*=6, *K*=7, and K=9. Individuals from location L shared membership with *N. australe* at all K values, particularly those from sites D1 and D4, although at *K*=6, some partial membership with Q was noted. Patterns observed within each species will be discussed below.

The same full microsatellite dataset was used to generate an individual-level Provesti distance matrix and Neighbor-Net. As seen in Figure 3.7, individuals tend to form clusters conforming to sampling sites, except for two larger groupings of sites, one in each species. Within *N. australe*, individuals from location B cluster with those from D5. In *N. rosulatum,* a large grouping of H, M, N, J, and P was segregated from G and Q. While there is poor resolution of broader splits, such as between species, a small split did separate *N. australe* and *N. rosulatum* sites. The within-species patterns will be discussed further below.

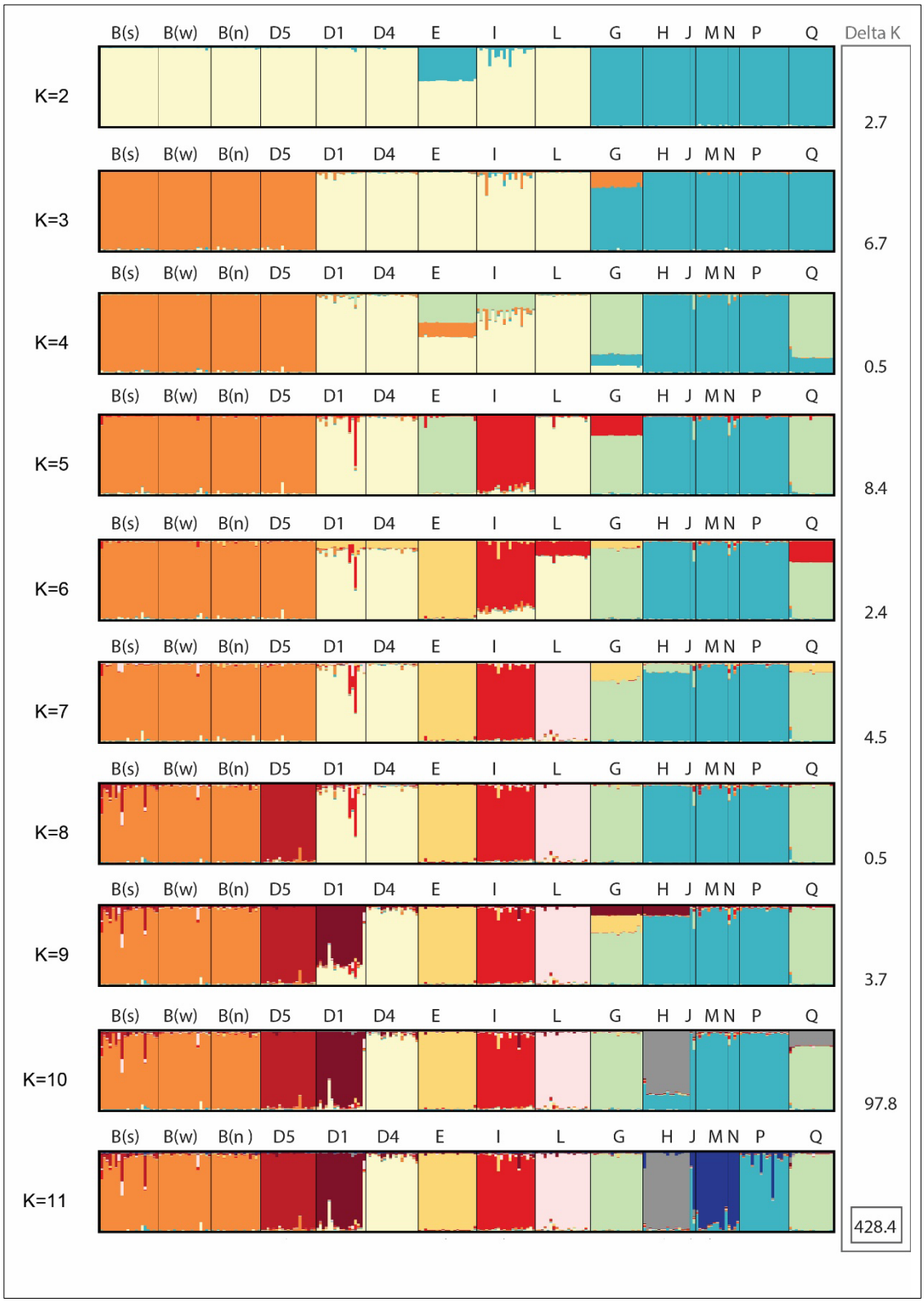


Figure 3.6: STRUCTURE results for the full dataset. K=2 to K=11 are reported here. Delta K values are reported on the right of the figure for each K value.

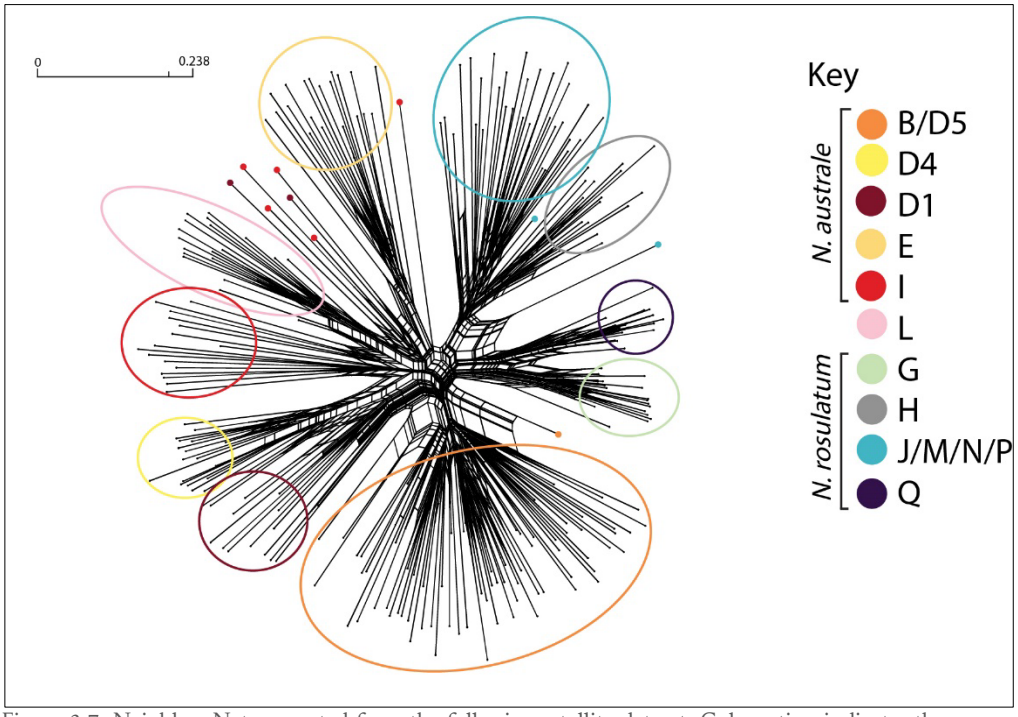


Figure 3.7: Neighbor-Net generated from the full microsatellite dataset. Colouration indicates the location/site of each individual.

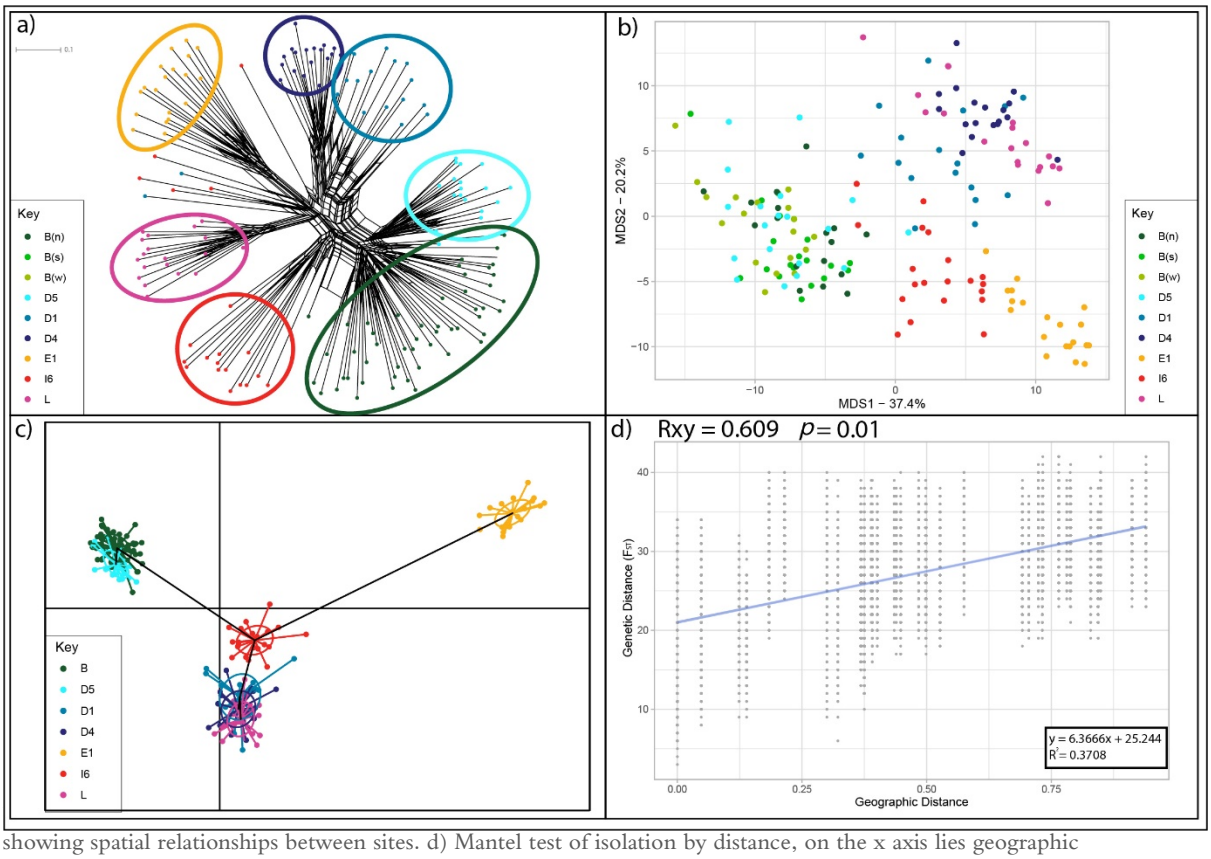
3.4.4 Determining the Distribution of Genetic Variation within *N. australe* and *N. rosulatum*

The full dataset was partitioned into two groups, one for each species, so that patterns of genetic variation within each species could be examined at higher resolution in isolation, and species-specific diversity statistics could be generated. Each dataset contains site L individuals, due to its morphological ambiguity; however, species-specific metrics excluded these samples.

3.4.4.1 *Notothlaspi australe*

Several patterns, such as the grouping of individuals from location B and site D5 (B/D5), were found in common in the Neighbor-Net, PCoA, DAPC, and STRUCTURE analyses. The largest split in the Neighbor-Net was between B/D5 and all other individuals in the dataset (Figure 3.8a). The PCoA, axis MDS1, which explained 37.4% of the variation, broadly separated individuals from B/D5 from all others (Figure 3.8b).

This was also seen in the DAPC, where B/D5 was spatially segregated from all other sites (Figure 3.8c). In the STRUCTURE plots, the groupings at *K*=2 split B/D5 and all other sites, after which, individuals from B/D5 shared membership to the same cluster up to *K*=5 (Figure 3.9). At *K*=7, the most supported *K* value, albeit weakly, according to the Evanno method, all sites have membership to different clusters except those within location B, which are grouped together up to *K*=9.



showing spatial relationships between sites. d) Mantel test of isolation by distance, on the x axis lies geographic distance, y axis is genetic distance.

Another pattern common among the four analyses was the grouping of D1 and D4 (D1/D4), which are distinct from one another but segregated from other sites. This is apparent in the Neighbor-Net (Figure 3.8a); the two are isolated from other groups by a moderately long branch, on which individuals from D4 are tightly clustered together. In the PCoA, D1/D4 and L (D1/D4/L) are separated from other individuals by both axes,

but do not cluster tightly (Figure 3.8b). In the DAPC plot, D1, D4, and L occupied the same space, most closely linked to I (Figure 3.8c). STRUCTURE supported this pattern of close association between individuals from D1, D4, and L, with all three sharing the same cluster from K=2–K=4, and D1/D4 sharing a cluster until K=6, to the exclusion of other sites (Figure 3.9).

The Neighbor-Net, PCoA, DAPC, and STRUCTURE all show individuals from E1 and I to be distinct from other sites. In the Neighbor-Net, there was clear clustering of individuals from each site, separate from other sites, which was also seen for individuals from location L (Figure 3.8a). The PCoA also separated these two sites out (Figure 3.8b). I was centred in the plot, with some individuals bridging the gap between B/D5 and D1/D4/L. Individuals from E1 mostly segregated away from other individuals in a loose cluster. The DAPC plot also put I in the middle of the *N. australe* spatial genetic distribution, most closely associated with D1/D4/L (Figure 3.8c). E is the most distant from all other sites in this plot (Figure 3.8c). In the STRUCTURE analysis, E1 and I initially clustered together at K=3, but following that, individuals within each site clustered exclusively on a site-specific basis (Figure 3.9).

