Investigating relationships in the New Zealand alpine *Ranunculus* using RNA sequencing

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

John Curteis Henry

2021

Abstract

Whether the alpine flora of New Zealand is resilient enough to withstand the effects of climate change is an important and unanswered question. Conspicuous amongst the alpine flora are the species of *Ranunculus* in section *Pseudadonis*. This monophyletic group of species is hypothesised to have rapidly diversified into distinct mountain habitats, with some species convergently evolving into similar habitats. Investigating how cryptic physiologies have convergently evolved in some *Ranunculus* species may provide insight into the adaptive potential of this group of plants. It has been argued that hybridisation is an important evolutionary process explaining the morphological and ecological variation of New Zealand alpine *Ranunculus* species. Hybridisation, and in particular introgression, has also been hypothesised elsewhere as an effective means for closely-related species to share genetic material and undergo rapid adaptation through selection of standing genetic variation. This research aimed to use RNA sequencing technology to address questions of physiology and phylogeny amongst four taxa of the alpine *Ranunculus* group. Habitat characterisation was carried out before plants were sampled and grown under standardised conditions in a common garden experiment. Bioinformatic approaches were used to analyse high-throughput sequencing data of RNA extracted from these laboratory-grown plants. This research illustrates the potential of RNA sequencing for studying non-model plant species. However, the conservative analytical approaches adopted, and noise within the data, limited inferences of physiological traits and evolutionary relationships. Analyses of heterozygosity and issues with *de novo* transcriptome assembly suggested greater numbers of gene variants than expected for these small, isolated populations of alpine plants. These gene variants likely occur because of the polyploid genomes of the New Zealand alpine *Ranunculus*. Further work is needed, however, to confirm this genetic diversity. Overall, this work reinforces the difficulties in studying non-model polyploid systems. Yet, it does hint at a genetic richness within the alpine *Ranunculus* that might aid survival of this clade during a rapidly changing future.

Acknowledgements

Many people have contributed to bringing this project to fruition. I am certain to forget and omit some wonderful people, but all the assistance I have received is greatly appreciated.

Firstly, I wish to express my gratitude to my research supervisor, Professor Peter Lockhart, for his encouragement and feedback throughout this project. I would also like to thank Dr Trish McLenachan for her valuable assistance in the laboratory, Emeritus Professor Anthony Larkum for his expertise with PAM, Dr James Hanley for taking time to enable the soil analyses, and Greg Dawson for his investigation into climate data layers.

So many Massey SFS staff have made this possible, but here are some much-needed acknowledgements: Deb, Ann, and Cynthia from finance and administration; Steve, from the engineering department; Pete and Keith in the electronics department.

Many thanks to Dini, at NeSI, for always being on hand to solve my HPC issues, Simon Zhang from Custom Science Ltd. for organising shipping of samples and sequencing, and Wouna for her valuable writing input.

Of course, I cannot go past thanking all the other denizens of the Boffin Lounge and surrounds: India, Tina, Sophie, Nikolai, Demet, Kelli, Amy, Annie, Dr Tina, and A/Prof. Patrick Biggs. A lot of laughs, food, and bouncing-off of ideas.

This project was generously supported by Government funding from the Marsden Fund Council, managed by Royal Society Te Apārangi.

Finally, thank you to Ruth, for everything.

Contents

Abbreviations

%GC	percentage of cytosine and guanine bases in a DNA sequence
Al	aluminium
BAM	binary sequence alignment/map format
BCF	binary variant call format
BD	bulk density of soil
BLAST	basic local alignment search tool
BUSCO	benchmarking universal single-copy orthologue
C	carbon
Ca	calcium
cDNA	complementary DNA
CEC	cation exchange capacity
CI	confidence interval
cm	centimetre/s
cpDNA	chloroplast DNA
DEG	differentially expressed gene
DGE	differential gene expression
DNA	deoxyribonucleic acid
dTTP	deoxythymidine triphosphate
dUTP	deoxyuridine triphosphate
Fv/Fm	maximum quantum yield of photosystem II
g	gram/s
g	gravitational force equivalents
GB	gigabyte/s
GBS	genotype by sequencing
GO	gene ontology
GPS	global positioning system
GWAS	genome-wide association studies
GWC	gravimetric water content
H	hydrogen
ICP-OES	inductively coupled plasma optical emission spectrometry
ID	identifier
ITS	nuclear ribosomal internal transcribed spacer
IUPAC	International Union of Pure and Applied Chemistry
K	potassium

L	litre/s
LENZ	Land Environments of New Zealand database
LOESS	locally weighted scatterplot smoothing
m	metre/s
M	mean
µE	micro Einstein/s
min	minute/s
Mg	magnesium
mg	milligram/s
mL	millilitre/s
ML	maximum likelihood
µL	microlitre/s
mm	millimetre/s
µM	micromole/s L-1
mRNA	messenger RNA
MSA	multiple sequence alignment
N	nitrogen
N50	the length of the shortest contig in a set of the longest contigs containing 50% of all assembled bases
Na	sodium
NCBI	National Center for Biotechnology Information
NeSI	New Zealand eScience Infrastructure
ng	nanogram/s
NGS	next generation sequencing
nm	nanometre/s
nt	nucleotides
OM	organic matter
ONT	Oxford Nanopore Technologies
P	phosphorous
PacBio	Pacific Biosciences
PAM	pulse amplitude modulated fluorometry
PAR	photosynthetically active radiation
PC	principal coordinate
PCA	principal coordinate analysis
PCR	polymerase chain reaction
PSII	photosystem II
QTL	quantitative trait locus

RNA	ribonucleic acid
RNA-seq	RNA sequencing
rpm	revolutions per minute
RQS	RNA quality score
s	second/s
SCO	single-copy orthologous gene
SD	standard deviation
SE	standard error
SNP	single nucleotide polymorphism
SSR	simple sequence repeat
TBS	total base saturation
TPM	transcripts per million
UV	ultraviolet
v/v	volume/volume
VCF	variant call format
VCSN	Virtual Climate Station Network
VMC	volumetric moisture content
WHC	water holding capacity of soil
ycf1	chloroplast *ycf1* gene

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Adaptive radiation of New Zealand alpine plants

Adaptive radiation is the phenomenon of rapid phenotypic and ecological diversification in response to biotic and/or abiotic pressures (Schluter 2000; Hodges and Derieg 2009; Lai *et al.* 2019). It is thought to underlie morphological and ecological diversity of the New Zealand alpine flora (Cockayne 1910; Winkworth *et al.* 1999; Mummenhoff *et al.* 2004), including the biology of the New Zealand alpine *Ranunculus* species (Fisher 1965; Lockhart *et al.* 2001) that are the subject of this thesis. In his critical appraisal of adaptive radiation, Schluter (2000) identified four important criteria against which a hypothesis of adaptive radiation can be tested:

(i) The radiating species must all share common ancestry.

(ii) A correlation must exist between phenotypic traits and environment.

(iii) Traits associated with an environment must have demonstrated utility in that environment.

(iv) Speciation has been rapid.

Hodges and Derieg (2009) have noted that the introduction of molecular phylogenetics has provided a wealth of evidence for ancestry and speciation rates. However, in practice it has been difficult to obtain evidence for (ii) and (iii), mainly because important phenotypes may involve cryptic physiologies (Becker *et al.* 2013; Joly *et al.* 2014) and experiments to demonstrate trait utility are difficult to perform (Voelckel *et al.* 2017; Suarez-Gonzalez *et al.* 2018). This situation might explain why, despite its hypothesised importance (Schluter 2000; Hodges and Derieg 2009; Lai *et al.* 2019; Nevado *et al.* 2019), there are relatively few established examples of adaptive radiation, or more generally adaptive diversification. Recognising this, Joly *et al.* (2014) used an alternative approach to test for adaptive radiation. These authors sought to identify unique predictions of adaptive radiation, and whether New Zealand alpine *Pachycladon* (Brassicaeae) represented an example of adaptive radiation. Specifically, they looked for evidence of rapid phenotypic evolution and ecological diversification in the form of niche shifts. A practical limitation of the approach adopted by Joly *et al*. (2014) is that it requires the distribution of species to be clearly determined. However, understanding the factors underlying an organism's distribution can be complex (O'Brien *et al.* 2017), and defining species' range limits is not a trivial exercise (Sexton *et al.* 2009), with high resolution data needed to accurately characterise habitat of taxa which may have complex ecological requirements (Franklin *et al.* 2013).

An alternative to looking for phenotype-environment associations is to investigate genotype-environment associations. Population genetics theory has provided a means for DNA sequence analyses that provide evidence for (ii) and (iii). Specifically, the neutral theory of molecular evolution developed by Kimura (1968) has contributed to the development of tests for positive selection and provided evidence for adaptation based on analyses of DNA sequence data (Goodswen *et al.* 2018; Lai *et al.* 2019; Derbyshire 2020). The neutral theory suggests that most changes in DNA sequences confer no selective advantage or disadvantage for individuals. A prediction of this hypothesis is that most changes to sequences, and fluctuations of allele frequencies, for most genes are neutral. Thus, the neutral theory provides a null hypothesis against which adaptive divergence of sequences can be distinguished from non-adaptive divergence and also a null hypothesis against which changes in allele frequencies due to selection can be distinguished from genetic drift (Raeymaekers *et al.* 2017; Schrieber *et al.* 2017). The association of specific allelic variants with specific environments is relevant for Schluter's criterion (ii) because it represents a genotype-environment association. The elevated frequency or fixation of specific allelic variants, that cannot be explained by genetic drift, is relevant to Schluter's criterion (iii) since it is best explained by the greater reproductive success of individuals with a particular genotype.

1.2 Climate change and the New Zealand alpine flora

An important question being asked by many researchers is whether evolutionary change of plants, involving genetic adaptation (Section 1.3.1) and plastic responses (Section 1.3.2) to environmental stimuli, will be sufficient to enable the future survival of species given anticipated climate change scenarios (Jump and Penuelas 2005; Jezkova and Wiens 2016; Román-Palacios and Wiens 2020). Climate change is likely to heavily impact the New Zealand alpine environments. Global greenhouse gas emissions are expected to result in higher average temperatures with commensurate reductions in accumulated alpine snowfall (Ministry for the Environment 2018). Additionally, weather events are predicted to become more extreme, with wet regions experiencing greater rainfall and dry areas becoming more drought prone (Ministry for the Environment 2018). Climate modelling indicates that in the future large areas of alpine habitat on the eastern side of the Southern Alps will be subjected to prolonged periods of water deprivation (Renwick *et al.* 2013; Caruso *et al.* 2017). The resilience of New Zealand alpine species, including their adaptive and plastic responses, to climate change is poorly understood. Hence, it is widely recognised that greater research is required for informed decision-making in conservation (Halloy and Mark 2003; McGlone and Walker 2011; Hoffmann *et al.* 2015).

1.3 Introgression

One means by which some species can rapidly adapt to environmental change is through introgressive hybridisation (commonly just 'introgression'). This term describes the integration of genetic material from one population or species into another via interbreeding and repeated backcrossing of hybrids to parental populations (Anderson 1953; Suarez-Gonzalez *et al.* 2018). This transfer of genetic material provides a rapid increase in the standing (existing) genetic variation of the recipient species through the introduction of novel alleles (Schmickl *et al.* 2017). Introgression also provides a potential mechanism by which organisms occupying new habitats can rapidly acquire traits that will help adapt them to local conditions (Whitney *et al.* 2015).

1.3.1 Genetic adaptation

For many years, there has been uncertainty over the relative importance of standing genetic variation for the evolutionary potential of a species (Tigano and Friesen 2016; Lai *et al.* 2019). However, in one recent study investigating factors driving adaptation of populations of vinous-throated parrotbill (*Sinosuthora webbiana*), Lai *et al*. (2019) showed that the contribution of standing genetic variation greatly outweighed that of novel mutations. While all alleles arise by novel genetic mutation, new beneficial alleles can be slow to reach a frequency in a population that makes them useful for responding to rapid environmental change (Barrett and Schluter 2008). However, introgression allows for alleles, pretested by natural selection, to be transferred from the donor species to a recipient within one or more generations (Mitchell *et al.* 2019). The rate at which introgression can facilitate adaptation in plants was demonstrated in a field experiment carried out by Mitchell *et al*. (2019). After only seven generations, a backcrossed population of a non-locally adapted annual sunflower (*Helianthus annuus*) (Asteraceae), incorporating genetic material from the locally adapted *H. deblis,* displayed greater fitness than its recipient parent for a range of physiological traits.

1.3.2 Plasticity

Plant biologists have long been aware that species grown under different conditions exhibit differences in their morphologies and physiologies. This process of acclimation, which represents a plastic response of a species to its environment, is an active area of research in many organisms: animals, plants, fungi, and bacteria (Gao *et al.* 2018; Leinweber *et al.* 2018;

Hokken *et al.* 2019; Bilandžija *et al.* 2020). Phenotypic plasticity involves phenotypic alteration without genetic change but is itself an evolved character (Scheiner *et al.* 2017). The plasticity of species has great importance for understanding the fitness of organisms in different habitats (Scheepens *et al.* 2018; Hiatt and Flory 2020). Notwithstanding potential fitness costs associated with phenotypic changes in response to novel environments (Snell-Rood *et al.* 2018; Fox *et al.* 2019), plasticity contributes significantly to the resilience of species facing environmental changes (Chevin *et al.* 2010; Fox *et al.* 2019). While investigation of this phenomenon was beyond the scope of the present study, its potential significance has been discussed in relation to the findings presented.