A pattern of isolation by distance within *N. australe* was weakly supported by the Mantel test (Figure 3.8d) with a low, positive R^2 value (0.3708, $P = 0.01$). The AMOVA showed that 59% of variance was found within individuals, 15% among individuals, and 26% among populations.

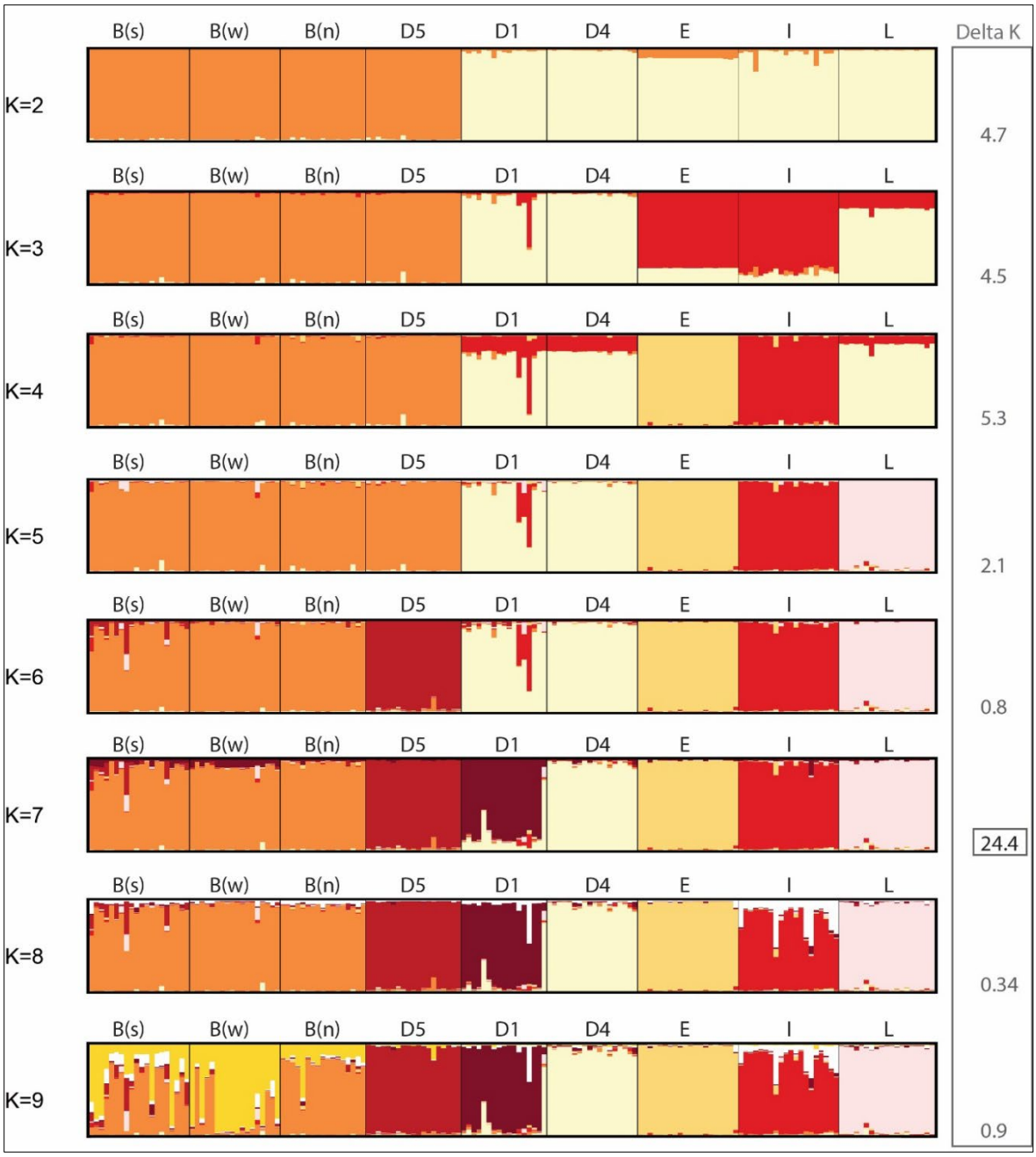


Figure 3.9: *Notothlaspi australe* STRUCTURE plots, K=2 to K=9. Delta K values on the right.

The summary statistics for each site are reported in Table 3.3. Mean number of alleles (Na) per locus in *N. australe* sites was 5.9; ranging 10.08 alleles (site I), to 3.92 (site D5). Site E1 had the highest number of private alleles (8), and site D5 had the lowest (1). The mean number of private alleles was 7.3. Ne was largest in site I (5.29), while site D4 had the lowest Ne (1.91). All *N. australe* sites had 100% polymorphic loci except D4

(91.67%). Mean observed heterozygosity was 0.50, ranging from 0.34 (site D4) to 0.65 (site I). Mean expected heterozygosity (H_E) in *N. australe* was 0.6. Again, site D4 had the lowest H_E value of 0.39, and site I has the highest at 0.74. Wright's fixation index showed a slight homozygote excess in all sites except B(s), sites are all nearly sitting at Hardy-Weinberg Equilibrium.

Table 3.3: Population genetics statistics. Number of individuals (n), number of alleles (A), private alleles (pA), number of effective alleles (Ne), percent polymorphic loci (%P), observed heterozygosity (H_O), effective heterozygosity (H_E), and Wright's fixation index (F_{IS}).

Species	Site code	Spatial spread (m)	n	A	pA	Ne	%P	H_O	H_E	F_{IS}
N. australe	B(n)	2530 transect	20	5.25	3	3.71	100	0.51	0.64	0.20
	B(s)	3700 transect	18	6.58	7	3.87	100	0.49	0.60	0.15
	B(w)	5x2	17	5.17	5	2.93	100	0.48	0.51	0.00
	D5	20x5	19	3.92	1	2.48	100	0.50	0.53	0.05
	D1	30x20	17	6.17	7	3.21	100	0.55	0.64	0.15
	D4	12x10	18	4.58	4	1.91	91.67	0.34	0.39	0.11
	E1	6x15	20	5.58	7	3.62	100	0.58	0.65	0.10
	I	35x10	20	10.08	20	5.29	100	0.65	0.74	0.11
N. australe Mean			18.5	5.9	7.3	3.2	98.6	0.5	0.6	
Unknown	L	30x15	19	5.17	10	2.56	100.0	0.55	0.53	-0.03
N. rosulatum	G	5x2	18	3.67	12	2.57	83.33	0.39	0.39	0.02
	H	20x30	16	3.67	1	2.22	91.67	0.41	0.46	0.14
	J	0.5x0.5	2	2.08	1	1.93	83.33	0.54	0.43	-0.27
	M4	5x9	11	5.42	3	3.57	100	0.55	0.63	0.10
	N	3x5	4	3.67	4	2.87	100	0.66	0.60	-0.12
	P	16x5	17	5.67	4	3.49	91.67	0.41	0.57	0.22
	Q	300x1	15	2.83	4	1.96	66.67	0.22	0.30	0.21
N. rosulatum Mean			11.9	3.9	4.1	2.7	88.1	0.5	0.5	

Pairwise F_{ST} (Table 3.4) within *N. australe* ranged from 0.04 (B(s) and B(n)) to 0.42 (L and B(w), and D4 and D5). The smallest values were found within the Cobb valley (B), while the largest values were found in pairwise comparisons with D4 and L. Sites D1, D5, and I all had low to moderate pairwise comparisons.

Table 3.4: Pairwise F_{ST} between sites. Boxes are around sites in the *N. australe* dataset, and those in the *N. rosulatum* dataset.

Site	B(s)	B(w)	B(n)	D1	D4	D5	E	I	L	G	H	J	M	N	P
B(w)	0.06														
B(n)	0.04	0.07													
D1	0.21	0.28	0.18												
D4	0.37	0.41	0.35	0.18											
D5	0.09	0.17	0.15	0.25	0.42										
E	0.30	0.35	0.26	0.24	0.37	0.33									
I	0.16	0.21	0.14	0.16	0.32	0.20	0.21								
L	0.36	0.42	0.33	0.24	0.34	0.38	0.35	0.26							
G	0.40	0.45	0.38	0.44	0.59	0.43	0.40	0.31	0.53						
H	0.35	0.42	0.33	0.29	0.47	0.40	0.34	0.31	0.39	0.53					
J	0.27	0.37	0.25	0.28	0.50	0.36	0.31	0.20	0.35	0.44	0.29				
M	0.29	0.36	0.25	0.26	0.44	0.35	0.31	0.23	0.33	0.46	0.23	0.13			
N	0.29	0.36	0.25	0.28	0.46	0.35	0.29	0.22	0.36	0.44	0.22	0.10	0.11		
P	0.34	0.40	0.28	0.30	0.47	0.38	0.35	0.28	0.37	0.50	0.26	0.17	0.08	0.19	
Q	0.42	0.47	0.41	0.45	0.58	0.48	0.46	0.34	0.53	0.52	0.52	0.48	0.44	0.46	0.50

3.4.4.2 *Notothlaspi rosulatum*

As in the analyses of the *N. australe* dataset, there were several patterns observed consistently across analyses. The first pattern was the strong grouping between individuals from sites J, M, N, and P (J/M/N/P), and sometimes H (H/J/M/N/P). The Neighbor-Net clusters J/M/N/P individuals together, with very little internal structure based on site identity (Figure 3.10a). The closest cluster to this exclusively contained H individuals. In the PCoA, H/J/M/N/P individuals are separated from all other sampled *N. rosulatum* and L individuals (Figure 3.10b). Five clusters, one with a combination of J/M/N/P, was shown to best fit the data for the DAPC analysis using the find.cluster function in ADEGENET (Figure 3.10c). This grouping was closely associated with H, but spatially distant from L, Q, and G. In the STRUCTURE plots, increasing *K* values showed the breakdown of the grouping (Figure 3.11). At *K*=2, individuals from L were assigned to the same cluster as H/J/M/N/P, at *K*=3 L individuals are assigned to a separate cluster. H falls out of the cluster at K=4, but individuals still retain some membership to the J/M/N/P cluster. Following this, the pairing between M and N is the strongest, while the

two J individuals are each assigned to different groupings, one with P and the other with M/N.

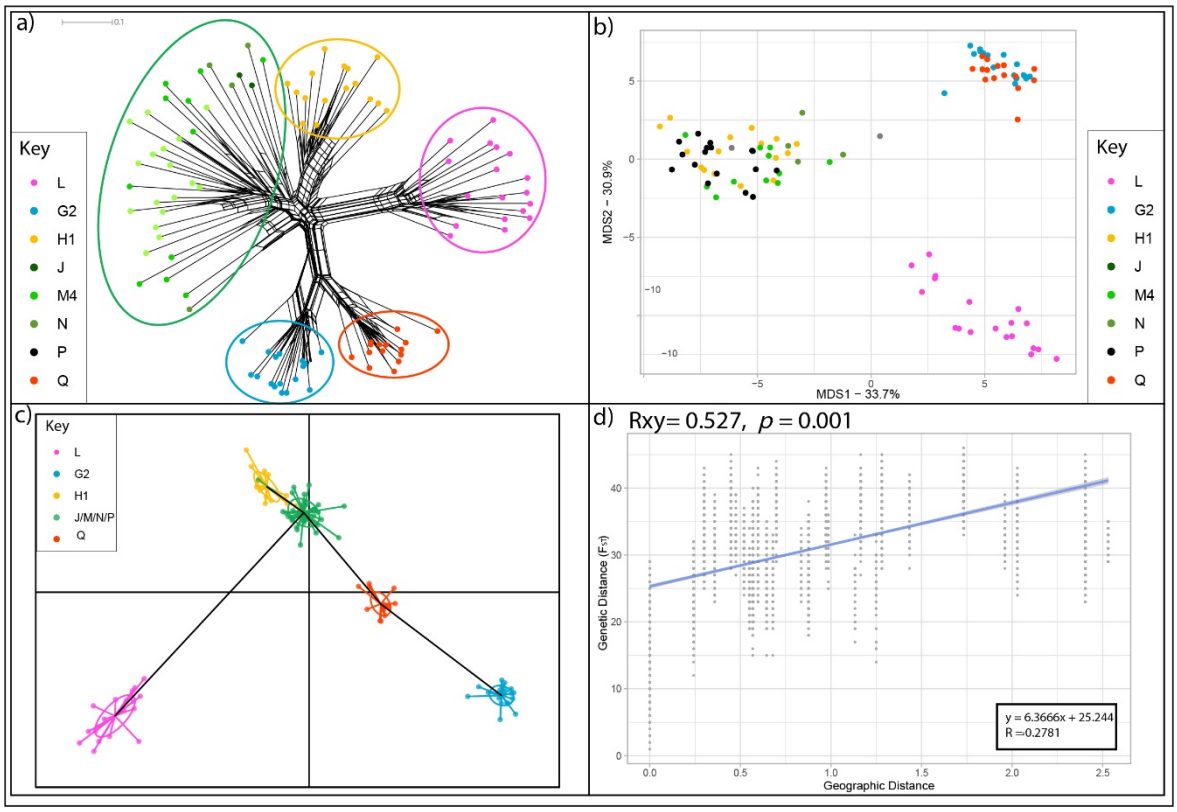


Figure 3.10 Analyses of the *N. rosulatum* dataset. a) Neighbor-Net, each node is coloured with site identity, and clusters from each site are encircled. b) PCoA with two axes explaining 33.7% and 30.9% of the data. c) DAPC plot showing spatial relationships between sites. d) Mantel test of isolation by distance, on the x axis lies geographic distance, y axis is genetic distance.

In each analysis, individuals from G and Q (G/Q) were more closely associated to each other than to other sites. In the Neighbor-Net, internal edges separated G/Q from other samples, although both G and Q still formed discrete clusters (Figure 3.10a). This association was similarly observed in the PCoA, where there was little apparent sorting within the G/Q cluster, and the group was separated from H/J/M/N/P by axis 1, and L by axis 2 (Figure 3.10b). However, the two were not clustered together in the DAPC plot (Figure 3.10c); instead, each formed a cluster segregated from other sites. In the

STRUCTURE plots, individuals from G and Q shared exclusive membership to the same cluster from K=2–K=5 (Figure 3.11).

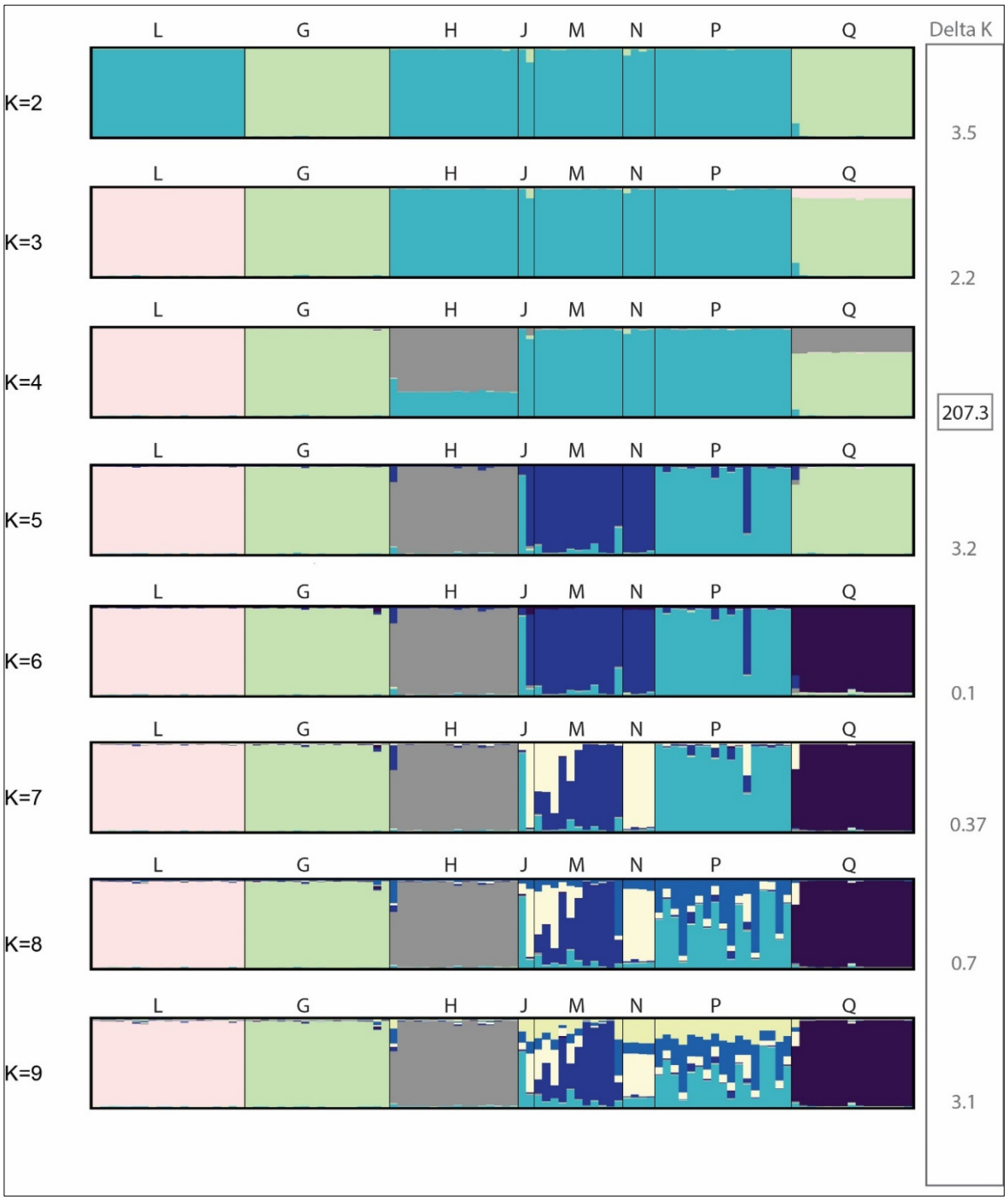


Figure 3.11: *Notothlaspi rosulatum* STRUCTURE analysis. Delta K values are on the right of the figure.

Patterns of isolation by distance among *N. rosulatum* sites were supported by the Mantel test, which returned a weak but significant R^2 value of 0.2781 ($P = 0.001$) (Figure 3.10d). The AMOVA analysis showed that 44% of variation was found within individuals, while 16% was found among individuals, and 40% among populations.

The mean number of alleles (A), private alleles (pA), effective alleles, and percent polymorphic loci (%P) were all lower in *N. rosulatum* compared with *N. australe* (Table 3.3) based on the nuclear microsatellite data. Overall, 149 *N. australe,* and 83 *N. rosulatum* individuals were used in these analyses so sample size may have an impact here. Two sites, J and N, had sample sizes below four, while the rest were above 11. Site M4 had the highest number of alleles (5.42), site J had the lowest (2.08), and the mean number was 3.9. Site G had the highest number of private alleles out of all sites sampled, at 12, while both sites H and J each had one. The mean *N. rosulatum* number of private alleles out of the sites is 4.1. Site M4 had the highest number of effective alleles (3.57), and J had the lowest (1.93); the overall mean was 2.7. Unlike in *N. australe*, only two sites (M4 and N) had 100% polymorphic loci, and the mean value was 88.1%. The lowest value was for site Q at 66.67%. Observed heterozygosity ranged from 0.66 (N) to 0.22 (Q), with a mean of 0.5. For expected heterozygosity the highest value was 0.63 (M4), and the lowest 0.3 (Q). Two sites had negative F_{IS} values, J (-0.27) and N (-0.12), and the rest had a mild to moderate homozygote excess.

Pairwise F_{ST} values within *N. rosulatum* ranged from 0.1 (M and P) to 0.53 (G and L, G and H, Q and L) (Table 3.4). G and Q were the most differentiated from other *N. rosulatum* sites, each with F_{ST} scores ranging from high to very high in all comparisons. L also had a moderate similarity to M, N, and P, but high or very high differentiation from other sites. Within H, J, M, N, and P, low to moderate levels of differentiation were found.

3.5 Discussion

This study describes the first population genetics investigation of New Zealand alpine plants, to our knowledge. Three species of the genus *Notothlaspi* were examined for spatial patterns of genetic diversity across the upper South Island of New Zealand. We have discovered a broad biogeographic pattern of genetic variation with *N. australe* to the west, *N. rosulatum* to the east, and a zone of overlap and potential hybridisation in the centre. Using a population genetics approach, interactions among the three species and within-species genetic diversity were investigated.

3.5.1 Detecting Genetic Boundaries within *Notothlaspi*

3.5.1.1 The geographic pattern of species boundaries

N. australe to the west (sites B and D), and *N. rosulatum* to the south east (sites M, N, P, and Q) of the Richmond Forest Park show no signs of admixture between species. Individuals from these locations appear to represent a clear disconnect based on historical divergence between the species, indicating a period of reproductive isolation. Divergence among lineages occurs through a lack of gene flow. While this can occur with (sympatry) or without (allopatry) contact between lineages, the current distributions of *N. australe* and *N. rosulatum* would suggest that it took place in isolation.

Without suitable amounts of chloroplast sequence data, no attempt has yet been made to time-calibrate the split between *N. australe* and *N. rosulatum,* however, during the Plio-Pleistocene period opportunities for both sympatric and allopatric divergence may have been available. Numerous radiations within the New Zealand flora have been dated to the Plio-Pleistocene period, where glacial cycles, volcanism, and tectonism altered plant distributions (LOCKHART *et al.* 2001; WINKWORTH *et al.* 2005; SHEPHERD *et al.* 2007; JOLY *et al.* 2009a; HEENAN AND MCGLONE 2013; MEUDT *et al.* 2015). At the last

glacial maximum (around 21,000ya (MIX *et al.* 2001)), glaciers were found throughout the present-day *Notothlaspi* distribution in the Kahurangi National Park, Richmond Forest Park, and Marlborough region (NEWNHAM *et al.* 2013).

3.5.1.2 Semi-permeable species boundaries.

Results indicate that between clear *N. australe* and *N. rosulatum* populations to the west and east lies a zone where sites have a mixture of *N. australe, N. rosulatum,* and *N. viretum* genotypes. The Richmond Forest Park hosts habitat suitable for all three species. *N. rosulatum* has been found on rocky scree-like slopes near Gordon's knob, along the Ben Nevis Ridge, Mt Patriarch, Slaty peak, Mt Robert, and Mt Fishtail according to herbarium records (AVH 2021. The Australasian Virtual Herbarium, Council of Heads of Australasian Herbaria, https://avh.chah.org.au, accessed 24 July 2021). In contrast, *N. australe* and *N. viretum* are restricted to open sites on ultramafic rock. Although the ranges overlap, *N. rosulatum* is not known to be sympatric with the other two species.

The signatures of mixed genotypes could result from different biological phenomena, such as incomplete lineage sorting (ILS), or hybridisation (JOLY *et al.* 2009b). Patterns of ILS are likely to have a widespread distribution across populations of the species (SEGATTO *et al.* 2014; GAO *et al.* 2015; ZHOU *et al.* 2017). This makes ILS an unlikely explanation for the shared genotypes of *N. australe* and *N. rosulatum,* for which mixed genotypes were nearly exclusively detected in the Richmond Forest Park. The only known *N. viretum* population is found in the Richmond Forest Park, therefore, ILS cannot be so easily dismissed for the shared chloroplast haplotype between *N. viretum* and *N. australe* from further down the ridge. Conversely, the mixed genotypes could be from admixture.

Hybridisation and introgression have several evolutionary consequences for populations, including allowing the transfer of adaptive alleles between groups, or reinforcing current species boundaries (HARRISON AND LARSON 2014). Gene flow can increase the evolutionary potential of the recipient population(s) by introducing pre-adapted genes (SUAREZ-GONZALEZ *et al.* 2018; MITCHELL *et al.* 2019). Compared with random mutation, this increases the speed at which alleles can be introduced and fixed (BARRETT AND SCHLUTER 2008). Conversely, new alleles can 'dilute' more fit alleles in locally adapted populations, which can lead to selection against migrants and to further divergence.

N. australe sites E and I, and *N. rosulatum* site G all have mixed nuclear genotypes that could have resulted from gene flow between the species. *N. australe* site E1, from near Dun Mountain, presents the strongest case for admixture based on nuclear data. In all analyses, individuals are clearly diverged from other *N. australe* sites, and, in cross-species comparisons there is clear association with the nearby *N. rosulatum* site G. At K=2 (Figure 3.6), individuals from site E1 have over one third membership to the *N. rosulatum* cluster. Some individuals from *N. australe* site I also had membership to the *N. rosulatum* cluster at K=2, up to about one-quarter. The nuclear STRUCTURE results and comparatively low F_{ST} scores indicate shared genotypes with *N. rosulatum* populations. There is no evidence of geneflow with H, the closest *N. rosulatum* site sampled, which could be a potential pollen source as it is upwind (according to the prevailing westerly wind (MACARA 2016)). Instead, G is a more likely candidate for geneflow out of the sampled sites. Individuals from location G have a clear *N. rosulatum* phenotype, with a terminal, pyramidal inflorescence (Supplementary Figure 1). Conversely, nuclear results show some shared genotypes with *N. australe*, particularly site E1, at eight of the eleven reported K values (Figure 3.6).

Interestingly, all three of these potential sites of admixture have high diversity statistics compared with others of the same species, particularly regarding unique alleles. Site I6 in particular has the highest number of private alleles, 20 compared with 1-7 private alleles in all other *N. australe* sites. Analyses of the *N. viretum* nuclear data may show if these private alleles are shared with *N. viretum.* As mentioned above, hybridisation can introduce novel combinations of alleles, potentially increasing genetic diversity within the recipient population. This study has not specifically examined any adaptive nuclear alleles, however, other studies, e.g. MITCHELL *et al.* (2019), present experiments showing clear examples of more rapid adaptive evolution in populations with recent hybridisation. In a New Zealand example, BECKER *et al.* (2013) linked past introgression of adaptive glucosinolate hydrolysis alleles with the survival of *Pachycladon* (Brassicaceae) species during the last glacial maximum.