1.3.3 Introgression and conservation

Understanding the evolutionary potential and consequences of introgression is important for the management of species (Mable 2019). It is generally accepted that climate change-induced environmental shifts mean that preserving genetic diversity is important because species that lack diversity have reduced evolutionary potential and may fail to adapt in the case of habitat change (Lai *et al.* 2019; Mable 2019). Furthermore, hybridisation has also been associated with increase in the genetic potential of some species and their evolutionary success (Hoffmann and Sgró 2011; Abbott *et al.* 2013; Becker *et al.* 2013). However, in some circumstances, the effects of introgression can be detrimental. An intentional effort to increase genetic diversity in a small population of mountain ibex (*Capra ibex ibex*) by interbreeding them with other ibex subspecies led to extinction when the resultant hybrids were of low fitness in the local environment (Rhymer and Simberloff 1996). Range expansion facilitated by introgression can also put species in direct competition with endangered relatives (Hata *et al.* 2019). Utilising conservation resources, particularly in the case of changing environmental conditions, to increase the resilience of lineages is highly desirable (Becker *et al.* 2013). Nevertheless, predicting when hybridisation will have positive and negative outcomes is considered a significant challenge of our time (Abbott *et al.* 2013).

1.3.4 Detecting introgression

Instances of introgression can remain cryptic, and a species may morphologically resemble the recipient parent but display a physiology more akin to that of the donor parent (Anderson and Hubricht 1938; Lewontin and Birch 1966). Early efforts to identify introgressants relied on identification of intermediate forms and examination—using statistical analyses—of morphological characters (Anderson 1953). However, without a clear relationship

between the introgressant and putative parents, identification wasn't possible (Heiser 1949, 1973; Anderson 1953). With increased use of nuclear and plastid genomic markers, evidence for introgression has become more readily accessible (Rhymer and Simberloff 1996) and the ubiquity of introgression in the evolution of species better understood (Mallet *et al.* 2016). Examination of bi-parentally inherited nuclear markers and (normally) maternally-inherited mitochondrial or chloroplast markers gives researchers evidence for putative introgression events, and can also allow the direction of introgression to be inferred (Rhymer and Simberloff 1996; Joly *et al.* 2009a) and hybridisation events to be dated (Llopart *et al.* 2014). However, in studies of adaptation, short neutral markers such as nuclear ribosomal internal transcribed spacers (ITS) can be limited in value. Analyses of additional markers are often necessary to confidently detect introgression; detection which becomes more difficult with increasing numbers of generations following hybridisation (Joly *et al.* 2009b; McFarlane and Pemberton 2019).

1.3.5 Evaluating the evolutionary significance of introgression

To determine the adaptive significance of introgression, information is needed on the functional significance of introgressed DNA that is maintained in successive populations through positive selection (Kirk and Freeland 2011). Various methods exist to test whether introgressed genes are under positive selection (Becker *et al.* 2013; Booker *et al.* 2017). Furthermore, frequencies of novel DNA variants in populations can be informative as to whether specific sequences confer functional significance for the introgressed DNA (Hoffmann *et al.* 2015; Whitney *et al.* 2015; Rana *et al.* 2019). Yet, methods—such as genome-wide association studies (GWAS) and quantitative trait locus (QTL) mapping—for associating allele frequencies with functional traits are impractical for many non-model plant species (Hoffmann *et al.* 2015; Voelckel *et al.* 2017). To date, in the study of New Zealand alpine plants, while adaptive diversification is thought to have been an important phenomenon that has shaped the evolution of the New Zealand flora (Fisher 1965; Wagstaff *et al.* 2002; Winkworth *et al.* 2005), and while hybridisation has long been thought to be a conspicuous feature of the flora (Cockayne 1923; Cockayne and Allan 1934; Connor 1967; Smissen and Heenan 2007; Smissen *et al.* 2014), there have been few studies that have investigated the functional significance of introgression for adaptive diversification of the New Zealand alpine flora (but see Becker *et al.* 2013).

1.4 Polyploidy

Polyploidy is an important phenomenon in the evolution of plant species. Characterised as possession of more than two sets of chromosomes (Comai 2005), polyploidy is typically partitioned into two broad classifications. Autopolyploids are produced from conspecific parents (Soltis and Soltis 1999). Allopolyploidy arises from interspecific hybridisation (Soltis and Soltis 1999), in which each parental species contributes a distinct subgenome (Cheng *et al.* 2018). Species can be further categorised as paleopolyploids, having arisen from an ancient duplication event and are now functionally diploid, or more recently arisen neopolyploids, which are species that have not fully undergone diploidisation (Blanc and Wolfe 2004). Polyploidy is viewed as ubiquitous in plants (Soltis and Soltis 2000), and at least two ancient whole genome duplication events are believed to have occurred in the evolution of flowering plants (Jiao *et al.* 2011). Evidence suggests that polyploidy is common in the New Zealand flora and has been an important force in the diversification of species (Meudt *et al.* 2021).

1.4.1 Genomic consequences of polyploidy

Changes in genomic content is a typical response to polyploidy. With tetraploid formation, the early polyploid must—if autopolyploid—navigate the immediate doubling of chromosomes (Baduel *et al.* 2018). In the case of allopolyploidy, relationships between genes of the disparate subgenomes (Kryvokhyzha *et al.* 2019) as well as novel cytonuclear interactions (Ferreira de Carvalho *et al.* 2019) present further challenges. The nascent polyploid typically incorporates increased numbers of homologous genes; ohnologues in the case of autopolyploids, and homeologues in the case of allopolyploids (Glover *et al.* 2016). Diploidisation, which implies chromosomal changes that result in a reduction from a number of alleles to only two alleles, has been found to be a compensatory strategy for polyploids (Wolfe 2001). It is a process that begins soon after polyploid formation (Tate *et al.* 2009) and continues through many generations (Schnable *et al.* 2011). An analysis of 3008 polyploid plant species found loss of DNA content was a typical response following polyploidisation (Leitch and Bennett 2004). However, this is not the case in all instances. Leitch *et al.* (2008) found no consistent trend of genome size reduction in *Nicotiana* (Solanaceae) species with polyploid origins dating from < 200,000 years ago to approximately 4–5 million years ago. In fact, a small number of taxa underwent genome size expansions. Total DNA content notwithstanding, gene losses, as part of the diploidisation process, are common and appear to occur in a non-random manner. Cheng *et al.* (2012) discovered dominant expression of genes, corresponding with a decreased frequency of frame-shift mutations, from one subgenome when researching genome fractionation in

Brassica rapa (Brassicaceae). In maize (*Zea mays*) (Poaceae), biased genome fractionation contributed to retention of genes, belonging to one subgenome, that encode interacting proteins (Schnable *et al.* 2011). Lastly, Edger *et al*. (2019) discovered significantly fewer transposable elements in the dominant subgenome of the octaploid cultivated garden strawberry (*Fragaria* × *ananassa*) (Rosaceae).

1.4.2 Adaptation and polyploids

Phenotypically, polyploids are often larger than their progenitor parents and can display novel traits. Increased ploidy has been linked to many changes such as cell enlargement, changes to cell wall composition, and overall increase in biomass (Corneillie *et al.* 2019). Transgressive phenotypic attributes such as these mean most commercial crop plants are polyploid (Udall and Wendel 2006; Renny-Byfield and Wendel 2014). In natural populations, polyploids can often exploit ecological niches more rapidly than their diploid progenitors (Baniaga *et al.* 2020), a feature which contributes to the invasiveness of polyploid weed species (Moura *et al.* 2021). Not surprisingly, the adaptive nature of polyploids has been attributed to the increase in genetic material (Ramsey 2011), but otherwise, their advantages remain an area of ongoing research (Baduel *et al.* 2018). The increased genetic diversity of polyploids might be utilised in two ways: i) the presence of duplicated genes means potential for relaxation of evolutionary constraint on gene copies, and subsequent evolution of novel gene variants, that could be important for adaptation (Cheng *et al.* 2018; Baniaga *et al.* 2020); and ii) the increased standing genetic variation—and in particular, allelic variation—means greater genetic variation upon which epigenetic regulation can influence plasticity (Zhang *et al.* 2013).

1.4.3 Studying polyploids

Phylogenetic study of polyploid organisms is more complex than for diploids. While commonly used for population studies, genomic single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) molecular markers provide challenges in polyploid organisms due to the duplicated subgenomes (Clevenger *et al.* 2018). This may result in detection of multiple loci instead of the single locus targeted (Bourke *et al.* 2018). Genotype by sequencing (GBS) (Elshire *et al.* 2011) has been successfully used to find phylogenetically informative molecular markers in polyploid species (Yang *et al.* 2017b; Campa and Ferreira 2018). However, inferring gene function from GBS data typically requires mapping sequences to a reference genome (Muktar *et al.* 2019).

Assembling genomes of polyploids is a complex issue. The introduction of next generation sequencing (NGS) technologies has substantially reduced the cost and laboratory resource use previously associated with genome assembly (Claros *et al.* 2012; Kyriakidou *et al.* 2018). However, because of the short reads typical of NGS, completion of high-quality assemblies remains a technical challenge, and most plant 'genomes' are fragmented and yet to be assembled to chromosome level (Jiao and Schneeberger 2017). The higher the ploidy of the genome, the more challenging assembly becomes (Kyriakidou *et al.* 2018). Long-read (third generation) sequencing, available through the Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) platforms, is helping to overcome some of the challenges faced when using NGS for genome reconstruction (Jiao and Schneeberger 2017). Kyriakidou *et al*. (2020) were able to improve fragmented assemblies of polyploid potato (*Solanum tuberosum*) (Solanaceae) genomes using PacBio sequencing. However, long-read sequencing is not a panacea. The technology is constantly improving (Kyriakidou *et al.* 2018), but long-held concerns remain about the accuracy of both PacBio and ONT sequencing (Watson and Warr 2019). This is why error correction using short NGS reads is still recommended for genomes assembled with long-read sequences (Zhang *et al.* 2020a). Multiple rounds of NGS, using different DNA fragment size libraries, as well as PacBio sequencing were required for the construction of a high-quality strawberry genome (Edger *et al.* 2019). Given the ploidy of the New Zealand alpine *Ranunculus* (see Section 1.5), genome assembly of these taxa would conceivably be a similarly demanding task.

The challenges inherent in genome assembly have meant alternative methods to generate large numbers of molecular markers have been developed, which also have uses in polyploid plant systems. Markers, comprising approximately 1,000 low-copy nuclear genes, were generated using target capture in research of *Rubus* (Rosaceae) evolution (Carter *et al.* 2019). However, previously assembled reference genomes were needed for the design of baits. Bait design can also be undertaken using messenger RNA (mRNA) sequences but reference transcript sequences are still required (Couvreur *et al.* 2019). The resourcing needs for target capture probe design were noted by Johnson *et al*. (2019) in their paper describing development of the (currently-named) myBaits® Angiosperms-353 Expert Panel enrichment kit (Daicel Arbor Biosciences, Ann Arbor, US). This kit targets 353 single-copy orthologous genes deemed to be highly conserved across the flowering plant phylogeny. In bait validation, a minimum of 100 sequences were recovered from each of the 42 taxa tested (Johnson *et al.* 2019). Issues arising from ploidy remain with target capture techniques. Polyploid taxa in the *Platanthera* subgenus *Limnorchis* (Orchidaceae) were excluded from phylogenetic analysis of *Limnorchis* over concerns of taxonomic complexity (Wettewa and Wallace 2021).

A promising alternative to genomics lies in the use of transcriptomics to uncover genetic markers that delimit taxa and are linked to traits of ecological significance. This

approach has proved useful in diploid species with few genomic resources. By testing samples against a transcriptome containing > 120,000 transcripts, Li *et al.* (2019b) discovered more than 3,000,000 SNPs, and 39,000 SSR loci to be used in further studies of insect defence mechanisms in the London planetree (*Platanus acerifolia*) (Platanaceae). Transcriptomic studies also provide a method for the bulk discovery of molecular markers in polyploid species. In a study of variation within blueberry (*Vaccinium* spp.) (Ericaceae), both SNPs and SSRs delimiting cultivars were identified through transcriptome analyses (Wang *et al.* 2019b). Sprenger *et al*. (2018) noted the difficulty of using standard genomic markers in polyploid organisms. Therefore, they successfully developed markers for selective breeding by identifying polymorphisms segregating with transcripts upregulated in drought tolerant potato cultivars. Interestingly, nucleotide sequences from mRNA transcripts are now being increasingly used for phylogenetic tree building (Wickett *et al.* 2014; Wang *et al.* 2017; Amaral *et al.* 2019; Quek and Huang 2019).

Importantly, RNA sequencing (RNA-seq) performs two tasks. It not only provides a mechanism for generating molecular markers but also allows inference of traits important in physiological responses through gene expression studies (da Fonseca *et al.* 2016). A comparative transcriptomic approach was used to investigate environmentally-induced changes in leaf morphology of the northern hemisphere species *Ranunculus trichophyllus* (Kim *et al.* 2018). Similar methods were used to identify genes responding to abiotic and biotic stresses in natural and cultivated populations of New Zealand alpine *Pachycladon* (Voelckel *et al.* 2008, 2010). For plant species without a reference genome, *de novo* transcriptome assembly methods provide a means of carrying out differential gene expression (DGE) studies (Liu *et al.* 2018; Zhang *et al.* 2020b).