Chloroplast capture is a relatively common occurrence among plants, where plastid DNA is captured through hybridisation and backcrossing (RIESEBERG AND SOLTIS 1991). Attributed to the ease of plastid transfer, numerous studies in New Zealand and overseas have observed geographic patterns of chloroplast haplotype distribution that do not conform to taxonomic expectations (WAN *et al.* 2017; LIU *et al.* 2020; RAWLENCE *et al.* 2021). This may be the case for three *Notothlaspi* sites in, or near to, the Richmond Forest Park. All sampled individuals and pools from *N. rosulatum* location G had haplotype, Ia, the most common in *N. australe*. Most *N. rosulatum* sites have haplotypes from cluster II, which is clearly distinguished from cluster I (Figure 3.4a), it seems unlikely but not unfeasible that this result is due to ILS. Due to the proximity to *N. australe* populations, it seems likely this is a case of chloroplast capture.

Interestingly, the two individuals from *N. rosulatum* location J had a unique chloroplast haplotype from cluster I, diverged from all extant sampled *N. australe* and *N.*

viretum chloroplast haplotypes. Although only two individuals were found at location J, there are no signs of nuclear admixture and the individuals closely cluster with the Marlborough *N. rosulatum* sites M, N and P, and are geographically isolated from all known *N. australe* populations. This result puts Cluster I haplotypes exclusively north of, and Cluster II haplotypes exclusively south of, latitude -41.85°, except for location H. However, more sampling is required to verify this and increase resolution in the northern Marlborough area.

Back in the Richmond Ranges, *N. viretum* share a chloroplast haplotype with *N. australe* nearby in the Red Hills (location I). This haplotype, Ic, is differentiated from the common *N. australe* haplotype and its derivative (Figure 3.4a). This could be explained if the haplotype originated in the differentiated *N. viretum* and was later captured through gene flow by local *N. australe.* Nuclear analysis comparing *N. australe* with *N. viretum* may further elucidate the relationship between the two species.

In the face of climate change, the major natural threat to *Notothlaspi* may be a rising treeline (HALLOY AND MARK 2003; HARSCH *et al.* 2009; DÖWELER *et al.* 2021). Tree species are nearly absent from the ultramafic sites of Dun Mountain (E) and the Red Hills (I), despite the low elevations of these locations (ROBINSON *et al.* 1996). These ultramafic environments may provide refugia for *N. australe* and *N. viretum* populations during warm periods (KOLÁŘ *et al.* 2012), particularly in the Richmond Forest Park which lacks high-elevation continuous habitat. Hybridisation between *N. australe* and *N. rosulatum* could introduce tolerance of these soils into *N. rosulatum* populations.

3.5.2 Determining the distribution of genetic variation within *N. australe* and *N. rosulatum*

3.5.2.1 Genetic structure within *N. australe*

Location L, with individuals of previously unknown species identity, has a far higher genetic affinity to *N. australe* than *N. rosulatum.* We found no evidence for admixture between Location L individuals and any sampled *N. rosulatum* sites. The indeterminate morphology of individuals from Location L may be a plastic response to environmental stimuli. In this study, a substantial amount of morphological heterogeneity has been observed in *N. australe*, even between individuals occurring at the same sites. For example, at multiple sites, individuals in shaded areas (e.g. Supplementary Figure 2) were observed to have greener laminas and longer internodes compared with individuals in the open that had shorter internodes and darker pigmentation. Nevertheless, these observations were not explicitly measured. Individuals from L have high diversity statistics, such as 10 private alleles, the second highest number out of *N. australe*. This site also segregates from other sites in Neighbor-Net and STRUCTURE analyses, without the obvious explanation of admixture. A period of reproductive isolation from other *N. australe* sites could explain this pattern.

Although based on limited sampling, an interesting biogeographic pattern of genetic variation was seen across the alpine landscape of *N. australe.* Location B (Cobb Valley), and site D5 (Biggs Tops) group together in all analyses. Based on our sampling, this departs from expectations under Isolation By Distance (IBD). At K=2, B/D5 segregate from all other *N. australe* individuals, however, D5 is geographically closer to D1 and D4. Locations B and D are found in the Kahurangi National Park; however, genetically, D1 and D4 are more closely affiliated with sites in the Nelson Lakes (L) and Richmond Forest Park (I, and E) (Figure 3.1). This biogeographic split in the lower Kahurangi National

Park has not been noted previously to our knowledge, although south of this, SMISSEN *et al.* (2003) describe an unexplained split between *Raoulia* (Asteraceae) in the Paparoa Range and those in the Kahurangi National Park. Several scenarios may explain the *Notothlaspi* result, such as ILS, IBE, or diverged ancestral populations, although none have been tested.

In heterogeneous environments, isolation by environment (IBE) affects genetic structure along with IBD. Sites D1 and D4 grow on marble- and granite-derived substrates, each with unique edaphic challenges (BELL 1973; RICHARDSON *et al.* 2012). It is possible that through adaptation to substrate, populations on Mt Owen and the Lookout Range have experienced selection against crossing with individuals from D5 (on sandstone substrate) to prevent the influx of unsuitable alleles (MCNEILLY AND ANTONOVICS 1968; VAN DER NIET *et al.* 2006). To the west and north of Mt Owen are several other alpine calcareous environments that host *N. australe* populations, such as along Matiri Range, or on Mt Arthur (RATTENBURY *et al.* 1998). Determining which group these populations affiliate with may indicate whether edaphic adaptation to marble is reproductively isolating site D1 from site D5.

D4 stands out as an interesting site genetically. With the exception of D1, high pairwise F_{ST} values characterise relationships with all other sites. Site D4 also has fairly low diversity statistics. A possible explanation is a founder event from the Mt Owen massif, which is upwind and less than 2 km away. Founder events occur when relatively few individuals establish a new population (TEMPLETON 2008), so the new population has only a subset of the variation found in the parental population (GAMACHE *et al.* 2003). *N. australe* appear fairly abundant in local sites near Mt Owen, so it is possible that some seed, able to tolerate the harsh conditions of granite sand, established. Therefore, D4 may

have lower diversity statistics due to more recent colonisation by a small number of individuals from Mt Owen, or other nearby sites.

A period of geographic or ecological isolation between the ancestors of B and those of other *N. australe* individuals could also explain the split. D5 could represent a more recent expansion from the north. Support for this hypothesis lies with the low number of private alleles in D5; however, diversity statistics are moderate. Conversely, chloroplast haplotypes that might have developed during a period of isolation were not detected. *N. rosulatum,* has four sampled haplotypes, so either substitution rates are far lower in *N. australe*, or haplotype Ia has swept to fixation across most *N. australe* sites. (BAI *et al.* 2010). If current *N. australe* populations are derived from a single isolated population, for example from a glacial refugium, haplotype Ia may have become fixed there and spread with range expansion, such as in JIA *et al.* (2011). *N. australe* has a perennial habit, and thus longer generation times than *N. rosulatum* (HEENAN 2019), so mutations may accumulate more slowly (LOVELESS AND HAMRICK 1984; EYRE-WALKER AND GAUT 1997). The switch to an annual habit has been implicated in increased diversification rates elsewhere (AZANI *et al.* 2019). This common haplotype among both the B/D5 and D1/D4/L groupings does not support an origin from two reproductively isolated sites. Further sampling, with more chloroplast loci, may give greater insight on the origin of this split.

Overall, genetic diversity within *N. australe* is high, with strong patterns of genetic structure based on geography and potential admixture. Genetic diversity is relatively evenly partitioned among sites. Although, site I has considerably higher diversity than others, perhaps due to the input of alleles from both *N. rosulatum* and *N. viretum* as discussed above. AMOVA indicated that the greatest diversity is partitioned within individuals, indicating high rates of heterozygosity. F_{IS} values are also close to Hardy-

Weinberg equilibrium, so it seems likely that outcrossing is a key strategy in *N. australe* reproduction. Patterns of diversity in the Kahurangi National Park (B/D) and Nelson Lakes (L) are unresolved. However, patterns of diversity in the 'admixed zone' of the Richmond Forest Park appear closely associated with mixture of genotypes shared with *N. rosulatum* and *N. viretum.*

3.5.2.2 Genetic structure within *N. rosulatum*

As mentioned above, *N. rosulatum* has twice as many chloroplast haplotypes as *N. australe,* with four haplotypes compared to two (one haplotype may have a *N. viretum* origin). Seventeen *N. australe* sites (including L) and eleven *N. rosulatum* sites were tested. Most *N. rosulatum* samples were collected from the north of the known species distribution, however populations of *N. rosulatum* have been found as far south as northern Otago, in the Ida Range (Figure 3.1). Further sampling is required to determine if this diversity is reflected throughout the *N. rosulatum* distribution, or if central Marlborough is the centre of chloroplast haplotype diversity. Studying the distribution of diversity throughout the 'species depauperate zone', otherwise known as the 'beech gap' in the central South Island (WARDLE 1963; MCGLONE 1985; WINKWORTH *et al.* 2005), may give valuable insights on refugia and recolonisation routes for alpine plants during the last glacial. There are few plant genetic studies comparing species that survive throughout the 'zone' as well as to the north and south (LOCKHART *et al.* 2001; HEENAN AND MITCHELL 2003; SMISSEN *et al.* 2003; BECKER *et al.* 2013), so a population genetic study of *N. rosulatum* throughout its range may help discern some of the impacts of glaciation and historic climate change on the distribution of alpine plants.

Interestingly, the nuclear data does not support diversification in Marlborough *N. rosulatum* sites. Several analyses tightly clumped the Marlborough samples, J, M, N, and P, together. Each site has moderate levels of within-site genetic diversity, but far fewer

private alleles than *N. australe* populations. Some geneflow between sites is supported by the low F_{ST} values between J, M, N, and P, although J and N each have small sample sizes. This could be partially due to ascertainment bias, as the nuclear markers were designed from a *N. australe* genome. Otherwise, the high chloroplast diversity, but low nuclear differentiation, could indicate pollen-mediated geneflow. The difference between results from nuclear and chloroplast markers highlights the importance of using information from multiple genomes to detect genetic patterns of geneflow (BAI *et al.* 2014). The mechanism of pollen transfer is unknown in *Notothlaspi*, although Coleoptera (beetles), Hymenoptera (ants), and Lepidoptera (butterflies or moths) were observed visiting fragrant *Notothlaspi* flowers on several occasions (personal observation). The distance these insects travel is unknown, as is the level of connectivity between *N. rosulatum* populations. The distribution of *N. rosulatum* populations is thought to be sparse and patchy (HEENAN 2019), perhaps due to the mobile nature of scree slopes and the biennial lifecycle of *N. rosulatum*. Nevertheless, within these sites is a great deal of within-individual diversity, and F_{IS} scores are close to Hardy-Weinberg equilibrium. This indicates low selfing rates, despite simple structured flowers with no apparent physical barrier to self-pollination.

The only site within the Richmond Forest Park without mixed-species genotypes is *N. rosulatum* location H. The other *N. rosulatum* sampled site in these mountains, G, has a cluster I haplotype common in *N. australe*. Instead, location H shares a chloroplast with site M1 and location P, despite a 69 km distance between H and M1. This makes H an exception to the north/south split of haplotype clusters. We propose this may represent a more recent wave of migration into the Richmond Forest Park by *N. rosulatum.* G and H have a very high pairwise F_{ST} score, indicating elevated genetic divergence between sites, so it seems unlikely they were part of the same wave into this

area, and were subsequently separated, unless active selection against interbreeding occurred. Diversity statistics are lower in H than in other sites with a similar sample size, which could indicate a geologically recent migration.

Remarkably, sites G and Q clump together in the PCOA and Neighbor-Net analyses, despite a 234 km distance between the sites. However, there is very high pairwise F_{ST} values between the two, and both sites also have high pairwise F_{ST} scores with all other sampled sites. These contrary results may represent conflicting patterns in the data. Due to their high mutation rate, and often stepwise mutation mechanism, sequence homoplasy is a common issue with microsatellites. Upon examination of the raw data, at three microsatellite loci (Na8426, Na11999, and Na0011), both Q and G have a high frequency of an allele not found in other *N. rosulatum* sites; therefore, homoplasy seems unlikely. Without further sampling, any explanations of this link are highly speculative, such as a rare long-distance dispersal event, an ancient migration route not intersecting other sampled populations, incomplete lineage sorting, or linkage to similarly conserved loci.

3.6 Summary

This is the first investigation into New Zealand alpine biogeography using a population genetics approach across the upper South Island. There is an east/west split with a potential central zone of admixture in the Richmond Forest Park, where all three *Notothlaspi* species reside. To the west, patterns of variation remain unresolved, but may link to isolation by environment or diverged ancestral groups within *N. australe*.

Hybridisation introduces novel combinations of alleles, that can increase the evolutionary potential of a population; individuals from Dun Mountain (E) and the Red Hills (I) each have higher diversity statistics than other sites. We sampled two *N.*

rosulatum sites within the admixed zone. Individuals from Mt Richmond (G) have a common *N. australe* chloroplast haplotype, and nuclear support for admixture with *N. australe* at Dun Mountain (E). Samples from Inwoods Ridge (H), however, have a common *N. rosulatum* haplotype and shows no evidence of admixture. *N. rosulatum* individuals at Mt Richmond (G) may represent descendants of earlier inhabitants of the Richmond Forest Park, and those from Inwoods Ridge (H), a new wave of migration.

To the east, *N. rosulatum* sites host twice as many chloroplast haplotypes as *N. australe*, segregating between Lake Tennyson (M4), Mt St Patrick (P), and Mt Chisholm (N). Despite diverse chloroplast haplotypes, sites in Marlborough cluster tightly according to nuclear microsatellite data, which could indicate the prevalence of pollen mediated geneflow.

Overall, our results have identified several biogeographic patterns across the species, and raised several interesting questions. Where and when did the three species split? Did *N. australe* populations survive in separate refugia during the Pleistocene glacial cycles? What is the mechanism of geneflow among *Notothlaspi* populations? Further population genetic studies of alpine New Zealand flora may determine if the biogeographic patterns uncovered in this study are genus-specific, or if similar arrangements of genetic variation are found in other groups, driven by similar environmental conditions.

3.7 References

Acosta, M. C., and A. C. Premoli, 2010 Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus,* Nothofagaceae). Molecular Phylogenetics and Evolution 54: 235-242.

Aguirre-Liguori, J. A., B. S. Gaut, J. P. Jaramillo-Correa, M. I. Tenaillon, S. Montes-Hernández *et al.*, 2019 Divergence with gene flow is driven by local adaptation to temperature and soil phosphorus concentration in *teosinte* subspecies (*Zea mays parviglumis* and *Zea mays* mexicana). Molecular Ecology 28: 2814-2830.

Alvarez, N., C. Thiel-Egenter, A. Tribsch, R. Holderegger, S. Manel *et al.*, 2009 History or ecology? Substrate type as a major driver of patial genetic structure in Alpine plants. Ecology Letters 12: 632-640.

Armstrong, T. T., and P. J. De Lange, 2005 Conservation genetics of *Hebe speciosa* (Plantaginaceae) an endangered New Zealand shrub. Botanical Journal of the Linnean Society 149: 229-239.

Augusto, L., D. L. Achat, M. Jonard, D. Vidal and B. Ringeval, 2017 Soil parent material—A major driver of plant nutrient limitations in terrestrial ecosystems. Global Change Biology 23: 3808-3824.

Azani, N., A. Bruneau, M. F. Wojciechowski and S. Zarre, 2019 Miocene climate change as a driving force for multiple origins of annual species in *Astragalus* (Fabaceae, Papilionoideae). Molecular Phylogenetics and Evolution 137: 210-221.

Bai, W.-N., W.-J. Liao and D.-Y. Zhang, 2010 Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. New Phytologist 188: 892-901.

Bai, W.-N., W.-T. Wang and D.-Y. Zhang, 2014 Contrasts between the phylogeographic patterns of chloroplast and nuclear DNA highlight a role for pollen-mediated gene flow in preventing population divergence in an East Asian temperate tree. Molecular Phylogenetics and Evolution 81: 37-48.

Barrett, R. D. H., and D. Schluter, 2008 Adaptation from standing genetic variation. Trends in Ecology & Evolution 23: 38-44.

Becker, M., N. Gruenheit, M. Steel, C. Voelckel, O. Deusch *et al.*, 2013 Hybridization may facilitate in situ survival of endemic species through periods of climate change. Nature Climate Change 3: 1039-1043.

Beever, R., G. Houliston and T. Armstrong, 2013 Geographic patterns of genetic variation in *Cordyline australis*; does fingerprinting recover phenotypic and phenological patterns? New Zealand Journal of Botany 51: 3-12.

Bell, C. J. E., 1973 Mountain Soils and Vegetation in the Owen Range, Nelson. New Zealand Journal of Botany 11: 73-92.

Beu, A. G., and A. R. Edwards, 1984 New Zealand Pleistocene and late Pliocene glacio-eustatic cycles. Palaeogeography, Palaeoclimatology, Palaeoecology 46: 119-142.

Chapuis, M.-P., and A. Estoup, 2006 Microsatellite Null Alleles and Estimation of Population Differentiation. Molecular Biology and Evolution 24: 621-631.

Chen, Y., Y. Liu, X. Fan, W. Li and Y. Liu, 2017 Landscape-Scale Genetic Structure of Wild Rice *Zizania latifolia*: The Roles of Rivers, Mountains and Fragmentation. Frontiers in Ecology and Evolution 5.

Chunco, A. J., 2014 Hybridization in a warmer world. Ecology and Evolution 4: 2019-2031.

Dąbrowski, M., S. Bornelöv, M. Kruczyk, N. Baltzer and J. Komorowski, 2015 'True' null allele detection in microsatellite loci: a comparison of methods, assessment of difficulties and survey of possible improvements. Molecular Ecology Resources 15: 477-488.

Dawson, M. I., and E. J. Beuzenberg, 2000 Contributions to a chromosome atlas of the New Zealand flora—36. Miscellaneous families. New Zealand Journal of Botany 38: 1-23.

Döweler, F., B. S. Case, H. L. Buckley and M. K. F. Bader, 2021 High light-induced photoinhibition is not limiting seedling establishment at abrupt treeline ecotones in New Zealand. Tree Physiology.

Dufresne, F., M. Stift, R. Vergilino and B. K. Mable, 2014 Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. Molecular Ecology 23: 40-69.

Earl, D. A., 2012 STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4: 359-361.

Eyre-Walker, A., and B. S. Gaut, 1997 Correlated rates of synonymous site evolution across plant genomes. Molecular Biology and Evolution 14: 455-460.

Gamache, I., J. P. Jaramillo-Correa, S. Payette and J. Bousquet, 2003 Diverging patterns of mitochondrial and nuclear DNA diversity in subarctic black spruce: imprint of a founder effect associated with postglacial colonization. Molecular Ecology 12: 891-901.

Gao, L., S. Tang, L. Zhuge, M. Nie, Z. Zhu *et al.*, 2012 Spatial genetic structure in natural populations of *Phragmites australis* in a mosaic of saline habitats in the Yellow River Delta, China. Plos One 7.

Gao, Y., B. Ai, H. Kong, M. Kang and H. Huang, 2015 Geographical pattern of isolation and diversification in karst habitat islands: a case study in the *Primulina eburnea* complex. Journal of Biogeography 42: 2131-2144.

Gillespie, J. H., 2000 Genetic Drift in an Infinite Population: The Pseudohitchhiking Model. Genetics 155: 909-919.

Glenny, D., 2004 A revision of the genus *Gentianella* in New Zealand. New Zealand Journal of Botany 42: 361-530.

Haldane, JB S An exact test for randomness of mating. Journal of Genetics 52: 631-635

Halloy, S. R., and A. F. Mark, 2003 Climate-change effects on alpine plant biodiversity: a New Zealand perspective on quantifying the threat. Arctic, Antarctic, and Alpine Research 35: 248-254.

Harrison, R., and E. Larson, 2014 Hybridization, introgression, and the nature of species boundaries. Journal of Heredity 105: 795-809.

Harsch, M. A., P. E. Hulme, M. S. McGlone and R. P. Duncan, 2009 Are treelines advancing? A global meta-analysis of treeline response to climate warming. Ecology Letters 12: 1040-1049.

Heenan, P., 2019 A taxonomic revision of *Notothlaspi* (Brassicaceae), a specialist alpine genus from New Zealand. Phytotaxa 399: 248-260.

Heenan, P., and M. McGlone, 2013 Evolution of New Zealand alpine and open-habitat plant species during the late Cenozoic. New Zealand Journal of Ecology: 105-113.

Heenan, P., and A. Mitchell, 2003 Phylogeny, biogeography and adaptive radiation of *Pachycladon* (Brassicaceae) in the mountains of South Island, New Zealand. Journal of Biogeography 30: 1737-1749.

Heenan, P. B., and M. S. McGlone, 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: 204-226.

Holzapfel, S., M. Z. Faville and C. E. Gemmill, 2002 Genetic variation of the endangered holoparasite *Dactylanthus taylorii* (Balanophoraceae) in New Zealand. Journal of Biogeography 29: 663-676.

Hulshof, C. M., and M. J. Spasojevic, 2020 The edaphic control of plant diversity. Global Ecology and Biogeography 29: 1634-1650.

Huson, D. H., and D. Bryant, 2006 Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23: 254-267.

Jakobsson, M., and N. A. Rosenberg, 2007 CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801-1806.

Jenny, H., 1941 *Factors of soil formation: a system of quantitative pedology*. Courier Corporation.

Jia, D.-R., T.-L. Liu, L.-Y. Wang, D.-W. Zhou and J.-Q. Liu, 2011 Evolutionary history of an alpine shrub *Hippophae tibetana* (Elaeagnaceae): allopatric divergence and regional expansion. Biological Journal of the Linnean Society 102: 37-50.

Joly, S., P. B. Heenan and P. J. Lockhart, 2009a A Pleistocene inter-tribal allopolyploidization event precedes the species radiation of *Pachycladon* (Brassicaceae) in New Zealand. Molecular Phylogenetics and Evolution 51: 365-372.

Joly, S., P. A. McLenachan and Peter J. Lockhart, 2009b A Statistical Approach for Distinguishing Hybridization and Incomplete Lineage Sorting. The American Naturalist 174: E54-E70.

Jombart, T., 2008 adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403-1405.

Jombart, T., S. Devillard and F. Balloux, 2010 Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11: 94.

Kalinowski, S. T., and M. L. Taper, 2006 Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. Conservation Genetics 7: 991-995.

Kamvar, Z. N., J. F. Tabima and N. J. Grünwald, 2019 Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ.

Kaweki, T. J., 1955 Demography of source—sink populations and the evolution of ecological niches. Evolutionary Ecology 9: 38-44

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung *et al.*, 2012 Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649.

Kim, M., M.-L. Cui, P. Cubas, A. Gillies, K. Lee *et al.*, 2008 Regulatory genes control a key morphological and ecological trait transferred between species. Science 322: 1116-1119.

Kimura, M., 1983 *The neutral theory of molecular evolution*. Cambridge University Press.

Kirkpatrick, M., Bartn, N. H., Evolution of a species' range. The American Naturalist 150: 1-23

Kolář, F., T. Fér, M. Štech, P. Trávníček, E. Dušková *et al.*, 2012 Bringing Together Evolution on Serpentine and Polyploidy: Spatiotemporal History of the Diploid-Tetraploid Complex of *Knautia arvensis* (Dipsacaceae). PLOS One 7: e39988.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg and I. Mayrose, 2015 Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15: 1179-1191.

Kumar, S., G. Stecher, M. Li, C. Knyaz and K. Tamura, 2018 MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution 35: 1547.

Lee, W. G., A. Mark and J. Wilson, 1983 Ecotypic differentiation in the ultramafic flora of the South Island, New Zealand. New Zealand Journal of Botany 21: 141-156.

Leigh, J. W., and D. Bryant, 2015 popart: full-feature software for haplotype network construction. Methods in Ecology and Evolution 6: 1110-1116.

Liu, B.-B., C. S. Campbell, D.-Y. Hong and J. Wen, 2020 Phylogenetic relationships and chloroplast capture in the *Amelanchier-Malacomeles-Peraphyllum* clade (Maleae, Rosaceae): Evidence from chloroplast genome and nuclear ribosomal DNA data using genome skimming. Molecular Phylogenetics and Evolution 147: 106784.

Lockhart, P. J., P. A. McLenachan, D. Havell, D. Glenny, D. Huson *et al.*, 2001 Phylogeny, radiation, and transoceanic dispersal of New Zealand alpine buttercups: molecular evidence under split decomposition. Annals of the Missouri Botanical Garden: 458-477.

Loveless, M. D., and J. L. Hamrick, 1984 Ecological determinants of genetic structure in plant populations. Annual Review of Ecology and Systematics 15: 65-95.

Macara, G. R., 2016 The climate and weather of Nelson and Tasman. , pp. 40 in *NIWA Science and Technology Series*.

Maguilla, E., M. Escudero, A. L. Hipp and M. Luceño, 2017 Allopatric speciation despite historical gene flow: Divergence and hybridization in *Carex furva and C. lucennoiberica* (Cyperaceae) inferred from plastid and nuclear RAD-seq data. Molecular Ecology 26: 5646-5662.

Markgraf, V., J. R. Dodson, A. P. Kershaw, M. S. McGlone and N. Nicholls, 1992 Evolution of late Pleistocene and Holocene climates in the circum-South Pacific land areas. Climate Dynamics 6: 193-211.

Martin, S. H., and C. D. Jiggins, 2017 Interpreting the genomic landscape of introgression. Current Opinion in Genetics & Development 47: 69-74.

McCulloch, G. A., G. P. Wallis and J. M. Waters, 2010 Onset of glaciation drove simultaneous vicariant isolation of alpine insects in New Zealand. Evolution: International Journal of Organic Evolution 64: 2033-2043.

McGlone, M. S., 1985 Plant biogeography and the late Cenozoic history of New Zealand. New Zealand Journal of Botany 23: 723-749.

McKown, A. D., R. D. Guy, J. Klápště, A. Geraldes, M. Friedmann *et al.*, 2014 Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *populus trichocarpa*. New Phytologist 201: 1263-1276.

McNeilly, T., and J. Antonovics, 1968 Evolution in closely adjacent plant populations IV. Barriers to gene flow. Heredity 23: 205-218.

Meudt, H. M., P. J. Lockhart and D. Bryant, 2009 Species delimitation and phylogeny of a New Zealand plant species radiation. BMC Evolutionary Biology 9: 1-17.

Meudt, H. M., J. M. Prebble and C. A. Lehnebach, 2015 Native New Zealand forget-me-nots (*Myosotis*, Boraginaceae) comprise a Pleistocene species radiation with very low genetic divergence. Plant Systematics and Evolution 301: 1455-1471.