1.5 The alpine *Ranunculus* of New Zealand

With a global distribution across temperate regions of the world, *Ranunculus* L. are herbaceous perennial plants of the Ranunculaceae (buttercup) family (Horandl *et al.* 2005). Within *Ranunculus*, the monophyletic section *Pseudadonis* is comprised of two Australian alpine species, *R. gunnianus* and *R. anemoneus*, and a group of nineteen New Zealand endemic species (de Lange *et al.* 2018), known as the 'New Zealand alpine *Ranunculus*' (or 'alpine *Ranunculus*') (Fisher 1965; Lockhart *et al.* 2001; Lehnebach 2008). The hexaploid (2n = 48, x = 8) alpine *Ranunculus* are a morphologically and ecologically diverse clade with a wide distribution throughout New Zealand mountain environments and have been considered the classic example of adaptive radiation among the New Zealand alpine flora (Fisher 1965).

1.5.1 Origins of the alpine *Ranunculus*

Little is known about the origin of the alpine *Ranunculus* group, and the location of the ancestor from which the clade evolved remains undetermined. Interestingly, New Zealand lowland species of *Ranunculus* show no close relationship to their alpine relatives (Lehnebach 2008; Emadzade *et al.* 2010). Fisher (1965) hypothesised the introduction of the group to New Zealand occurred via transoceanic dispersal from South America. Raven (1973), on the other hand, argued for New Guinea and Australia acting as stepping stones during long-distance dispersal from the Northern Hemisphere. However, more recent phylogenetic analyses have found little support for either of these hypotheses (Horandl *et al.* 2005; Lehnebach 2008). The alpine *Ranunculus* share chromosome number and ITS DNA sequence similarities with species of Europe, South America, and sub-Antarctic islands (Horandl *et al.* 2005; Lehnebach 2008). Nevertheless, determination of an immediate ancestor, or even the closest relatives of the group, remains elusive (Lehnebach 2008).

Evidence suggests radiation of the group occurred recently and early divergence within the group has led to two reproductively isolated clades. Diversification of the alpine *Ranunculus* into alpine habitats, and the evolution of extant species, likely occurred following onset of the Pleistocene (i.e. within the last 2.6 million years) (Heenan and McGlone 2013). This was a geological period marked by many repeating glacial and interglacial cycles (Suggate 1990; Newnham *et al.* 1999). Fisher (1965) studied morphological affinities of the alpine *Ranunculus* species and concluded the group arose from a single ancestor situated in the Southern Alps. Subsequent molecular systematic studies by Lockhart *et al*. (2001), Horandl *et al*. (2005), and Lehnebach (2008) supported the hypothesis of monophyly. Postulated by Fisher (1965), and subsequently reinforced by analyses of chloroplast and nuclear markers (Lockhart *et al.* 2001), radiation of the alpine *Ranunculus* appears to be characterised by a primary division into three or four distinct lineages compatible with Fisher's inference of two distinct breeding groups (Lockhart *et al.* 2001). From within one of these lineages the Australian species *R. anenomeus* and *R. gunnianus* have evolved in the Australian mountains following transoceanic dispersal (Lockhart *et al.* 2001; Lehnebach 2008). Some present-day species were not recognised by Fisher (1965) at the time of his study. These included *R. scrithalis* (which he considered an unusual form of *R. sericophyllus*), *R. viridis*, as well as distinct morphotypes assigned to *R. haastii* and *R. insignis* (Webb *et al.* 1988; Heenan *et al.* 2006; Lehnebach 2008).

A molecular clock date estimate was made by Lockhart *et al*. (2001) for the primary divergence by analysing partial chloroplast *ycf1*—denoted the J_{SA} region in Lockhart *et al.* (2001)—and internal transcribed spacer 2 (ITS2) sequences. The variance on this clock estimate was large due to the short sequence lengths, and only the lower bound is consistent with an inference made by McGlone and Heenan (2013) that strict alpine habitats in New Zealand have

formed within the last million years. McGlone and Heenan speculate that while formation of New Zealand alpine habitats began approximately five million years ago due to tectonic uplift, rising and falling tree lines linked to cycles of warming and cooling have meant that permanent alpine habitat has existed for less than one million years. Subsequent to formation of persistent alpine areas, repeated cycles of glaciation took place in what is now the Southern Alps (Suggate 1990). These changes in the alpine environment likely facilitated multiple instances of range expansion and contraction that have been linked to increases in alpine plant diversification rates (Winkworth *et al.* 2005).

1.5.2 Reproductive biology

Based on cultivation experiments testing the breeding compatibility of species in a common garden experiment, Fisher (1965) determined that most of the 14 species he identified could be separated into two distinct breeding groups consistent with his primary dichotomy hypothesis. Different degrees of interspecific fertility occurred within what Fisher deemed the many-petalled and few-petalled lines, and only sterile hybrids resulted from crosses made between the groups.

While predominantly outcrossing, inbreeding is a feature of alpine *Ranunculus* reproduction biology. Fisher (1965) noted the flowers are protogynous. That is, stigmas become receptive before the anthers mature and dehisce. Considered typical for *Ranunculus* (Steinbach and Gottsberger 1994; but see Kipling and Warren 2014), pollination of the alpine *Ranunculus* flowers is routinely carried out by non-specific insect pollinators such as Syrphidae flies (Fisher 1965). Experimentally, alpine *Ranunculus* individuals display almost no self-incompatibility (Fisher 1965). Fisher (1965) believed a significant proportion of successful pollination, and therefore inbreeding, occurs when a visiting insect visits flowers of different maturities on the same flower stem. Beyond the work conducted by Fisher, there is a paucity of data for the New Zealand alpine *Ranunculus* breeding systems. However, in a study of five Australian alpine *Ranunculus* by Pickering (1997), flowers were found to be self-fertile when hand pollinated. Additionally, protogyny was not complete. If pollination did not occur within the first one to three days of stigma maturity, pollen, either carried by an insect pollinator or transferred directly from the dehiscing anthers, was likely to cause self-fertilisation (Pickering 1997).

Self-fertilisation may lead to inbred populations (Rodrigues *et al.* 2019). Certainly, seed dispersal is unlikely to promote gene flow between populations of alpine *Ranunculus*. The North American alpine species *R. adoneus* seed dispersal was found to be limited to within 30 cm of the parent plant over a two year period (Scherff *et al.* 1994), and (Lehnebach 2008) detected no structures on any New Zealand alpine *Ranunculus* seed which could contribute to

wind-mediated or animal-mediated dispersal. Typically, there are concerns that inbreeding can lead to a lack of genetic diversity (Bona *et al.* 2019) and an increase in homozygosity (Haag and Ebert 2004), which has been linked to the phenomenon of inbreeding depression (East 1908; Schrieber *et al.* 2021; but see Undin *et al.* 2021). Any inbreeding arising from self-fertilisation is presumably exacerbated by instances of vegetative reproduction from rhizomes (Fisher 1965). This is because increased levels of clonal propagation in otherwise sexually reproducing plant populations have been noted to result in decreased genotypic richness (Dorken and Eckert 2001).

1.5.3 Habitat characterisation

Fisher's (1965) field observations led him to suggest that there were five distinct habitat types in which alpine *Ranunculus* species are found. Although these habitats were not rigorously described by Fisher, he emphasised differences in the moisture requirements of species. Based on his compatibility studies, he concluded that some species of the different breeding groups have convergent phenotypes and occur in similar habitats. However, precise ecological requirements differ between these convergently-evolved species, and fine-scale niche sharing is minimised even in cases of geographic overlap (Fisher 1965). A systematic assessment of species' niche preferences was attempted by Lehnebach (2008). However, it was concluded that the Land Environments of New Zealand (LENZ) database (Manaaki Whenua – Landcare Research, Lincoln, New Zealand) used for these analyses lacked the accuracy necessary to detect fine-scale habitat variation. More recently, preliminary habitat characterisation, measuring a range of environmental variables, has indicated that species with overlapping ranges—*R. nivicola*, *R. insignis,* and *R. verticillatus*—occupy niches primarily delineated by distinct soil moisture characteristics (Becker 2020).

1.5.4 Hybridisation within the clade

Introgressive hybridisation is believed to have played an important role in the evolution of the alpine *Ranunculus*. Fisher (1965) explained morphological variation of species by hypothesising first generation and subsequent generation hybridisation events, as well as hybrid origins for some species. This argument for the importance of hybridisation is consistent with preliminary research using available molecular evidence (Lockhart *et al.* 2001, 2014). For example, Lockhart *et al*. (2001) reported phylogenetic inferences for ITS2 and *ycf1* sequences consistent with allopolyploid origins of *R. nivicola.* Lockhart *et al*. (2014) also inferred local instances of introgressive hybridisation between *R. haastii* and at least two other species, based

on ITS2 and y*cf1* sequence data. Hybridisation has long been hypothesised to be common in the evolution of many New Zealand alpine plants (Cockayne 1923; Connor 1985). However, experimental evidence is lacking (Connor 1985); which is why increased use of molecular markers has been advocated for strengthening support for these hypotheses (Morgan-Richards *et al.* 2009).

1.6 Project background

Since the pioneering work of Cockayne (1910), which initiated some of the earliest discussions of adaptative diversification in the New Zealand flora, relatively little progress has been made in testing hypotheses of plant adaptation in New Zealand; but see investigations into crypsis in alpine plants (Strauss *et al.* 2015; Niu *et al.* 2017) and discussions on the evolutionary drivers of divarication (Greenwood and Atkinson 1977; McGlone and Webb 1981; Lusk *et al.* 2020). As previously mentioned, Fisher (1965) did suggest that New Zealand *Ranunculus* species in section *Pseudadonis* had unique morphological features consistent with shared common ancestry. Lockhart *et al*. (2001) went further to provide evidence from DNA sequence analyses suggesting that all species of section *Pseudadonis* were derived from a common shared ancestor (monophyletic) and were recently evolved (Schluter's criteria i and iv). Fisher (1965) also suggested that some morphological features had convergently evolved in species occupying similar habitats such as those on grassland or screes. However, associations between specific phenotypic traits and specific environmental features were not rigorously evaluated (Schluter's criterion ii). Nor was the utility of traits demonstrated or rigorously investigated (Schluter's criterion iii), although Fisher (1952) did speculate on the adaptative advantage of the convergent phenotypes (e.g. deep tap roots and waxy coating of leaves) in Canterbury scree species from different plant genera, including *Ranunculus*.

To extend the work of Fisher (1965) and Lehnebach (2008) on adaptive diversification, the present study investigated alpine *Ranunculus* species in the Canterbury region. The microhabitats of species at Mount Hutt, Porters Pass, and Lake Tennyson were characterised, and transcriptome analyses were undertaken on plants cultivated from these sites. While specific questions of interest included whether there was evidence for local adaptation, and if so, whether this was achieved through hybridisation and introgression, further important issues addressed the application of transcriptomics for studying non-model polyploid alpine plants.

1.6.1 Taxa of interest

Two forms of the species *R. insignis*, and *R. crithmifolius* (both members of Fisher's 'few petalled line') were the focus of study. Taxonomic revision of the alpine *Ranunculus* (Fisher 1965; Webb *et al.* 1988) collapsed *R. insignis* Hook. f., *R. monroi* Hook. f. (Figure 1.1, A), and *R. lobulatus* (Kirk) Cockayne (Figure 1.1, B) into a single species: *R. insignis*. Although, later evidence has indicated strong support for the reinstatement of *R. monroi* to species level, with weaker support for reinstatement of *R. lobulatus* (Lehnebach 2008). In this work, the names *R. monroi* and *R. lobulatus* are used to distinguish between *R. insignis* forms. All three forms of *R. insignis* are found in the mountains of the South Island of New Zealand (Figure 1.2) and were referred to by Fisher (1965) as inhabiting sheltered situations. A feature of open semi-stable scree slopes, *R. crithmifolius* (Figure 1.1, C) has an overlapping geographic distribution with *R. monroi* and *R. lobulatus* (Figure 1.2).

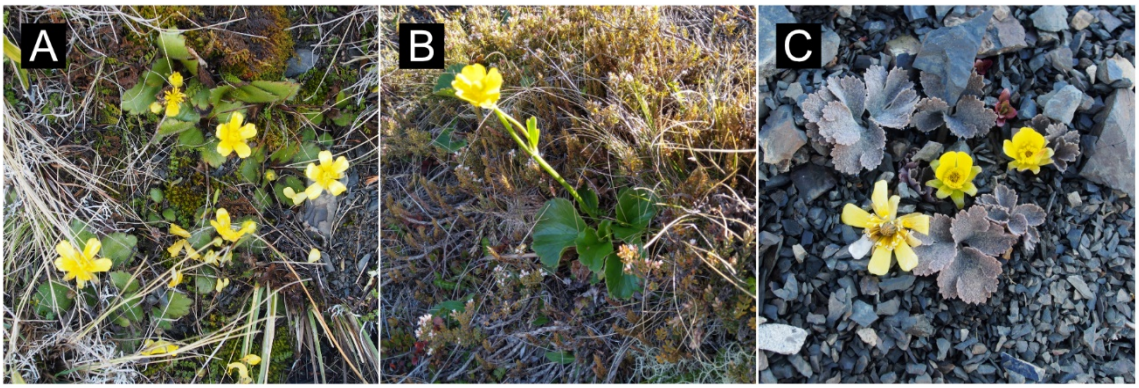


Figure 1.1 Members of the alpine *Ranunculus* clade under investigation. A, the monroi form of *R. insignis*. B, the lobulatus form of *R. insignis*. C, *R. crithmifolius*. In this study, A and B are referred to as *R. monroi* and *R. lobulatus* respectively.

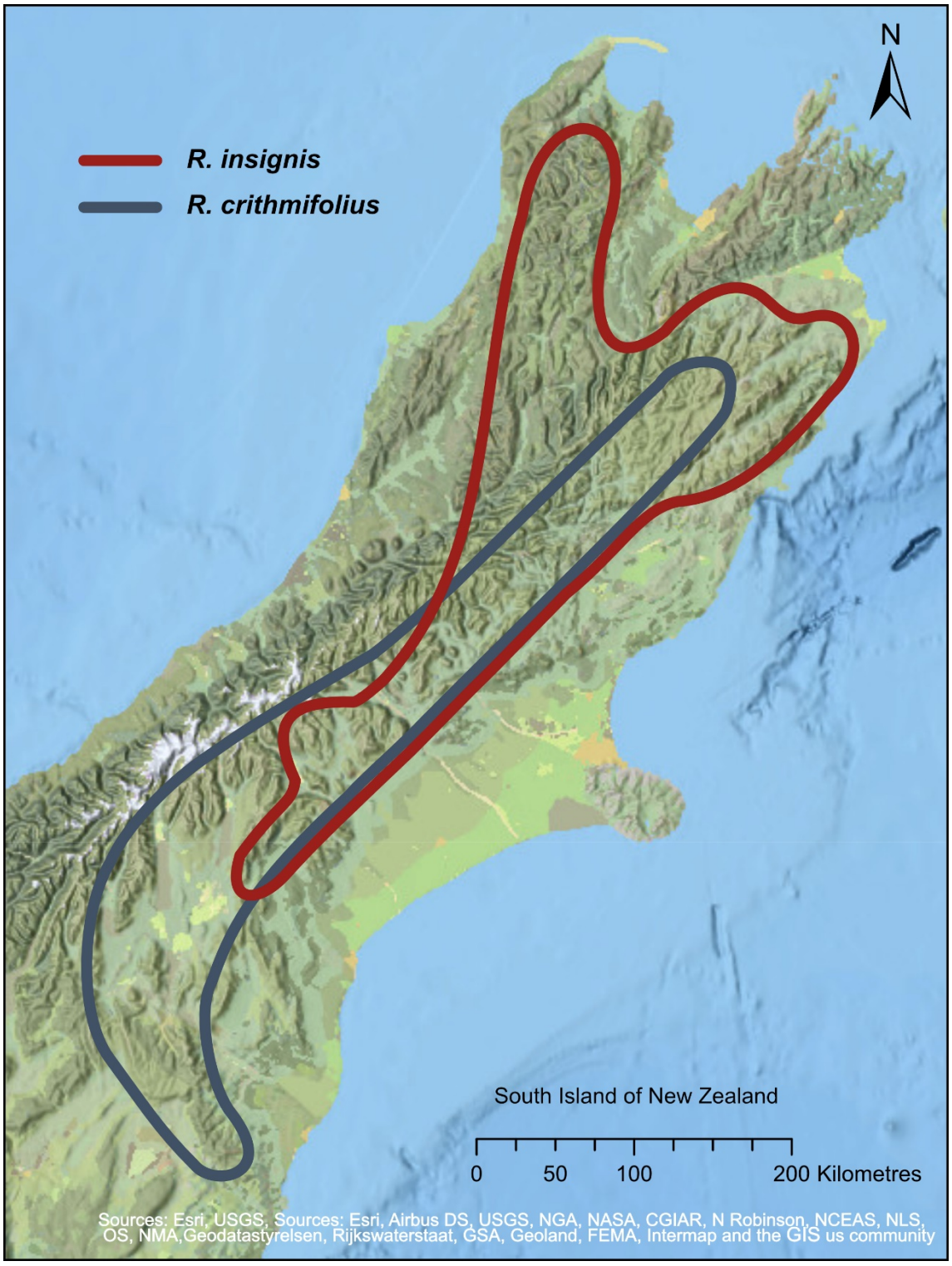


Figure 1.2 South Island distributions of *R. insignis* and *R. crithmifolius*. This map, adapted from Fisher (1965), shows the overlapping distributions of *R. insignis and R. crithmifolius* in the South Island of New Zealand. All three forms of *R. insignis* are present within different regions of the overall *R. insignis* range.

Some interesting observations concern *R. monroi* and *R. crithmifolius* from Mount Hutt in the Southern Alps. While phenotypic analyses (Lehnebach 2008) and ITS2 sequence analysis (Carter 2006; Lehnebach 2008) indicate that *R. monroi* is most closely related to *R. insignis* and *R. lobulatus*, chloroplast *ycf1* sequences from samples of *R. monroi* from Mount Hutt have instead clustered this taxon with *R. crithmifolius* (Carter 2006; Lehnebach 2008). At Mount Hutt, the *R. monroi* habitat (Figure 1.3, A) appears at least superficially similar to the open scree niche occupied by *R. crithmifolius* (Figure 1.3, B) and different from the sheltered conditions typically associated with *R. monroi* (Fisher 1965; Lehnebach 2008) at Porters Pass (Figure 1.3, C). These observations suggest convergent local adaptation that might be explained by hybridisation-facilitated niche shift of species as hypothesised by Fisher (1965).

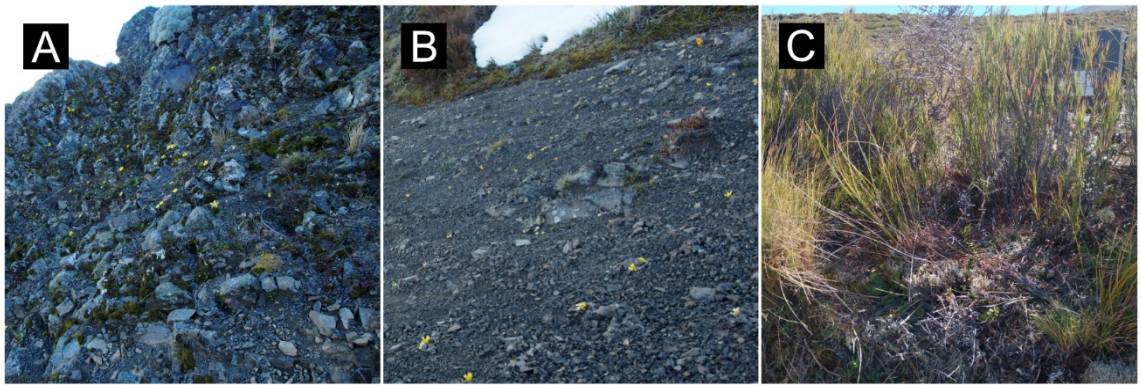


Figure 1.3 Habitat comparison between *R. crithmifolius* at Mount Hutt, *R. monroi* at Mount Hutt, and *R. monroi* at Porters Pass. A, exposed habitat of *R. monroi* at Mount Hutt. B, typical *R. crithmifolius* habitat on open scree at Mount Hutt. C, typically-described sheltered habitat of *R. monroi* at Porters Pass.

1.6.2 Research hypothesis

Introgression of genetic material from *R. crithmifolius* to *R. monroi* has facilitated a niche shift of *R. monroi* into open habitat at Mount Hutt.

1.6.3 Research aims

To help test this hypothesis the study aims included:

- Characterisation of the microhabitats of *R. monroi and R. crithmifolius* at Mount Hutt, *R. monroi* at Porters Pass, and *R. lobulatus* at Lake Tennyson.
- Establishment of a bioinformatics pipeline for differential expression and nucleotide variation analyses using RNA-seq data from closely-related species.
- Development and evaluation of a phylogenetic protocol for identifying putatively introgressed genes, as well as genes that potentially contribute to adaptive physiological trait differences.

1.6.4 Study overview

The current work investigates processes of adaptation in the New Zealand alpine *Ranunculus.* It builds on systematic studies of this group (Fisher 1965; Lockhart *et al.* 2001; Heenan *et al.* 2006; Lehnebach 2008). Hypotheses of physiological similarities and gene flow between *R. monroi* at Mount Hutt and *R. crithmifolius* are tested. Inclusion of *R. lobulatus*, as another form of *R. insignis*, enables four-taxon comparisons to be made and provides an opportunity to test the null hypothesis that there is little difference between the two *R. monroi* taxa (Mount Hutt and Porters Pass). Habitat characterisation was carried out to objectively assess observed differences between study sites. A common garden experiment was used to mitigate environmental influences on plant responses. Then, by utilising RNA-seq, analyses of gene expression and allelic variation were used to identify traits and genes both consistent and anomalous with respect to phylogenetic expectations. That is, there is an expectation the two *R. monroi* taxa should share similar gene expression profiles and DNA sequence similarities. However, at Mount Hutt, gene expression underlying traits important in allowing *R. monroi* to colonise exposed habitat should be more similar between Mount Hutt *R. monroi* and *R. crithmifolius*. Subsequently, genetic linkage means analyses of nucleotide variation might help identify candidate genes or regulatory regions underlying these similar gene expression patterns.

Chapter 2: Materials and Methods

2.1 General notes

2.1.1 Permitted activities

Research activities carried out on New Zealand public conservation land were permitted by the Department of Conservation (Permission Number: 72743-RES).

2.1.2 Software codes, software scripts, and supplementary materials

Where applicable, software scripts, software code, and supplementary materials have been uploaded to folders and files at https://github.com/jc-henry/Masters/. For brevity, these items are denoted in this text as '/<folder>/<file>'. Example scripts and code have been edited for clarity.

2.1.3 Computer hardware

The computing requirements of this project meant many processes were carried out on the Mahuika high-performance computing cluster of New Zealand eScience Infrastructure (NeSI). All other computation work was carried out on a Hewlett Packard Elitebook 830 G5 with Intel® Core™ i7-8650U CPU @ 1.90GHz 2.11 GHz and 32 GB RAM. On this laptop, any software requiring a Linux operating system was run in the Microsoft Windows Subsystem for Linux (WSL) 2. Scripts and commands for running software on NeSI have been saved to the '/Masters/nesi/' directory. Other scripts and commands have been saved to '/Masters/local/'.

2.1.4 Statistics

All statistical testing for significance was carried out in R v4.0.2 (R Core Team 2020) using RStudio v1.3.959 (RStudio Team 2020). Before any tests for statistical significance were conducted, data were assessed for normality using the Shapiro-Wilk test. The resulting p value for all tests was below the chosen significance level set to 0.05. The null hypotheses were rejected, and all data deemed to be not normally distributed. Therefore, further tests for the

significance of differences between groups were carried out using non-parametric alternatives to standard tests such as *t*-tests and analysis of variance (ANOVA).

2.1.5 Graphs and diagrams

Scatter plots, line diagrams, volcano plots, and bar charts were plotted using ggplot2 v3.3.3 (Wickham 2011). Principal component analysis (PCA) plots and Violin plots were constructed with ggbiplot v0.55 (https://github.com/vqv/ggbiplot) and Vioplot v0.3.6 (Adler and Kelly 2019) respectively. Heatmaps were generated using gplots v3.1.1 (Warnes *et al.* 2016). Plot colouring was inspired by the palettes of Manu v0.01 (https://github.com/G-Thomson/Manu/).

2.1.6 Other R packages

Data wrangling and string manipulation in R relied heavily on Tidyverse v1.3.0 (Wickham *et al.* 2019) packages and strex v1.4.1 (https://github.com/rorynolan/strex). Manipulation of phylogenetic trees required phangorn v2.6.2 (Schliep 2011).

2.2 Plant locations

2.2.1 Sample site selection

Sites were chosen based on several criteria. Primarily, dense populations (> 100 plants per site) of the plants under investigation were necessary to ensure the local environments were representative of environmental niches. Additionally, the sites needed to be accessible for the purposes of equipment installation and plant sampling, yet hidden from public view to deter vandalism of equipment.

A location at Mount Hutt, near the Mount Hutt Skifield Road, but not visible from the road, was the principal site for this study (Figure 2.1; Supplementary Table 1). At this locale, *R. monroi* was found growing on fringes of the exposed scree faces where *R. crithmifolius* occurred (Figure 2.2). It should be noted that elsewhere, at the Mount Hutt site, additional small pockets of *R. monroi* can be found growing among the *R. crithmifolius* where rock outcrops or other factors have stabilised the scree slope. A second location was chosen at Porters Pass (Figure 2.1; Supplementary Table 1). Here, a healthy population of *R. monroi* was found growing on the southern slopes of Foggy Peak overlooking State Highway 73. Lastly, at Lake

Tennyson, a site was established 200 m up a steep slope on the eastern side of the lake (Figure 2.1; Supplementary Table 1) where *R. lobulatus* is found.