Millar, T. R., P. B. Heenan, A. D. Wilton, R. D. Smissen and I. Breitwieser, 2017 Spatial distribution of species, genus and phylogenetic endemism in the vascular flora of New Zealand, and implications for conservation. Australian Systematic Botany 30: 134-147.

Mitchell, A. D., and P. B. Heenan, 2000 Systematic relationships of New Zealand endemic Brassicaceae inferred from nrDNA ITS sequence data. Systematic Botany 25: 98-105.

Mitchell, N., G. L. Owens, S. M. Hovick, L. H. Rieseberg and K. D. Whitney, 2019 Hybridization speeds adaptive evolution in an eight-year field experiment. Scientific Reports 9: 6746.

Mix, A. C., E. Bard and R. Schneider, 2001 Environmental processes of the ice age: land, oceans, glaciers (EPILOG). Quaternary Science Reviews 20: 627-657.

Montgomery, S., 1985 Gene Flow in Natural Populations. Annual Review of Ecology and Systematics 16: 393-430.

Nevado, B., N. Contreras-Ortiz, C. Hughes and D. A. Filatov, 2018 Pleistocene glacial cycles drive isolation, gene flow and speciation in the high-elevation Andes. New Phytologist 219: 779-793.

Newnham, R., M. McGlone, N. Moar, J. Wilmshurst and M. Vandergoes, 2013 The vegetation cover of New Zealand at the Last Glacial Maximum. Quaternary Science Reviews 74: 202-214.

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

Peakall, R., and P. E. Smouse, 2006 GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288-295.

Pilkington, K. M., 2014 A population genetics approach to species delimitation in the genus *Selliera* (Goodeniaceae): a thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Plant Biology at Massey University, Palmerston North, New Zealand, pp. Massey University.

Prevosti, A., J. Ocana and G. Alonso, 1975 Distances between populations of *Drosophila subobscura*, based on chromosome arrangement frequencies. Theoretical and Applied Genetics 45: 231-241.

Pritchard, J. K., M. Stephens and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Rahbek, C., M. K. Borregaard, A. Antonelli, R. K. Colwell, B. G. Holt *et al.*, 2019 Building mountain biodiversity: Geological and evolutionary processes. Science 365: 1114-1119.

Rattenbury, M. S., R. A. Cooper and M. R. Johnston, 1998 Geology of the Nelson area: scale 1:250,000 Institute of Geological & Nuclear Sciences 1:250,000 geological map 9 pp.

Rawlence, N. J., B. C. Potter, N. Dussex, L. Scarsbrook, D. A. Orlovich *et al.*, 2021 Plio-Pleistocene environmental changes shape present day phylogeography of New Zealand's southern beeches (Nothofagaceae). New Zealand Journal of Botany 59: 55-71.

Ren, G., R. G. Mateo, J. Liu, T. Suchan, N. Alvarez *et al.*, 2017 Genetic consequences of Quaternary climatic oscillations in the Himalayas: *Primula tibetica* as a case study based on restriction site-associated DNA sequencing. New Phytologist 213: 1500-1512.

Richardson, S. J., P. A. Williams, N. W. H. Mason, R. P. Buxton, S. P. Courtney *et al.*, 2012 Rare species drive local trait diversity in two geographically disjunct examples of a naturally rare alpine ecosystem in New Zealand. Journal of Vegetation Science 23: 626-639.

Rieseberg, L. H., and D. Soltis, 1991 Phylogenetic consequences of cytoplasmic gene flow in plants. American Journal of Botany.

Robinson, B. H., R. R. Brooks, J. H. Kirkman, P. E. Gregg and P. Gremigni, 1996 Plant-available elements in soils and their influence on the vegetation over ultramafic (" serpentine") rocks in New Zealand. Journal of the Royal Society of New Zealand 26: 457-468.

Rogers, G., S. Courtney and P. B. Heenan, 2018 *The calcicolous vascular flora of New Zealand : life forms, taxonomy, biogeography and conservation status*. Creative Services Team, Department of Conservation.

Rosenberg, N. A., 2004 distruct: a program for the graphical display of population structure. Molecular Ecology Notes 4: 137-138.

RStudio-Team, 2020 RStudio: Integrated Development for R., pp., RStudio, PBC, Boston, MA.

Ryan, M., G. Dunbar, M. Vandergoes, H. Neil, M. Hannah *et al.*, 2012 Vegetation and climate in Southern Hemisphere mid-latitudes since 210 ka: new insights from marine and terrestrial pollen records from New Zealand. Quaternary Science Reviews 48: 80-98.

Sambatti, J. B., and K. J. Rice, 2006 Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). Evolution 60: 696-710.

Schuster, T. M., S. D. Setaro, J. F. Tibbits, E. L. Batty, R. M. Fowler *et al.*, 2018 Chloroplast variation is incongruent with classification of the Australian bloodwood eucalypts (genus *Corymbia*, family Myrtaceae). PloS One 13: e0195034.

Seehausen, O., G. Takimoto, D. Roy and J. Jokela, 2008 Speciation reversal and biodiversity dynamics with hybridization in changing environments. Molecular Ecology 17: 30-44.

Segatto, A. L. A., Cazé, C. Turchetto, U. Klahre, C. Kuhlemeier *et al.*, 2014 Nuclear and plastid markers reveal the persistence of genetic identity: A new perspective on the evolutionary history of *Petunia exserta*. Molecular Phylogenetics and Evolution 70: 504-512.

Shepherd, L. D., and T. G. McLay, 2011 Two micro-scale protocols for the isolation of DNA from polysaccharide-rich plant tissue. Journal of Plant Research 124: 311-314.

Shepherd, L. D., and L. R. Perrie, 2011 Microsatellite DNA analyses of a highly disjunct New Zealand tree reveal strong differentiation and imply a formerly more continuous distribution. Molecular Ecology 20: 1389-1400.

Shepherd, L. D., L. R. Perrie and P. J. Brownsey, 2007 Fire and ice: volcanic and glacial impacts on the phylogeography of the New Zealand forest fern *Asplenium hookerianum*. Molecular Ecology 16: 4536-4549.

Smissen, R., I. Breitwieser and J. Ward, 2006 Genetic diversity in the New Zealand endemic species *Helichrysum lanceolatum* (Asteraceae: Gnaphalieae). New Zealand Journal of Botany 44: 237-247.

Smissen, R., I. Breitwieser, J. Ward, P. McLenachan and P. Lockhart, 2003 Use of ISSR profiles and ITS-sequences to study the biogeography of alpine cushion plants in the genus *Raoulia* (Asteraceae). Plant Systematics and Evolution 239: 79-94.

Sork, V. L., F. W. Davis, R. Westfall, A. Flint, M. Ikegami *et al.*, 2010 Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. Molecular Ecology 19: 3806-3823.

Suarez-Gonzalez, A., C. Lexer and Q. C. Cronk, 2018 Adaptive introgression: a plant perspective. Biology Letters 14: 20170688.

Suggate, R. P., 1990 Late pliocene and quaternary glaciations of New Zealand. Quaternary Science Reviews 9: 175-197.

Tamura, K., 1992 Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+ C-content biases. Molecular Biology and Evolution 9: 678-687.

Taylor, S. A., E. L. Larson and R. G. Harrison, 2015 Hybrid zones: windows on climate change. Trends in Ecology & Evolution 30: 398-406.

Templeton, A. R., 2008 The reality and importance of founder speciation in evolution. Bioessays 30: 470-479.

Temunović, M., J. Franjić, Z. Satovic, M. Grgurev, N. Frascaria-Lacoste *et al.*, 2012 Environmental heterogeneity explains the genetic structure of continental and Mediterranean populations of *Fraxinus angustifolia* Vahl. PLOS One 7.

Thórsson, Æ. T., E. Salmela and K. Anamthawat-Jónsson, 2001 Morphological, cytogenetic, and molecular evidence for introgressive hybridization in birch. Journal of Heredity 92: 404-408.

Tomasello, S., K. Karbstein, L. Hodač, C. Paetzold and E. Hörandl, 2020 Phylogenomics unravels Quaternary vicariance and allopatric speciation patterns in temperate-montane plant species: A case study on the *Ranunculus auricomus* species complex. Molecular Ecology 29: 2031-2049.

Tsitrone, A., M. Kirkpatrick and D. A. Levin, 2003 A model for chloroplast capture. Evolution 57: 1776-1782.

Turner, T. L., M. W. Hahn and S. V. Nuzhdin, 2005 Genomic islands of speciation in *Anopheles gambia*. PLoS Biology 3: e285.

van der Niet, T., S. D. Johnson and H. P. Linder, 2006 Macroevolutionary data suggest a role for reinforcement in pollination system shifts. Evolution 60: 1596-1601.

Van Oosterhout, C., W. F. Hutchinson, D. P. Wills and P. Shipley, 2004 MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4: 535-538.

Via, S., 2009 Natural selection in action during speciation. Proceedings of the National Academy of Sciences 106: 9939-9946.

Wagstaff, S. J., and I. Breitwieser, 2004 Phylogeny and classification of *Brachyglottis* (Senecioneae, Asteraceae): an example of a rapid species radiation in New Zealand. Systematic Botany 29: 1003-1010.

Wagstaff, S. J., and P. J. Garnock-Jones, 1998 Evolution and biogeography of the Hebe complex (Scrophulariaceae) inferred from ITS sequences. New Zealand Journal of Botany 36: 425-437.

Walden, N., D. A. German, E. M. Wolf, M. Kiefer, P. Rigault *et al.*, 2020 Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. Nature Communications 11: 3795.

Wan, Q., Z. Zheng, K. Huang, E. Guichoux and R. J. Petit, 2017 Genetic divergence within the monotypic tree genus *Platycarya* (Juglandaceae) and its implications for species' past dynamics in subtropical China. Tree Genetics & Genomes 13: 73.

Wang, I. J., and G. S. Bradburd, 2014 Isolation by environment. Molecular Ecology 23: 5649-5662.

Wang, Y.-H., H. P. Comes, Y.-N. Cao, R. Guo, Y.-R. Mao *et al.*, 2017 Quaternary climate change drives allo-peripatric speciation and refugial divergence in the *Dysosma versipellis-pleiantha* complex from different forest types in China. Scientific Reports 7: 1-13.

Wardle, P., 1963 Evolution and distribution of the New Zealand flora, as affected by Quaternary climates. New Zealand Journal of Botany 1: 3-17.

Weston, K., and B. Robertson, 2015 Population structure within an alpine archipelago: strong signature of past climate change in the New Z ealand rock wren (*Xenicus gilviventris*). Molecular Ecology 24: 4778-4794.

Whitney, K. D., R. A. Randell and L. H. Rieseberg, 2006 Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. The American Naturalist 167: 794-807.

Whitney, K. D., R. A. Randell and L. H. Rieseberg, 2010 Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. New Phytologist 187: 230-239.

Wickham, H., 2011 ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics 3: 180-185.

Williams, P. W., M. McGlone, H. Neil and J.-X. Zhao, 2015 A review of New Zealand palaeoclimate from the Last Interglacial to the global Last Glacial Maximum. Quaternary Science Reviews 110: 92-106.

Winkworth, R. C., S. J. Wagstaff, D. Glenny and P. J. Lockhart, 2005 Evolution of the New Zealand mountain flora: Origins, diversification and dispersal. Organisms Diversity & Evolution 5: 237-247.

Wright, G. A., and F. P. Schiestl, 2009 The evolution of floral scent: the influence of olfactory learning by insect pollinators on the honest signalling of floral rewards. Functional Ecology 23: 841-851.

Wright, S., 1943 Isolation by distance. Genetics 28: 114.

Young, A. G., G. Schmidt-Adam and B. G. Murray, 2001 Genetic variation and structure of remnant stands of pohutukawa (*Metrosideros excelsa*, Myrtaceae). New Zealand Journal of Botany 39: 133-140.

Zhang, X., Y. Sun, J. B. Landis, J. Zhang, L. Yang *et al.*, 2020 Genomic insights into adaptation to heterogeneous environments for the ancient relictual *Circaeaster agrestis* (Circaeasteraceae, Ranunculales). New Phytologist 228: 285-301.

Zhou, Y., L. Duvaux, G. Ren, L. Zhang, O. Savolainen *et al.*, 2017 Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. Heredity 118: 211-220.

Chapter 4

General Discussion

4.1 Introduction

The New Zealand alpine flora is characterised by a series of recent radiations thought to be driven by climatic and tectonic upheaval (WAGSTAFF *et al.* 1999; LOCKHART *et al.* 2001; WAGSTAFF AND BREITWIESER 2004; JOLY *et al.* 2009; HEENAN AND MCGLONE 2013; MEUDT *et al.* 2015). It is therefore an excellent place to study evolutionary processes like local adaptation, divergence, and hybridisation. There are very few micro-evolutionary studies of the New Zealand flora (YOUNG *et al.* 2001; HOLZAPFEL *et al.* 2002; ARMSTRONG AND DE LANGE 2005; SMISSEN *et al.* 2006; SHEPHERD *et al.* 2007; SHEPHERD AND PERRIE 2011; BEEVER *et al.* 2013; PILKINGTON 2014); at this scale, population genetic studies can infer past impacts on species distribution and effective populations sizes, along with present patterns of genetic structure spatially and temporally (VIA 2009). This information can advise the allocation of conservation resources to hotspots of diversity and shed light on the evolutionary processes important in the recent history of New Zealand.