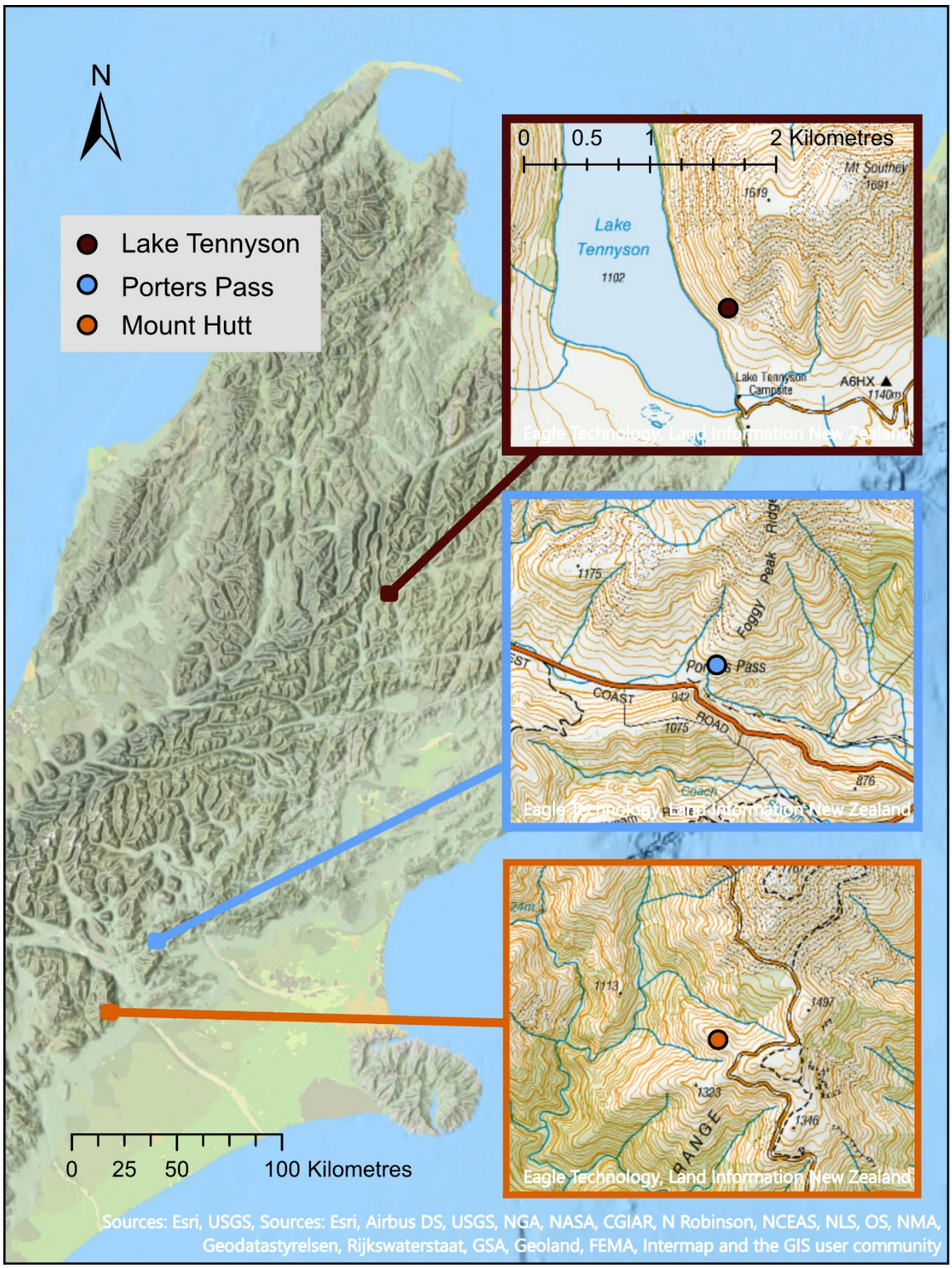


Figure 2.1 Sampling sites for this study. Insets show topographical information of sites at 1:50,000 scale.

Figure 2.2 *Ranunculus monroi* and *R. crithmifolius* at Mount Hutt. This picture shows where plants were sampled at Mount Hutt. Coloured outlines indicate distribution of taxa. The gravel and soil patches on the less mobile rock face provides habitat for *R. monroi*. The scree face (outlined in green) contains large numbers of *R. crithmifolius*. Little to no overlap of taxa is observed here.

2.2.2 Physical and vegetation factors

Each site was evaluated for physical and vegetative characteristics. These were adapted from 'The Recce method for describing New Zealand vegetation – Field Protocols' (Hurst and Allen 2007). Altitude was recorded from global positioning system (GPS) readings, while aspect and slope were measured using a compass and clinometer respectively. Terrain features were noted, and estimates made of vegetation coverage. Cataloguing vegetation diversity was outside the scope of this study.

2.2.3 Soil moisture

Data loggers were established at each location to continuously monitor soil moisture. During November 2019, a single HOBO Micro Station Data Logger #H21-002 with four EC5 Soil Moisture Smart Sensors #S-SMC-M005 (Onset Computer Corporation, http://onset.com) was installed at each site. The data loggers were fastened inside weather-proof custom aluminium housings, which were powder coated brown to reduce visibility. These housings were secured to steel Y-posts ('waratahs') driven into the ground (Figure 2.3, A). The EC5 Soil Moisture Smart Sensors measure the volumetric moisture content (VMC) of soil with an accuracy of ± 3% without calibration for soil type or mineral content. Data loggers were located such that the moisture probes could be placed among actively growing *Ranunculus*. At Mount Hutt, the sensors were fixed in an area of heavy *R. crithmifolius* growth. A second data logger was installed at Mount Hutt during November 2020 to capture soil moisture data more effectively for the *R. monroi* plants located there. To record soil moisture at the *Ranunculus* root zones, moisture sensors were buried at their minimum effective depth, approximately 25 mm under the soil surface (Figure 2.3, B). These sensors were vertically orientated to minimise site disturbance. Soil moisture was sampled by each sensor every minute with mean values logged every 30 minutes. Loggers were downloaded periodically during site visits, and the equipment checked for continuing function.

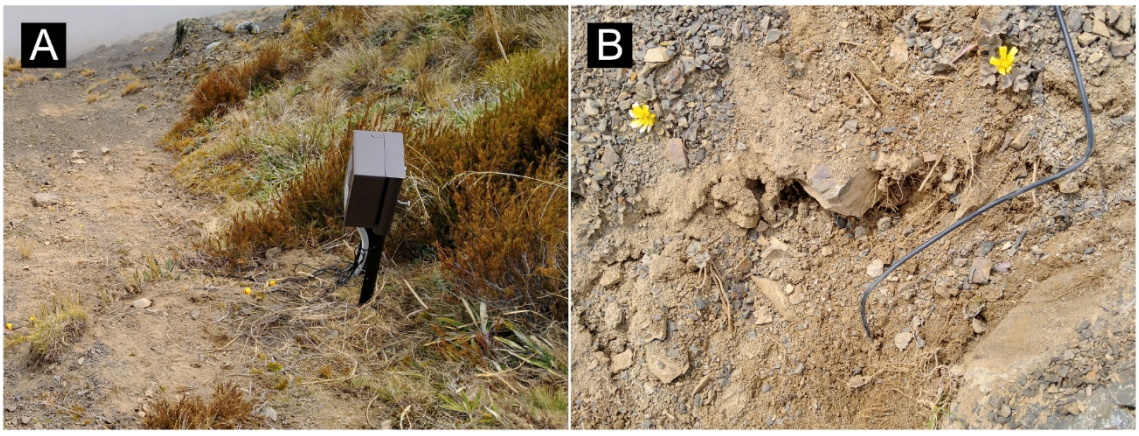


Figure 2.3 Installation of data logger and soil moisture sensors. A, data logger housing positioned, in an area of relative stability, with sensors extending out into the *R. crithmifolius* habitat of the scree face. B, installation of a soil moisture sensor among *R. crithmifolius* plants.

2.2.4 Soil physical characteristics

Soil particle size distributions were examined to make inferences of soil water holding capacity. Five replicates of soil samples were removed from each plant sampling site. Excavations were made at depths of 0–5 cm and 5–10 cm, with approximately 150 cm^2 of soil extracted from each vertical section. These depths were chosen based on observations of alpine *Ranunculus* root distributions in the soil. Samples were air dried for a period of five weeks before being passed through a series of sieves, with diminishing mesh sizes (16.00 mm, 8.00 mm, 5.60 mm, 4.00 mm, 2.00 mm, 1.00 mm, 0.50 mm, 0.25 mm), to separate the samples into nine size fractions (/supplementary/soil_raw.csv). These were weighed and the proportion of the total sample mass each fraction comprised was calculated.

2.2.5 Soil chemical composition

In addition to particle sizes, the chemical composition of soil at each site was also investigated. After soil sieving was completed, all fractions comprised of particles less than 2.00 mm in size were recombined and sent to Hill Laboratories (Hamilton, New Zealand) for analyses. Two replicates (replicate one and replicate three) from the Mount Hutt *R. monroi* site did not contain sufficient mass of soil individually so were combined to make up the 40 g necessary. To maintain equal numbers of replicates among sites, only four replicates (one to four) were analysed from the Mount Hutt *R. crithmifolius* site and Porters Pass. Additionally, no samples from Lake Tennyson were above 40 g in weight, so, for this site, replicates one, two, and three were combined to attain sufficient mass for analyses. A number of extractions and analyses were carried out by Hill Laboratories to characterise the samples. A 1:2 (v/v) soil:water slurry was made and pH determined using a potentiometer. Olsen (plant available) phosphorous (P) was extracted using the Olsen method and quantified with molybdenum blue colourimetry. Total phosphorous was determined by inductively coupled plasma optical emission spectrometry (ICP-OES) after nitric/hydrochloric digestion. Sodium (Na), potassium (K), magnesium (Mg), and calcium (Ca) were extracted and quantified using 1 mol L^{-1} neutral ammonium acetate followed by ICP-OES. Total nitrogen (N) and total carbon (C) were determined using near-infrared spectroscopy. The ratio of C to N (C:N) was calculated to better understand the plant availability of N. Percentage of organic matter (OM) was inferred to be 1.72x total C. Summation of extractable Na, K, Mg, and Ca, with extractable acidity, produced a value for cation exchange capacity (CEC). Total base saturation (TBS) was calculated from extractable cations and CEC.

To mitigate complexity of the multivariate data in the soil chemical composition dataset, PCA were carried out (using the R function 'prcomp') to investigate the important chemical factors underlying similarities and differences between sites. Variables were centred and scaled (/local/pca.R). Because of the direct correlation, OM values were retained and total C was excluded in the PCA. As some sample variables in the raw data (/supplementary/chem_raw.csv) were determined to be below detectable limits, these data were entered as the value of detectable limit to complete the PCA (/supplementary/chem_clean.csv).

2.3 Plant growth and taxon delineation

2.3.1 Plant sampling and growth

Ranunculus were sampled from the field and transported to Massey University. Difficulty in cultivating the alpine *Ranunculus* from seed meant that extraction of mature plants from natural populations was necessary. Plants at each site were sampled at the same time as data logger installation. Plants were dug up with soil still encasing the roots. Roots (and soil) were wrapped in moist paper towels and the plants placed into plastic bags. All plant roots were maintained in a watered state and transported to the School of Fundamental Sciences, Massey University, Palmerston North.

All *Ranunculus* plants were processed and planted into pots for ongoing growth. On 29 November 2019, each plant was root-washed to remove any matter remaining from its collection site. An initial pilot study, undertaken at the beginning of the project, found a standard nutrient-rich potting mix unsuitable for cultivating *R. crithmifolius* because plants rapidly flowered and then died. For this reason, the growth medium chosen had reduced nutrient levels. The potting mix used was Daltons Base Mix (Daltons, Matamata, New Zealand), comprising 50% *Pinus radiata* bark with calcium and ammonium nitrate additive, 30% coconut fibre, and 20% pumice (7 mm). Osmocote® Exact (AICL Specialty Fertilisers, Tel Aviv, Israel) slow release fertiliser was applied to the potting mix at a rate of 1.5 g/L. Plants were weighed before being potted into 2 L square plastic pots with 800 g of potting mix. To each pot, 300 mL of water was added, to make a total weight of approximately 1,150 g, and the plants placed into a Contherm 620 RHS Growth Chamber (Contherm Scientific, Lower Hutt, New Zealand).

Soil bulk density (BD) was calculated to later determine potting mix moisture levels. Early testing had indicated the sensors used for measuring soil moisture at the sampling sites were unreliable when used in the plastic pots. Additionally, repeated insertion and removal of the sensors to facilitate watering added cavities to the potting mix. Therefore, potting mix moisture was monitored manually through a system of weighing and drying. To calculate BD,

potting mix was added to three glass beakers and compacted, with the same pressure used to pot the plants, to the 500 mL mark. The potting mix was then dried for 48 hours at 80 °C, a temperature chosen to minimise loss of organic material (O'Kelly 2005), before weighing. Bulk density was then calculated by dividing the mass of dry soil by the compacted volume and estimated at 356 g/L (/supplementary/bulk_density.csv).

Conditions within the growth chamber were set to imitate summer environmental conditions at Mount Cook Village in the Southern Alps (NIWA 2010). This area was chosen since it is host to a number of different alpine *Ranunculus* species but not *R. monroi* nor any *R. insignis* forms (Fisher 1965; Lehnebach 2008). Daylength, with lighting provided by an alternating mix of Gro-Lux® and Cool White fluorescent tubes (Sylvania, Budapest, Hungary) was set to a 16-hour light – 8-hour dark cycle (180 μE light intensity setting). Temperature varied from 9 °C in the dark phase to 20 °C during the light phase. Relative humidity was constant at 75%. Validation of cabinet temperature, relative humidity, and photosynthetically active radiation (PAR) was undertaken by measuring these factors at different positions within the growth chamber using a HOBO Micro Station Data Logger #H21-002 with Temperature/RH Smart Sensor S-TBH-M00x and Photosynthetically Active Radiation Smart Sensor # S-LIA-M003 sensors (Onset Computer Corporation).

Plants were subjected to a regular watering regimen. After initial acclimation, every four days (excepting a single instance of a five-day interval prior to stress assessment) plants were removed from the growth chamber, inspected, photographed, weighed, and rewatered to a total weight of 1,050 g. To randomise the position of plants in the chamber after each watering, and to generate any other random assignments in this experiment, the R function 'sample' was used. Tap water used for watering was left to sit for four days to allow excess chlorine to evaporate. At the second watering, Plantmate™ (Egmont Commercial, Auckland, New Zealand), a bio-fungicide, was added to the water at a rate of 1g/L for prophylactic suppression of fungal infestation. Water was administered to the plants with a 60 mL syringe. This allowed substantial, yet precise, amounts of water to be deposited across all areas of the potting mix. Often, a common garden experiment watering regimen involves constant addition of repeated small amounts of water sufficient to maintain a previously-established soil moisture content (Fracasso *et al.* 2016; Puértolas *et al.* 2017). This may potentially result in a more stable moisture content more closely approximating field conditions, but it has been argued this approach causes zones of different soil moisture, and responses of individual plants will vary based on root density in these zones (Lawlor 2012). Twice, throughout the course of the experiment, insect infestation was detected, and all plants were treated with Mavrik® Insect & Mite Spray (Yates, Clayton, Australia)—an insecticide containing the active ingredient tau-fluvalinate at 0.1 g/L .

2.3.2 DNA barcoding

Sequencing of chloroplast *ycf1* and ITS2 DNA barcodes, and phylogenetic analyses were carried out to inform comparisons of plants. These markers are commonly used in plant phylogenetics due to high copy number and discriminatory power among closely-related taxa (Li *et al.* 2011; Dong *et al.* 2015). Furthermore, both markers have been previously successfully used in phylogenetic analyses of the alpine *Ranunculus* (Lockhart *et al.* 2001; Lehnebach 2008). Lehnebach (2008) highlighted the presence of different chloroplast haplotypes in *R. monroi.* A lack of information regarding the functional significance of these chloroplast variants made it prudent to restrict downstream analyses of Porters Pass *R. monroi* to a single chloroplast haplotype. DNA extraction was carried out, using the proprietary tools and reagents prescribed for use with *phyto*GEM kits (MicroGEM, Southampton, UK) in the *phyto*GEM plant DNA extraction protocol. Briefly, leaf tissue was sampled using a Crusher Tool and phytoCards. Dispensed into each PDQeX cartridge was DNA extraction buffer comprising: 10 µL GREEN+ buffer, 2 µL *prep*GEM, 2 µL *phyto*GEM A, 10 µL *phyto*GEM B, and 76 µL Invitrogen™ Nuclease-Free Water (Thermo Fisher Scientific, Waltham, USA). The cartridges were placed into a PDQeX-2400 after the addition of three phytoCard punched discs. The PDQeX was then run at 35 °C for 5 min, followed by 75 °C for 5 min, and 115 °C for 2 min.

Two separate polymerase chain reaction (PCR) amplifications were carried out to target the barcoding loci. PCR amplifications were performed in 20 µL reaction volumes containing 1x EmeraldAmp® GT PCR Master Mix (Takara, Shiga, Japan), 3 µL of 100-fold diluted DNA template, and 0.5 µM of each amplification primer. Thermocycling was carried out using a T1 Thermocycler (Biometra, Analytik Jena, Jena, Germany). Amplifying part of the chloroplast *ycf1* region utilised J_{SA} primers described by Lockhart *et al.* (2001) and cycling conditions with initial 3 min denaturation at 94 °C, then 35 cycles of 94 °C for 30 s, 50 °C for 30 s, and 72 °C for 30 s, with final extension at 72 °C for 5 min. PCR primers ITS4 (White *et al.* 1990) and ITS5HP (Hershkovitz and Zimmer 1996) were used to amplify the ITS2 region. Thermo cycling was set to: 3 min denaturation at 94 °C, then 35 cycles at 94 °C for 30 s, 48 °C for 30 s, 72 °C for 30 s, and final extension at 72 °C for 5 min. Amplification products were visualised following agarose gel electrophoresis using Invitrogen™ SYBR™ Safe DNA Gel Stain (Thermo Fisher Scientific) with a Uvidoc HD6 (Uvitec, Cambridge, UK).

PCR products were prepared for DNA sequencing. Five units of Exonuclease I (New England Biolabs, Ipswich, USA) and 0.5 units of Shrimp Alkaline Phosphatase (New England Biolabs) were added to 12 µL of PCR reaction and made up to 15 µL total volume with Invitrogen™ Nuclease-Free Water. The cleanup reaction was carried out in a T1 Thermocycler run at 37 °C for 15 min then 80 °C for 15 min.

Cleaned products were used as templates for Sanger sequencing (Sanger *et al.* 1977). Sequencing was performed using a BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) in a T1 Thermocycler. Forward and reverse reactions were carried out for each locus. Four µL of template was added to reaction mixture containing: 1 µL of BigDye™ Terminator v3.1 Ready Reaction Mix, 3.5 µL of 5X buffer, 1 µL of primer, and 11.5 µL of Invitrogen™ Nuclease-Free Water. Thermo cycling conditions were: 2 min denaturation at 98 °C, followed by 28 cycles of 98 °C for 10 s, 50 °C for 10 s, and 60 °C for 4 min. Capillary electrophoresis was carried out by the Massey Genome Service (Palmerston North, New Zealand), using an Applied Biosystems™ 3730 Genetic Analyzer (Thermo Fisher Scientific). DNA sequences were subsequently checked, and multiple sequence alignments (MSAs), conducted manually, using Geneious v9.1.8 (Kearse *et al.* 2012). Ambiguous sites were retained using International Union of Pure and Applied Chemistry (IUPAC) nucleic acid notation (Cornish-Bowden 1985). Phylogenetic networks were constructed using the Neighbor-Net method (Bryant and Moulton 2004) in SplitsTree v4.17.0 (Huson and Bryant 2006), with uncorrected *p*-distances and ambiguous states averaged, to visualise taxa relationships and sequence incompatibilities.

2.3.3 Plant health

Approximately 13 weeks after potting, plants were assessed for stress by checking photosynthetic efficiency. On 19 February 2020, all plants were removed from the chamber at 10.10 a.m. and placed in a dark room to be acclimated for 30 min. Growth chamber settings meant the plants had been exposed to light for 2 hours 45 min at this time. A Pocket-PAM Quantum Yield Analyser (Heinz Walz GmbH, Effeltrich, Germany) was used to make measurements of Fv/Fm (Kitajima and Butler 1975), or maximum quantum yield of photosystem II (PSII) in dark-adapted leaves. Pocket-PAM settings were: Light setting = 6, Interval = 600 milliseconds, Leaf factor = 0.84. Plants were measured in a random order. Generally, two leaves were measured per plant, and two measurements were taken per leaf. Plants were then watered and replaced in the growth chamber. Two hours 40 minutes later, the plants were again placed into a dark room, dark adapted for 30 min, and the Fv/Fm measurements repeated.

2.3.4 Plant harvest

All plants were harvested, and tissue stored, after three months of standardised growth conditions. Harvest took place on 27 February 2020. Plants were removed from the growth

chamber and processed in random order. Using a scalpel, each plant was separated into distinct tissue types: mature leaves, immature leaves, and roots. Each tissue type was placed into a 50 mL Falcon tube and immersed in liquid nitrogen. The snap-frozen tissue was then transferred to a freezer and stored at minus 80 °C.

To assess how water usage varied across samples, final potting mix moisture levels were calculated at the time of plant harvest. After each plant was processed, the remaining potting mix was placed into a sealable plastic bag and stored at 4 °C. Once harvesting was completed, the potting mix was weighed before being oven-dried for 48 hours at 80 °C. Samples were then reweighed. Gravimetric water content (GWC) at the time of harvest was calculated by dividing mass of water lost by the mass of dry soil. The GWC was then converted to VMC (the metric collected by the soil moisture data loggers) by multiplying GWC and potting mix BD.

2.4 RNA extraction and sequencing

2.4.1 RNA extraction

Extraction of total RNA was carried out on mature leaves of 12 individual plants representing the four taxa. A minimum of three biological samples is typically required for making population-level inferences of gene expression differences (Conesa *et al.* 2016). Neither immature leaf tissue nor new root tissue was available for *R. crithmifolius.* From the four individuals of Porters Pass *R. monroi* with the same chloroplast haplotype, and the individuals of the remaining taxa, three plants of each taxon were randomly selected as representative.

The RNA extraction procedure mainly relied on the E.Z.N.A.® Plant RNA Kit (Omega Bio-tek, Norcross, USA) with alterations to the manufacturers protocol necessary to successfully process these alpine *Ranunculus*. Primarily, the alterations consisted of bead-beating with different sized beads to rapidly pulverise tough plant tissue, the addition of a reagent to alleviate the inhibitory effects of plant tissue secondary compounds, and additive to overcome buffer foaming. Samples (four at a time) were removed from the freezer and placed into liquid nitrogen until needed. Extraction buffer consisting of 500 µL RB Buffer, 11 µL 2-Mercaptoethanol (Bio-Rad Laboratories, Hercules, USA), 50 µL PSS Solution (Qiagen, Hilden, Germany), and 3 µL Reagent DX (Qiagen) was pipetted into a 2 mL Micro Tube (Sarstedt AG & Co, Nümbrecht, Germany) into which one 5 mm stainless steel bead (Qiagen) and six 2.3 mm zirconium beads (BioSpec Products, Bartlesville, USA) had been placed. Once the tubes were chilled on ice, approximately 50 mg of tissue was added to each and immediately disrupted in a MagNA Lyser (Roche, Basel, Switzerland) for 30 s at 6,000 rpm, followed by 60 s at 5,000 rpm. The stainless steel beads were removed, and the tubes incubated for 5 min at

room temperature. Lysate was transferred to Homogenizer Mini Columns and centrifuged for 5 min at 14,000 *g*. The cleared lysate was then transferred to Axygen® 1.7 mL RNase and DNase-free Microcentrifuge tubes (Corning, Glendale, USA) and 1 volume of 70% ethanol added to each. Tubes were vortexed at maximum speed for 20 secs, and 600 µL of sample at a time transferred to HiBind® RNA Mini Columns by centrifuging for 1 min at 12,000 *g* until all sample was transferred. A 500 µL volume of RNA Wash Buffer I was added to each binding column and centrifuged at 10,000 *g* for 30 s. Another wash step was carried out three times by adding 600 µL of RNA Wash Buffer II and again centrifuging for 30 s at 10,000 *g*. The empty binding column was then centrifuged at maximum speed (16,000 *g*) for 2 min. To elute the RNA, 71 µL of Invitrogen™ Nuclease-Free Water (Thermo Fisher Scientific) was pipetted directly onto the binding columns and incubated for 1 min at room temperature before being transferred to 1.7 mL microcentrifuge tubes by centrifuging for 1 min at maximum speed.

To remove residual genomic DNA contamination from the RNA eluate, a DNase Max Kit (Qiagen) was used following the manufacturer's instructions. To the RNA eluate, 9 µL of DNase I solution was added prior to incubation for 10 min at 37 °C. An 8 µL aliquot of DNase Removal Resin was then added and the Resin kept in suspension by gentle agitation for 10 min at room temperature. After pelleting the Resin by centrifuging for 1 min at 13,000 *g*, the remaining sample was divided into 30 µL for RNA sequencing, 12 µL for quantification and quality checking, and the remainder (approximately 30 µL) retained for storage. RNA quantification was carried out using the Invitrogen™ Qubit™ RNA BR Assay Kit (Thermo Fisher Scientific) and an Invitrogen™ Qubit™ 2.0 (Thermo Fisher Scientific). Contamination from proteins (260/280 ratio) and other contaminants (260/230 ratio) was assessed on a NanoDrop One (Thermo Fisher Scientific). Lastly, samples were checked for RNA degradation on a LabChip® GX Touch HT (PerkinElmer LAS, Llantrisant, UK) operated by the Massey Genome Service.

2.4.2 RNA sequencing

RNA samples were air-freighted, on dry ice, to Novogene (Hong Kong, China) for strand-specific, paired-end mRNA sequencing as recommended by Conesa *et al.* (2016). Messenger RNA was enriched using oligo(dT) beads. Then, fragmentation buffer was added to randomly fragment the transcripts before cDNA synthesis using random primers. Second-strand synthesis substituted dUTPs for dTTPs. Transcript end repair, adenine tailing, and sequencing adaptor ligation was carried out before the second strands were subjected to uracil–DNA–glycosylase degradation. A size selection step and PCR enrichment completed the sequencing library preparation. All samples were pooled to equal concentrations and sequencing was

carried out on a single lane of a NovaSeq™ 6000 (Illumina, San Diego, USA) with expected output of ≥ 60 million paired-end reads (150 bases long) per sample.

For clarity and ease of analyses, sequence read files were renamed prior to any processing. Table 2.1 shows the designations given to each sample.

Table 2.1 Plant IDs and samples names used for RNA analyses

Plant ID	Taxon	Analysis ID
C4	*R. crithmifolius*	*R. crithmifolius* _1
C7	*R. crithmifolius*	*R. crithmifolius* _2
C8	*R. crithmifolius*	*R. crithmifolius* _3
M2	Mount Hutt *R. monroi*	Mt Hutt-*R. monroi* _1
M10	Mount Hutt *R. monroi*	Mt Hutt-*R. monroi* _2
M1	Mount Hutt *R. monroi*	Mt Hutt-*R. monroi* _3
M3	Porters Pass *R. monroi*	Porters Pass-*R. monroi* _1
M7	Porters Pass *R. monroi*	Porters Pass-*R. monroi* _2
M12	Porters Pass *R. monroi*	Porters Pass-*R. monroi* _3
L7	*R. lobulatus*	*R. lobulatus* _1
L5	*R. lobulatus*	*R. lobulatus* _2
L6	*R. lobulatus*	*R. lobulatus* _3

2.4.3 RNA sequence read quality

Sequencing read quality was assessed prior to, and after, quality filtering by using FastQC v0.11.9 (Andrews 2010) (/nesi/fastqc_raw_reads.sl; /nesi/fastqc_trimmed_reads.sl). Rcorrector v1.0.4 (Song and Florea 2015) was used in accordance with best practices recommended by Freedman (2016) to correct or flag any RNA sequencing reads containing erroneous kmers (/nesi/Rcorrector.sl). Rcorrector appends information to each header in the FASTQ files. To remove this information, which can cause issues with later processes (Freedman 2016), and remove any read pairs for which one read was deemed unfixable by Rcorrector, the script FilterUncorrectabledPEfastq.py from the Harvard Informatics GitHub repository TranscriptomeAssemblyTools (https://github.com/harvardinformatics/TranscriptomeAssemblyTools) was used (/nesi/filter_corrected_reads.sl). The final step in sequencing read filtering utilised TrimGalore! v0.6.4 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). TrimGalore! was set to remove Illumina adapter contamination from reads, and trim bases from read ends that fell below a threshold of Phred quality score 20 (/nesi/trimgalore.sl).