Local adaptation to substrate can be critical to the survival of sessile organisms like plants. The northern third of the South Island of New Zealand has narrow bands of different basement rock types with numerous examples of edaphic endemic flora (LEE *et al.* 1983; RICHARDSON *et al.* 2012; ROGERS *et al.* 2018). Intraspecific study of plant groups growing on numerous different rock types may provide insight into the development of

locally adapted populations. Research on spatial patterns of diversity would help detect if local adaptation to these soil types may be driving divergence.

Notothlaspi australe populations occur across a wide range of soil types, including soils derived from ultramafic, calcareous, volcanic, and granitic rock. Its distribution is limited to the northern South Island of New Zealand, where the herb mostly grows above the tree line in rock ledges, fellfields, and bare ridgelines. *Notothlaspi viretum* and *N. rosulatum* have an overlapping range with *N. australe* in the Richmond Ranges and Nelson Lakes regions. These three congeners are the only known species in the endemic tribe of Notothlaspideae. Here, a population genetics study of *N. australe* was originally designed to investigate potential patterns of divergence that may be indicative of local adaptation. Given the overlapping distributions of its congeners and the potential for hybridization, sampling included all three *Notothlaspi* species. The objectives of this study were to:

1. Develop intra-specific and intra-generic markers suitable for *Notothlaspi,*
2. Determine the distribution of genetic variation across the alpine landscape within and among *Notothlaspi* species,
3. Infer historical influences on the adaptive and evolutionary history of *Notothlaspi.*

4.2 Findings

Twelve microsatellite markers were designed, and the trnL-F locus was trialled for use in *Notothlaspi* to characterise the distribution of genetic variation. The main findings are summarised in previous chapters: 'Marker development'; and 'Using Population

Genetics to investigate the Spatial Structure of Genetic Variation in *Notothlaspi'*. This section will summarise the results of the study:

4.2.1 Objective 1:

Three marker systems were trialled for utility across the three *Notothlaspi* species: nuclear microsatellite markers; five universal chloroplast markers; and a RAPDseq scheme to identify single nucleotide polymorphisms. Primers for thirty microsatellite loci were designed from *N. australe* sequences from an Illumina HiSeq run by researchers at the Centre for Organismal Studies, University of Heidelberg (WALDEN *et al.* 2020). Of these, twelve loci were selected as they mainly amplified a maximum of two alleles per individual, are polymorphic, and easily scorable. Due to high concentrations of PCR product, they could be multiplexed six at a time, which considerably reduced genotyping costs. There was a greater proportion of missing data in *N. rosulatum,* and these markers did not behave as simple diploid codominant loci in *N. viretum* due to multiploid allele amplification. Of the five universal chloroplast markers, only the trnL-F and psbA-trnH loci produced clean, usable sequences. Sequences from both markers could only be read in a single direction due to homopolymers within the sequences. Using the RAPDseq scheme, 41 loci were identified that revealed polymorphisms within and between species and are present in a suitable number of individuals. Unfortunately, results for the RAPDseq experiment arrived too late to develop the method beyond the initial trialling stage.

4.2.2 Objectives 2 and 3:

The twelve microsatellite markers designed in chapter 2 and the trnL-F chloroplast locus, were utilised to undertake a population genetics study of *Notothlaspi*. Within *N. australe*, two geographically segregate genetic groups were identified. One in the upper

and mid Kahurangi National Park, the other comprising of southern Kahurangi National Park sites and those in the Nelson Lakes and Richmond Forest Park. The two genetic groups meet north of Mt Owen, where the genetic distance between individuals from Biggs tops and Mt Owen is much greater than expectations based on geographic distance. The causes for the split remain unexplained, however, two possibilities are isolation by environment in the lower Kahurangi National Park, and diverged ancestral groupings. Interestingly, both groups share a chloroplast haplotype. Only two *N. australe* groups, I and E3, do not have the common haplotype Ia.

Six chloroplast haplotypes were identified within *N. rosulatum*, although two are likely sourced from admixture with *N. australe.* These haplotypes readily distinguish populations in central Marlborough, however nuclear microsatellite results find little differentiation between these populations. This could indicate a prevalence of pollen mediated geneflow across valleys. An interesting but unexplained finding was genetic similarity between the southernmost sampled site in the Craigieburn Range with the northernmost site on Mt Richmond in the Richmond Forest Park. Overall, nuclear genetic data showed *N. rosulatum* sites as less diverse than in *N. australe*.

Across the broader landscape of the upper South Island, a zone of admixture has been uncovered where the three *Notothlaspi* species ranges overlap in the Richmond Forest Park. All but one sampled site in these mountains has a mixture of genotypes from two or more species. These sites have higher diversity statistics and diverged genotypes from others of the same species, which may be attributed to variation introduced through hybridisation. The two sampled *N. rosulatum* locations in the zone are strongly diverged from one another at both chloroplast and nuclear loci. They may represent two waves of migration into the Richmond Forest Park. Most *N. viretum* individuals sampled share a

chloroplast haplotype with *N. australe* from nearby in the Red Hills, this may be through recent gene flow, a recent split between the species, or incomplete lineage sorting.

4.3 Limitations

Several factors have imposed limitations on this study. One of the primary issues is sample size, particularly within *N. rosulatum*. At Mt Altimarloch only two individuals were found, and only four near Mt Chisholm. These four individuals may be outliers in the surrounding area, skewing results. Both sites have unique chloroplast haplotypes, an interesting result that is cannot be verified without further sampling. Other *N. rosulatum* sites also have sample sizes below 20, which was the targeted sample size for microsatellite analysis. This is partially due to poor amplification of some *N. rosulatum* sites for specific loci, leading to several individuals being removed from the analysis.

Poor amplification for some loci in *N. rosulatum* led to a high degree of missing data (~13%), but it was less of an issue in *N. australe* (~4.5% missing data). Missing data can lead to overestimates of homozygosity, $F_{ST,}$ and admixture, however it is relatively common in microsatellite studies (CARLSSON 2008; REEVES *et al.* 2016). Such missing data can be due to several factors, including human error, null alleles, and poor-quality template DNA. In *N. rosulatum*, specific loci failed to amplify for nearly all individuals in sites G, H, and Q. This could indicate ascertainment bias, due to genetic distance from the *N. australe* individual from which the markers were designed. Some data was coded as missing when there were more than two alleles per locus. This was particularly severe for locus Na0043 and site L. In pairwise F_{ST} estimates, these alleles were accounted for using the ENA method (CHAPUIS AND ESTOUP 2007), however not in other analyses.

The presence of more than two alleles per individual at microsatellite loci excluded *N. viretum* samples from nuclear analysis. This limits findings on interspecific relationships, as only chloroplast data from a single locus could be used for this species. *Notothlaspi australe* site I6 has an unusually high number of private alleles, so there is a chance some of these are sourced from hybridization with the nearby *N. viretum* population. *Notothlaspi viretum* nuclear data would help put this variation within Red Hills (location I) *N. australe* into context.

4.4 Future Work

This research has uncovered more questions than answers. Nevertheless, it provides a good biogeographic platform for future plant population genetic studies in the alpine regions of the South Island. To verify and strengthen results, further sampling across the range of all three species would be useful. Within *N. australe*, sampling is limited in the Kahurangi national park. Sampling around location D would provide resolution on the proposed biogeographic boundary between D1 and D5. Specifically, *N. australe* have been sighted on the Matiri range, Mt Patriarch, the Allen Range, and the Radiant Range (AVH 2021. The Australasian Virtual Herbarium, Council of Heads of Australasian Herbaria, https://avh.chah.org.au, accessed 4th August 2021). However, the records from the Allen and Radiant Ranges are as old as 1947, and the sites may be challenging to access without a helicopter. In *N. rosulatum*, small sample sizes from near Mt Chisholm and Mt Altimarloch have led to inconclusive results. These sites are easily accessible when there is no snow, therefore more samples could be collected to verify the observations reported here. Greater sampling in the Richmond Ranges would also increase resolution in the hybrid zone. Most sampled *N. rosulatum* sites are from

Marlborough, however *N. rosulatum* populations are found as far south as northern Otago. A population genetics study of *N. rosulatum* throughout its range could provide insights on key biogeographic questions that have puzzled New Zealand naturalists for decades: particularly, how and where plant lineages survived during quaternary climate fluctuations (WARDLE 1963; RAVEN 1973; MOAR AND SUGGATE 1979; MILDENHALL 1980; MCGLONE 1985; WARDLE 1985; HEADS 1998; MCGLONE *et al.* 2001; NEWNHAM *et al.* 2013).

Present-day sampling of *N. viretum* would be beneficial to extend this study. The *N. viretum* samples used here are herbarium specimens from 1980-1997, yet a consistent sampling method across all three species from a similar generation would give greater comparative power. Individuals from both *N. australe* and *N. viretum* at the *N. viretum* site could be sampled to investigate potential hybridization between the two species. While nuclear data from *N. viretum* was not used in this research, it is important to consider this species in studies of the evolution of *Notothlaspi*. This rare edaphic-endemic species is not immediately under threat from human habitat modification, however hybridisation with sympatric *N. australe* may shape the future of the group (HEENAN 2019). *N. australe* collected in the Red Hills and most sampled *N. viretum* individuals shared haplotype Ic, but the direction of the proposed chloroplast capture is unconfirmed. Incomplete lineage sorting, or a recent split between the species could also explain this result. Nuclear data may help elucidate the relationship further. Treating the microsatellite data as presence/absence (DUFRESNE *et al.* 2014), or using a different marker system, such as RAPDseq, could allow for a direct comparison between species.

Chromosome counts for both *N. australe* and *N. viretum* come from the Red Hills, and both were placed within the 2n=90-100 range (DAWSON AND BEUZENBERG 2000). The consistent multiplication in number of microsatellite alleles within *N. viretum*

compared with *N. australe,* including those from the Red Hills, may indicate a higher ploidy for *N. viretum.* Chromosome counts for *N. australe* individuals from other parts of the range would be valuable information for any future studies on the genus. Some individuals from both *N. australe* and *N. rosulatum* had to be removed from analysis due to microsatellite results that revealed more than two alleles per locus. Although this, in itself, is not proof of higher ploidy individuals.

Beyond the *Notothlaspi* genus, investigating other species with similar ranges to *Notothlaspi* in the upper South Island alpine could determine if the observed patterns are present in other groups. If this was the case, it would be valuable information for conservation efforts, allowing resources to be focused on areas with greater genetic variation. In the field, multiple plant groups were observed alongside *Notothlaspi* populations, such as *Leucogenes grandiceps* (Hook.f.) Beauverd. (Compositae), *Gaultheria depressa* Hook.f. (Ericaceae), *Gentianella* Moench. (Gentianaceae), *Chionochloa* Zotov. (Gramineae), and *Epilobium* L. (Onagraceae). Similar to *N. australe*, some of these species, particularly in the genus *Gentianella*, also grow on a wide range of substrates (GLENNY 2004). This would be an excellent study genus to further investigate the impact of soil types on the evolution of New Zealand alpine flora, as various species are already believed to be edaphic endemics, while others are found nearby across numerous substrates. Studies could compare sister species to uncover the impact of soil type on diversification in the New Zealand flora, as has begun with edaphic endemics overseas (ANACKER AND STRAUSS 2014; BABST-KOSTECKA *et al.* 2014; SIANTA AND KAY 2019; NAGASAWA *et al.* 2020).

4.5 References

Anacker, B. L., and S. Y. Strauss, 2014 The geography and ecology of plant speciation: range overlap and niche divergence in sister species. Proceedings of the Royal Society B: Biological Sciences 281: 20132980.

Armstrong, T. T., and P. J. De Lange, 2005 Conservation genetics of *Hebe speciosa* (Plantaginaceae) an endangered New Zealand shrub. Botanical Journal of the Linnean Society 149: 229-239.

Babst-Kostecka, A. A., C. Parisod, C. Godé, P. Vollenweider and M. Pauwels, 2014 Patterns of genetic divergence among populations of the pseudometallophyte *Biscutella laevigata* from southern Poland. Plant and Soil 383: 245-256.

Beever, R., G. Houliston and T. Armstrong, 2013 Geographic patterns of genetic variation in *Cordyline australis;* does fingerprinting recover phenotypic and phenological patterns? New Zealand Journal of Botany 51: 3-12.

Carlsson, J., 2008 Effects of Microsatellite Null Alleles on Assignment Testing. Journal of Heredity 99: 616-623.

Chapuis, M.-P., and A. Estoup, 2007 Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution 24: 621-631.

Dawson, M. I., and E. J. Beuzenberg, 2000 Contributions to a chromosome atlas of the New Zealand flora—36. Miscellaneous families. New Zealand Journal of Botany 38: 1-23.

Dufresne, F., M. Stift, R. Vergilino and B. K. Mable, 2014 Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. Molecular Ecology 23: 40-69.

Glenny, D., 2004 A revision of the genus *Gentianella* in New Zealand. New Zealand Journal of Botany 42: 361-530.

Heads, M., 1998 Biogeographic disjunction along the Alpine fault, New Zealand. Biological Journal of the Linnean Society 63: 161-176.