2.5 Transcriptome assemblies

2.5.1 Assembly

To capture a set of contigs representative of each taxon, *de novo* transcriptome assemblies were constructed for each of the taxa by incorporating sequencing reads from all three samples (Haas *et al.* 2013). The Trinity v2.11.0 (Grabherr *et al.* 2011) platform was chosen for its accuracy in the reconstruction of transcripts (Telfer *et al.* 2018; Hsieh *et al.* 2019; Zhao *et al.* 2019) and for the suite of integrated analysis tools that come packaged. Trinity has performed well in benchmarking and provides consistent coverage under different conditions (Zhao *et al.* 2011; Hölzer and Marz 2019). Also, in studies of plant gene expression, Trinity is reportedly used with success (Liu *et al.* 2018; Xu and Huang 2018; Ye *et al.* 2018), even when constructing transcriptomes derived from polyploid genomes (Zhou *et al.* 2019). Trinity clustering of contigs into putative genes was also found to outperform Corset v1.09 (Davidson and Oshlack 2014) clustering of rnaSPAdes v3.14.0 (Bushmanova *et al.* 2019) *de novo*-assembled contigs (/nesi/rnaspades.sl; /nesi/salmon_corset.sl; /nesi/corset.sl) in preliminary testing conducted for this study. Trinity parameters were set with strand-specific library type as 'RF' and minimum contig length of 300 bases. Additionally, to increase the speed of assembly, the Trinity run was split into two phases. Phase 1: processes which require multiple threads and large amounts of RAM (/nesi/trinity_phase_1.sl), and phase 2: processes requiring a single compute thread and small amounts of RAM, which can be batch processed (/nesi/trinity_phase_2.sl; /nesi/SLURM.conf). Subsequent Trinity assemblies for individual samples were constructed with the same parameters excepting reductions in compute resources. Throughout this text, assemblies constructed by incorporating all taxon sample reads are called aggregated assemblies (Table 2.2). Trinity assemblies constructed from individual sample reads are denoted individual assemblies (Table 2.2).

Table 2.2 Assembly type nomenclature	
Assembly	Assembly Type
R. crithmifolius	Aggregated
Mt Hutt-*R. monroi*	Aggregated
Porters Pass-*R. monroi*	Aggregated
R. lobulatus	Aggregated
R. crithmifolius _1	Individual
R. crithmifolius _2	Individual
R. crithmifolius _3	Individual
Mt Hutt-*R. monroi* _1	Individual
Mt Hutt-*R. monroi* _2	Individual
Mt Hutt-*R. monroi* _3	Individual
Porters Pass-*R. monroi* _1	Individual
Porters Pass-*R. monroi* _2	Individual
Porters Pass-*R. monroi* _3	Individual
R. lobulatus _1	Individual
R. lobulatus _2	Individual
R. lobulatus _3	Individual

2.5.2 Transcriptome validation

Assessing quality of the Trinity *de novo* assemblies consisted of several methods. Due to computational and time constraints, only validation of the aggregated assemblies utilised all these methods. However, basic information, such as total contig number and N50 (the length of the shortest contig in a set of the longest contigs containing 50% of all assembled bases), was gathered for all assemblies using the TrinityStats.pl script. For the aggregated assemblies, Bowtie2 v2.3.5 (Langmead and Salzberg 2012), using end to end alignment in 'sensitive' mode with strand-specific options enabled (/nesi/bowtie2.sl), aligned sequencing reads used for transcriptome assembly back to the completed assemblies to calculate the percentage of correctly mapped reads. This indicated how many reads were correctly incorporated into a given assembly. Additionally, transcriptome completeness was inferred by BUSCO v4.1.4 (Simão *et al.* 2015) through recovery of expected single-copy orthologous genes (SCOs) from the OrthoDB v10 (Kriventseva *et al.* 2018) Embryophyta database (/nesi/busco.sl).

2.5.3 Transcript quantification

Quantification of aggregated assembly transcripts was necessary for downstream DGE analyses and for selecting representative transcripts from each Trinity cluster. Trinity generates

multiple contigs or 'isoforms' that are clustered together into a 'gene' level. Trinity supports the running of RSEM v1.3.3 (Li and Dewey 2011) though the align_and_estimate_abundance.pl script. RSEM takes the read counts output from alignment software such as Bowtie2 and implements its expectation maximisation algorithm to handle ambiguously mapping reads (Li and Dewey 2011). RSEM then reports both transcript-level and gene-level counts. The Trinity align_and_estimate_abundance.pl script internally manages parameters for RSEM and Bowtie2, so these were run at default (strand-specific library option notwithstanding) when preparing the RSEM reference (/nesi/rsem_prepreference.sl) and carrying out quantification (/nesi/rsem_abundance.sl).

To run different types of analyses, transcript quantification was carried out multiple times. First, for pairwise DGE analyses, transcript quantification was performed by aligning reads against reference transcriptomes. The three transcriptomes were: *R. crithmifolius*, *R. lobulatus*, and Porters Pass-*R. monroi*. This involved preparing an RSEM reference for each of these taxa then aligning reads and estimating abundances for all 12 samples against these references. In other words, the RSEM abundance step was performed 36 times. Second, a Mt Hutt-*R. monroi* reference was prepared, and the Mount Hutt *R. monroi* samples used to estimate transcript abundances for this assembly. This meant RSEM estimation was carried out for all RNA-seq samples against the *R. crithmifolius*, *R. lobulatus*, and Porters Pass-*R. monroi* assemblies, and all samples were self-mapped against their own aggregated assemblies.

2.5.4 Gene functional annotation

To enable the functional annotation of genes and enable the searches for SCOs, it was necessary to choose a representative contig from each Trinity cluster. The software developers of Trinity recommend selecting the isoform with the highest expression in a cluster in preference to selection based on length (https://github.com/trinityrnaseq/trinityrnaseq/wiki), and this recommendation was adopted. For each aggregated assembly, the transcripts per million (TPM) (Wagner *et al.* 2012) normalised contig counts from RSEM were used to select the representative contigs. Expression matrices, each with transcript ID and TPM values, were generated for each aggregated assembly from its constituent self-mapped samples. This was accomplished by using R functions and packages within a Jupyter Notebook (Kluyver *et al.* 2016) (/nesi/rsem_TPM_sum.ipynb) on the JupyterLab v2.2.4 web interface. The Trinity script filter_low_expr_transcripts.pl, set to output the isoform with the highest TPM value from each cluster, was used to generate a FASTA file for each assembly containing only these representative contigs (/nesi/rsem_contig.sl). These filtered assemblies are referred to as representative assemblies (or transcriptomes) throughout the rest of this text.

To generate representative assemblies for the individual assemblies, a different approach was taken. Computing resource constraints meant a lightweight alternative to RSEM was needed. Salmon v1.3.0 (Patro *et al.* 2017) was used to self-map sample reads against each individual assembly to generate TPM-normalised transcript counts (/nesi/taxon_x_salmon.sl). Salmon was run with library type set as strand-specific, and with the 'gcBias' and 'validateMappings' settings enabled as recommended in the Salmon documentation (https://salmon.readthedocs.io/en/latest/salmon.html#using-salmon). Once the Salmon quantification was completed, lists of the most highly expressed contig per cluster for each assembly were generated once again using R within a Jupyter Notebook (/nesi/taxon_123_salmon_contig.ipynb). The programme seqtk v1.3 (https://github.com/lh3/seqtk) was then used to filter the initial individual assemblies by the lists of highly supported contigs to generate FASTA files of representative transcriptomes for the individual assemblies (/nesi/bash.html).

TransDecoder v5.5.0 (https://github.com/TransDecoder) was used to identify and extract protein coding sequences from the representative contigs of all assemblies. The programme was run in strand-specific mode and set to output only the longest open reading frame per contig (/nesi/transdecoder.sl).

Functional annotation of genes was carried out to gain an understanding of gene function. *Ranunculus lobulatus* was chosen as the representative taxon. Contigs from the *R. lobulatus* aggregated assembly and the *R. lobulatus*_1 individual assembly were annotated against the *Arabidopsis thaliana* (Brassicaceae) transcriptome. This transcriptome was chosen ahead of that for the more closely-related model organism *Aquilegia coerulea* (Ranunculaceae) because it performed better in preliminary analyses. Specifically, while basic local alignment search tool (BLAST) v2.10 (Altschul *et al.* 1990) searches of *R. lobulatus* contigs against the Phytozome (Goodstein *et al.* 2012) *A. coerulea* v3.1 protein sequences resulted in more hits, the resources available for *A. coerulea* are relatively poor, and fewer of these contigs could be annotated with gene ontology (GO) terms. To complete gene annotation, the protein sequences for *A. thaliana* from the Araport 11 genome release (Cheng *et al.* 2017), 'Araport11_genes.201606.pep.fasta.gz', were uploaded to NeSI in FASTA format. From these sequences a BLAST database was constructed. Peptide sequences, generated by TransDecoder, of the representative contigs were then searched against the database with the BLASTp function (/nesi/blast_arabidopsis.sl). A threshold of E-value 1e-05 was used to exclude poor BLAST matches, and a '-max_hsps' setting of one was used to output only a single matched alignment per BLAST hit. The single best BLAST hit per sequence was desired. Therefore, to prevent BLAST outputting the first merely good match (Shah *et al.* 2019), the '-max_target_seqs' parameter was set to five and subsequent filtering carried out to select matches with the lowest E-value (/local/lowest_evalue.R). It is worth noting, that the issue raised by Shah *et al*. (2019)

has subsequently been addressed, and later versions of BLAST are deemed to more reliably return the best hit from searches using '-max_target_seqs' set to one (Madden *et al.* 2019).

To investigate noted differences in the numbers of TransDecoder outputs between individual assemblies, peptide sequences of representative contigs for all *R. crithmifolius* and Mount Hutt *R. monroi* individual assemblies were searched against the *A. thaliana* protein sequences with BLASTp. Because number of hits, and not necessarily the absolute best hits, were of interest, '-max_hsps' and '-max_target_seqs' were set to one (/nesi/contig_number.sl). The number of unique matches to the *A. thaliana* proteins were then counted to determine if the increased number of TransDecoder peptide sequences was correlated with increased numbers of distinct genes or isoforms being expressed (/nesi/bash.html).

2.5.5 Chloroplastic sequences

Anomalies were noted in the length distributions of contigs, as well as in the numbers of gene clusters and predicted proteins. In particular, a number of very long contigs were detected in the Mt Hutt-*R. monroi* and Mt Hutt-*R. monroi*_3 transcriptomes. To determine the identity of these, the sequence of the longest contig from the Mt Hutt-*R. monroi* assembly was uploaded to the NCBI website and searched, using the BLASTn function, against the non-redundant nucleotide sequences database (Pruitt *et al.* 2007). Based on the BLASTn results, which suggested a chloroplast sequence, the TransDecoder predicted protein of this contig was searched against the NCBI non-redundant protein sequences database. Subsequently, sequencing reads of all samples were mapped to the *Ranunculus sceleratus* genome with Bowtie2, using '--very-sensitive-local' alignment (/nesi/sceleratus.sl) to quantify reads derived from chloroplast DNA (cpDNA).

2.6 Gene expression and phylogenetics

Phylogenetic patterns of gene expression were investigated. A criticism of traditional pairwise comparisons of differential gene expression analyses, when comparing between species, is that no allowance is made for phylogenetic relationships and population structure. Both population structure and phylogenetic relationships can contribute to inferences of differential expression (Koenig *et al.* 2013; Landis *et al.* 2013; Rohlfs *et al.* 2013; Rohlfs and Nielsen 2015). This observation can be leveraged to illuminate evolutionary relationships based on gene expression profiles and identify patterns of differential expression unexpected given these evolutionary relationships. Three different approaches to DGE analyses were taken. First, to investigate if standard approaches to gene expression analyses were applicable to the alpine

Ranunculus dataset, pairwise comparisons were made by aligning reads to a reference transcriptome. Second, more explicit testing for phylogenetic signals in the gene expression patterns was undertaken for genes of the reference transcriptome. Third, to alleviate potential read alignment biases, transcript quantification was carried out on SCOs common to all taxa. Of primary interest in this study was the relationship of Mount Hutt *R. monroi* to the remaining three taxa. Therefore, DGE analyses requiring reference transcriptomes were carried out using the *R. crithmifolius*, *R. lobulatus*, and Porters Pass-*R. monroi* assemblies. Because gene expression analyses have been demonstrated to be more accurate when conducted at the gene rather than transcript level (Soneson *et al.* 2015; Conesa *et al.* 2016), all analyses were conducted at the gene level.