Heenan, P., 2019 A taxonomic revision of *Notothlaspi* (Brassicaceae), a specialist alpine genus from New Zealand. Phytotaxa 399: 248-260.

Heenan, P., and M. McGlone, 2013 Evolution of New Zealand alpine and open-habitat plant species during the late Cenozoic. New Zealand Journal of Ecology: 105-113.

Holzapfel, S., M. Z. Faville and C. E. Gemmill, 2002 Genetic variation of the endangered holoparasite *Dactylanthus taylorii* (Balanophoraceae) in New Zealand. Journal of Biogeography 29: 663-676.

Joly, S., P. B. Heenan and P. J. Lockhart, 2009 A Pleistocene inter-tribal allopolyploidization event precedes the species radiation of *Pachycladon* (Brassicaceae) in New Zealand. Molecular Phylogenetics and Evolution 51: 365-372.

Lee, W. G., A. Mark and J. Wilson, 1983 Ecotypic differentiation in the ultramafic flora of the South Island, New Zealand. New Zealand Journal of Botany 21: 141-156.

Lockhart, P. J., P. A. McLenachan, D. Havell, D. Glenny, D. Huson *et al.*, 2001 Phylogeny, radiation, and transoceanic dispersal of New Zealand alpine buttercups: molecular evidence under split decomposition. Annals of the Missouri Botanical Garden: 458-477.

McGlone, M., R. Duncan and P. Heenan, 2001 Endemism, species selection and the origin and distribution of the vascular plant flora of New Zealand. Journal of Biogeography 28: 199-216.

McGlone, M. S., 1985 Plant biogeography and the late Cenozoic history of New Zealand. New Zealand Journal of Botany 23: 723-749.

Meudt, H. M., J. M. Prebble and C. A. Lehnebach, 2015 Native New Zealand forget-me-nots (*Myosotis*, Boraginaceae) comprise a Pleistocene species radiation with very low genetic divergence. Plant Systematics and Evolution 301: 1455-1471.

Mildenhall, D. C., 1980 New Zealand late Cretaceous and cenozoic plant biogeography: A contribution. Palaeogeography, Palaeoclimatology, Palaeoecology 31: 197-233.

Moar, N. T., and R. P. Suggate, 1979 Contributions to the Quaternary history of the New Zealand flora 8. Interglacial and glacial vegetation in the Westport District, South Island. New Zealand Journal of Botany 17: 361-387.

Nagasawa, K., H. Setoguchi, M. Maki, H. Goto, K. Fukushima *et al.*, 2020 Genetic consequences of plant edaphic specialization to solfatara fields: Phylogenetic and population genetic analysis of *Carex angustisquama* (Cyperaceae). Molecular Ecology 29: 3234-3247.

Newnham, R., M. McGlone, N. Moar, J. Wilmshurst and M. Vandergoes, 2013 The vegetation cover of New Zealand at the Last Glacial Maximum. Quaternary Science Reviews 74: 202-214.

Pilkington, K. M., 2014 A population genetics approach to species delimitation in the genus *Selliera* (Goodeniaceae): a thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Plant Biology at Massey University, Palmerston North, New Zealand, pp. Massey University.

Raven, P. H., 1973 Evolution of subalpine and alpine plant groups in New Zealand. New Zealand Journal of Botany 11: 177-200.

Reeves, P. A., C. L. Bowker, C. E. Fettig, L. R. Tembrock and C. M. Richards, 2016 Effect of error and missing data on population structure inference using microsatellite data. 080630.

Richardson, S. J., P. A. Williams, N. W. H. Mason, R. P. Buxton, S. P. Courtney *et al.*, 2012 Rare species drive local trait diversity in two geographically disjunct examples of a naturally rare alpine ecosystem in New Zealand. Journal of Vegetation Science 23: 626-639.

Rogers, G., S. Courtney and P. B. Heenan, 2018 *The calcicolous vascular flora of New Zealand : life forms, taxonomy, biogeography and conservation status*. Creative Services Team, Department of Conservation.

Shepherd, L. D., and L. R. Perrie, 2011 Microsatellite DNA analyses of a highly disjunct New Zealand tree reveal strong differentiation and imply a formerly more continuous distribution. Molecular Ecology 20: 1389-1400.

Shepherd, L. D., L. R. Perrie and P. Brownsey, 2007 Fire and ice: volcanic and glacial impacts on the phylogeography of the New Zealand forest fern *Asplenium hookerianum*. Molecular Ecology 16: 4536-4549.

Sianta, S. A., and K. M. Kay, 2019 Adaptation and divergence in edaphic specialists and generalists: serpentine soil endemics in the California flora occur in barer serpentine habitats with lower soil calcium levels than serpentine tolerators. American Journal of Botany 106: 690-703.

Smissen, R. D., I. Breitwieser and J. M. Ward, 2006 Genetic diversity in the New Zealand endemic species *Helichrysum lanceolatum* (Asteraceae: Gnaphalieae). New Zealand Journal of Botany 44: 237-247.

Via, S., 2009 Natural selection in action during speciation. Proceedings of the National Academy of Sciences 106: 9939-9946.

Wagstaff, S. J., and I. Breitwieser, 2004 Phylogeny and classification of *Brachyglottis* (Senecioneae, Asteraceae): an example of a rapid species radiation in New Zealand. Systematic Botany 29: 1003-1010.

Wagstaff, S. J., P. B. Heenan and M. J. Sanderson, 1999 Classification, origins, and patterns of diversification in New Zealand *Carmichaelinae* (Fabaceae). American Journal of Botany 86: 1346-1356.

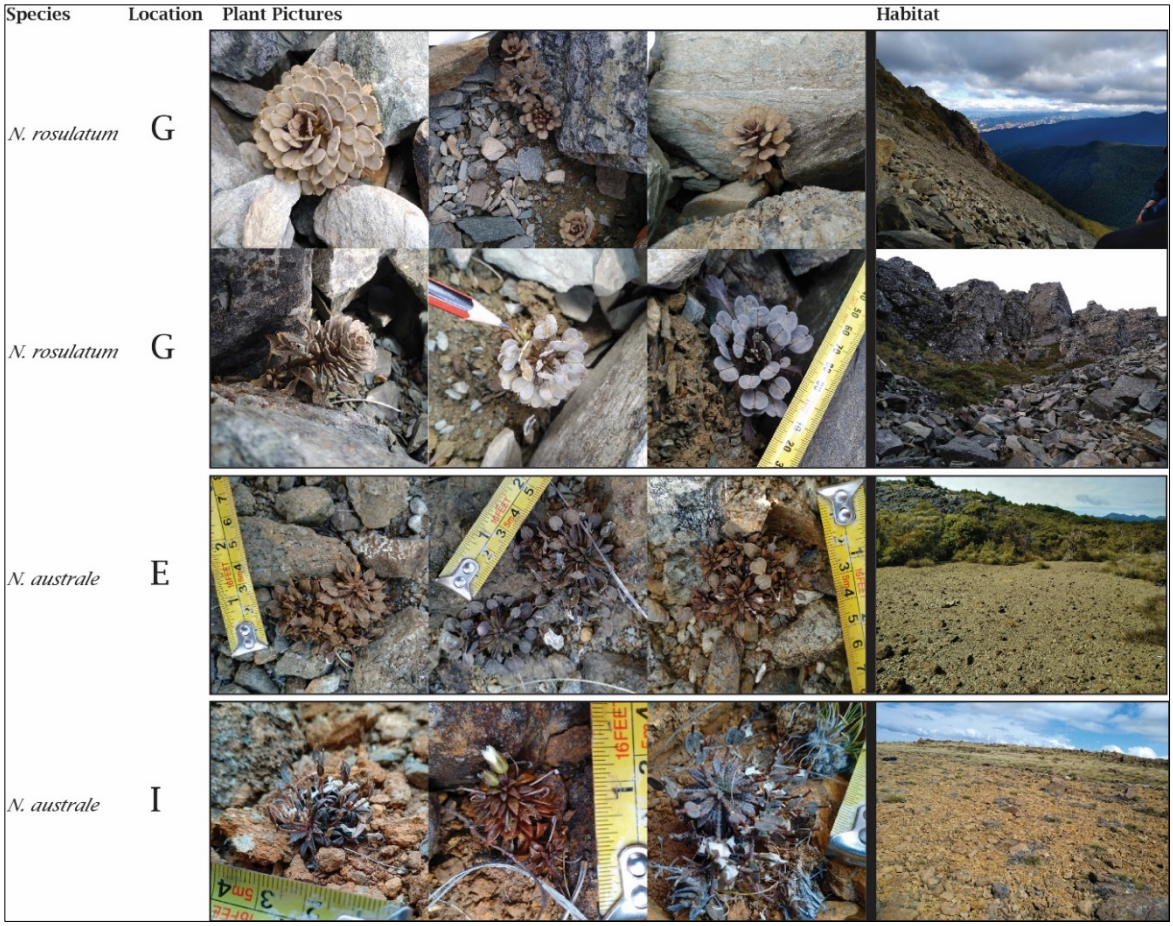
Walden, N., D. A. German, E. M. Wolf, M. Kiefer, P. Rigault *et al.*, 2020 Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. Nature Communications 11: 3795.

Wardle, P., 1963 Evolution and distribution of the New Zealand flora, as affected by Quaternary climates. New Zealand Journal of Botany 1: 3-17.

Wardle, P., 1985 Environmental influences on the vegetation of New Zealand. New Zealand Journal of Botany 23: 773-788.

Young, A. G., G. Schmidt-Adam and B. G. Murray, 2001 Genetic variation and structure of remnant stands of pohutukawa (*Metrosideros excelsa*, Myrtaceae). New Zealand Journal of Botany 39: 133-140.

Supplementary Information



Supplementary Figure 1: Morphology of *N. rosulatum* at site G compared with that of *N. australe* from the Richmond Ranges. A picture of the general environment at each of the sites is shown.

Supplementary Figure 2: Morphology of *N. australe* in the open and in shaded spots

Supplementary Table 1: Geographic coordinates for each site, and MPN code from the Allan Herbarium

Location	Site	Latitude	Longitude	Altitude	MPN
Cobb Valley (north)	B2	-41.102013	172.610314	1698	
	B3	-41.101932	172.609711	1715	
	B4	-41.099037	172.607299	1665	
	B5	-41.098871	172.606947	1664	
	B6	-41.096063	172.606212	1619	
	B7	-41.087502	172.59174	1439	
Cobb Valley (south)	B8	-41.151796	172.6027105	1543	
	B9	-41.1493995	172.5975003	1536	
	B10	-41.14775567	172.5963933	1585	
	B11	-41.14585683	172.5939948	1602	
	B12	-41.147201	172.5955333	1598	
	B13	-41.14471817	172.5938977	1615	
	B14	-41.13841067	172.5838155	1532	
	B15	-41.13676133	172.5806835	1558	52407
	B16	-41.1302055	172.582675	1474	
Cobb Valley (West)	B17	-41.0478495	172.4992443	1637	52398
Mt Owen	D1	-41.54519971	172.5445129	1588	52411
Lookout Range	D4	-41.53738542	172.592016	1542	
Biggs Tops	D5	-41.41373241	172.4154311	1400	
Dun Mountain	E1	-41.36636183	173.3537805	840	
	E2	-41.36776317	173.3410362	884	52399
	E3	-41.34957667	173.3657617	966	
Mt Richmond	G1	-41.47538515	173.389187	1466	52404
	G2	-41.48053033	173.3932162	1391	
	G3	-41.48071849	173.3936347	1407	
	G4	-41.48020132	173.3948562	1499	

Location	Site	Latitude	Longitude	Altitude	MPN
Inwoods Lookout	H1	-41.591145	172.959002	1384	52397
	H2	-41.594223	172.961279	1500.00	
Red Hills	I1	-41.723341	173.004821	1131	52406
	I2	-41.71041817	173.0187827	1168	
	I3	-41.704958	173.0328405	1279	
	I4	-41.6979245	173.039496	1363	
	I4	-41.69031883	173.0437092	1411	
	I6	-41.67869183	173.0614747	1469	
	I7	-41.67920883	173.059274	1507	
	I8	-41.6822705	173.0561165	1515	52402
	I10	-41.685954	173.0521718	1492	
	I11	-41.68673667	173.0500602	1466	
	I12	-41.69090517	173.0446403	1422	
Mt Altimarloch	J1	-41.75782268	173.7784575	1564	
Arnaud Ridge	L1	-41.825772	172.885303	1718	
	L2	-41.829825	172.884338	1670	
	L3	-41.831715	172.883995	1699	
Robert Ridge	L4	-41.843375	172.797181	1588	52403
	L5	-41.84305	172.79916	1591	
	L6	-41.847279	172.79151	1615	
Rag n Famish Gully	M1	-42.1472210	172.85650	1178	52401
	M2	-42.1459700	172.8567660	1180	
Island Pass	M3	-42.171735	172.794412	1420	
Lake Tennyson	M4	-42.195551	172.733452	1160	
Near Mt Chisholm	N1	-42.03825038	173.279573	1500	
Mt St Patrick	P1	-42.437102	172.730411	1568	
	P2	-42.436544	172.730002	1580	52400

Location	Site	Latitude	Longitude	Altitude	MPN
	P3	-42.435667	172.7294	1547	
	P4	-42.437135	172.730908	1562	
Cheeseman Ski Field	Q1	-43.160164	171.672802	1451	