2.6.1 Standard pairwise comparisons of gene expression

The effect of read alignment biases arising from mapping to a reference were investigated before carrying out DGE analyses. The RSEM outputs for each sample were filtered (/nesi/rsem_TPM_gene.ipynb) to produce files that contained only the Trinity gene IDs with associated TPM values. Then, an assessment of mapping biases was made to inform the pairwise comparisons of gene expression. Accordingly, for the *R. crithmifolius*, *R. lobulatus*, and Porters Pass-*R. monroi* assemblies, heat maps were generated using hierarchical clustering of the normalised (TPM) gene expression values (/local/heatmap.R). Importantly, TPM is a more robust method of RNA-seq data normalisation than the commonly-used reads/fragments per kilobase of transcript per million reads mapped (RPKM/FPKM) methods when comparing between samples (Wagner *et al.* 2012; Abrams *et al.* 2019; Zhao *et al.* 2020). Computer memory restrictions prevented production of heatmaps representing all genes so filtering was carried out using a Kruskal-Wallis test to retain only genes that showed significantly different (at $p < 0.05$) expression in at least one taxon. To investigate potential read alignment biases at the individual level, RSEM transcript quantification against the Porters Pass-*R. monroi*_1 and Mt Hutt-*R. monroi*_1 assemblies was carried out. All parameters for RSEM were maintained as above. Utilising a reference assembly taxonomically equidistant from the study taxa was also investigated. However, preliminary read alignments against the Phytozome *Aquilegia coerulea* v3.1 transcriptome produced poor mapping statistics.

Subsequent pairwise expression analyses were carried out with RSEM gene counts from reads aligned to the *R. lobulatus* assembly. Anomalous gene expression levels resulting from biases due to heterospecific read alignment were a consideration (Voelckel *et al.* 2012). However, based on the alpine *Ranunculus* taxonomy (Fisher 1965; Webb *et al.* 1988), inferences of shared expression levels between Mount Hutt *R. monroi* and *R. crithmifolius* were

assumed less likely to be influenced by these biases when using a *R. lobulatus* reference than would be the case in using either Porters Pass *R. monroi* or *R. crithmifolius*. DESeq2 v1.30.1 (Love *et al.* 2014) was used to carry out the DGE analyses (/local/deseq2.R). Many tools are available for differential expression analysis. However, it has been asserted that software such as DESeq2 and edgeR (Robinson *et al.* 2010), which assume a negative binomial distribution to model count data, outperform software using different models of RNA-seq count data (Finotello and Di Camillo 2015). Benchmarking studies have validated the performances of these two programmes (Costa-Silva *et al.* 2017; Lamarre *et al.* 2018), particularly when testing fewer than 12 biological replicates (Schurch *et al.* 2016).

Three DGE comparisons were made. First, gene expression levels of the two *R. monroi* taxa were compared with *R. crithmifolius* and *R. lobulatus*. Second, Mount Hutt *R. monroi* and *R. crithmifolius* were grouped together and compared with Porters Pass *R. monroi* and *R. lobulatus*. Third, Mount Hutt *R. monroi* and *R. lobulatus* were compared against Porters Pass *R. monroi* and *R. crithmifolius*. Because DESeq2 performs internal normalisation, raw gene counts were imported into DESeq2 with tximport v1.18.0 (Soneson *et al.* 2015). Genes were considered to be differentially expressed if the *p* value, corrected for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg 1995), was below a threshold of 0.05. Levels of fold change in expression were not considered since similarities of gene expression, rather than magnitudes of difference, were being tested.

A filtering step was required to overcome apparent issues arising from using a single reference. The initial DGE analyses were carried out by excluding any genes that didn't have at least 10 reads aligned. Yet this filtering appeared insufficient. Therefore, data were re-analysed by retaining only genes which had at least one read aligned by each of the 12 samples.

2.6.2 Phylogenetic analyses of reference assembly expression levels

Investigation was undertaken to determine if phylogenetic signals were contained in the gene expression data derived from aligning reads to the *R. lobulatus* reference (/local/rank_reference.R). Unlike using DESeq2, this method did neither internally account for library sizes nor transcript lengths. Therefore, the data used were normalised gene TPM values contained within the 'all_gene_TPM.tsv' file generated by /local/heatmap.R. Each row of the dataframe (corresponding to a gene) was input into a loop. Sample TPM values were ranked from lowest to highest. Then, the 6th and 7th values (the middle two values) were compared, and if these were identical, the gene was rejected on the basis that the 12 samples could not be resolved into two distinct and equally sized groups. For any gene not rejected, occurrences of the different taxa, in the lowest six rankings were counted. If three or more taxa were counted

then, again, the gene was rejected for not displaying two clear taxonomic groupings. If only two taxa were counted, then those taxa were deemed to share similar expression levels for that gene and the gene was added the appropriate cluster. For example, if the three Mount Hutt *R. monroi* and three Porters Pass *R. monroi* samples, or the three *R. crithmifolius* and three *R. lobulatus* samples were counted in the six lowest rankings, then this gene was determined to cluster Mount Hutt *R. monroi* with Porters Pass *R. monroi*.

3.6.3 Single copy orthologue detection

Reference-free gene expression analyses, and phylogenetic tree building based on DNA sequence evolution, meant identification of SCOs was required. Orthologues are genes present in different species, which have evolved from an ancestral gene after speciation. This contrasts with paralogous genes which arise through gene duplication (Fitch 1970). Importantly, gene duplications occurring after speciation can lead to multiple copies of orthologous genes (co-orthologues), so there might be one to many, or many to many (co-)orthologues (Gabaldón and Koonin 2013) as depicted in a simplified manner in Figure 2.4. Duplication can lead to changes in evolutionary constraints on co-orthologues (Johnstun *et al.* 2021). These changes can violate the common evolution process assumed for orthologues, which has been the foundation of model-based phylogenetic inference (Steel and Penny 2000). The consequence of this model violation is that it can mislead phylogenetic reconstruction (Lockhart *et al.* 1996). The relaxation of selective pressures also facilitates increased sequence and functional diversity among paralogous gene sets (Hahn 2009; Lafond *et al.* 2018; Johnstun *et al.* 2021). On the other hand, sequence conservation and functional conservation is common for 1:1 (single-copy) orthologues even between evolutionarily distant organisms (Han *et al.* 2014; Kachroo *et al.* 2015). One approach to SCO finding utilises BLAST to identify reciprocal best BLAST hits (Moreno-Hagelsieb and Latimer 2008). Briefly, if transcript sequences from two databases report one another as the best hit then these are categorised as reciprocal best BLAST hits. However, this approach becomes more difficult to implement as the number of taxa increases (Sanders and Cartwright 2015), necessitating a more computationally efficient approach.

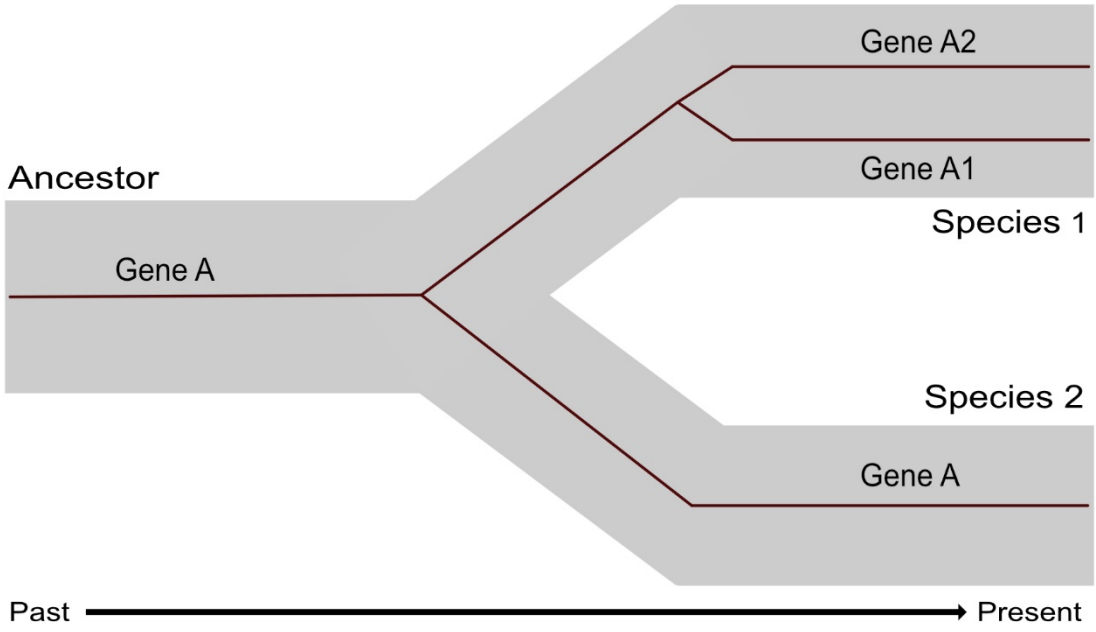


Figure 2.4 Orthologue evolution. This figure represents orthologous genes arising in sister taxa following divergence of species. After lineage splitting, Gene A (which was present in the ancestral species) is found in both Species 1 and Species 2. Duplication of Gene A in Species 1 produces genes (A1 and A2) which are paralogous. Both A1 and A2 are orthologous with respect to Gene A in Species 2 but are not single-copy orthologues. Duplication can cause selective pressure to be relaxed for either A1 or A2, which means models used in analyses of orthologues would poorly fit these sequence data.

OrthoFinder (Emms and Kelly 2019), noted for its accuracy (Nichio *et al.* 2017), has been proposed as an alternative approach to orthologue detection and was used in the present study to detect SCOs common to all taxa. For the aggregated assemblies, this meant genes common to the four taxa. For the individual assemblies, the SCOs needed to be present in all 12 samples. To trace each assemblies' genes throughout the process, assembly identifiers were first appended to each TransDecoder predicted nucleotide coding sequence and its corresponding peptide sequence using bioawk 1.0 (https://github.com/lh3/bioawk) (/nesi/bash.html). Because early testing showed orthologue detection using nucleotide sequences resulted in fewer SCO clusters, OrthoFinder v2.4.1 was run on the predicted protein sequences using default settings. OrthoFinder output a single FASTA file, containing the peptide sequences, for each SCO cluster. Further steps were then required to generate the same cluster files with nucleotide sequences. For both the aggregated and individual assemblies, all TransDecoder nucleotide ('.cds') FASTA files were merged together (/local/bash.html). Next, the /local/extract_cds.sh bash script was used to generate orthologue clusters containing nucleotide sequences. Briefly, for each SCO cluster, this script searched for gene identifiers in the merged nucleotide sequences file and saved the found sequences into another file.

2.6.4 Phylogenetic analyses of aggregated assembly SCO expression levels

To mitigate issues with heterospecific read mapping, gene expression analyses were carried out only on genes identified as SCOs. To do so, expression values for the SCOs had to be produced. First, for each of the aggregated assemblies, the SCO cluster files output by OrthoFinder were parsed for the assembly name and then a list of SCOs was generated that had the cluster identifier (file name) appended (/local/bash.html). Next, using the R script /local/sco_TPM.R, the orthologue transcript IDs were searched for in the RSEM gene-level TPM outputs generated by /nesi/rsem_TPM_gene.ipynb, to produce a file of sample TPM values for each SCO.

A heat map was constructed of hierarchically clustered normalised (TPM) expression values for all SCOs (/local/sco_TPM.R). Subsequently, expression levels of each SCO TPM value were tested using a Kruskal-Wallis test, and only SCOs in which at least one taxon was significantly differentially expressed (at $p < 0.05$) were retained. From these significantly different genes, another heatmap was generated (/local/sco_sigdiff.R).

Expression similarities between taxa were examined using the same ranking approach taken with the *R. lobulatus* reference mapping data (/local/rank_sco.R). However, because the SCOs were found in all taxa, no pre-filtering to ensure all samples had reads aligning was required.

2.6.5 Phylogenetic analyses of aggregated assembly SCO nucleotide variation

Nucleotide sequences of the SCOs were used to construct unrooted phylogenetic trees, representing the most strongly supported relationships among orthologues, from the aggregated assemblies. To process the large number of SCO clusters, most processes were carried out in 'for' loops. First, to ameliorate issues with aligning sequences of different lengths, each SCO FASTA file was processed to remove any sequence with length shorter than 70% of the longest sequence. The perl script /local/Percent.pl (script credit to Leonna Szangolies) produced new FASTA files with any short sequences removed. These files were searched and any with fewer than four sequences remaining were deleted (/local/bash.html). Multiple sequence alignments were carried out using MUSCLE v3.8.1551 (Edgar 2004) with default parameters (/local/bash.html). To then remove gaps from the MSAs, trimAL v1.4 (Capella-Gutiérrez *et al.* 2009) was run using the '-nogaps' flag (/local/bash.html). After gap removal, another filtering step was undertaken to remove any SCO sequences comprised of fewer than 200 nucleotides (nt) or with no nucleotide variation between taxa (/local/bash.html). Because later assessment of tree topologies required standard naming conventions, Trinity gene names were removed from

the FASTA headers of the SCO files, so all headers comprised only assembly names (/local/bash.html).

Phylogenetic tree building was carried out on the final, filtered MSAs. The maximum likelihood (ML) phylogenetic inference software RAxML-NG v0.9.0 (Kozlov *et al.* 2019) was used to construct phylogenetic trees. This approach was chosen for tree building as ML is one of the most reliable methods for reconstructing unrooted trees from site pattern data (Swofford *et al.* 2001). There were few taxonomic relationships to reconstruct, and nucleotide substitution model selection using the Akaike information criterion (AIC) (Bozdogan 1987; Posada and Buckley 2004) on such a large number of MSAs was impractical, so ML tree searching was performed from a single random start tree using the HKY85 (Hasegawa *et al.* 1985) model (/local/bash.html). One hundred rounds of non-parametric bootstrapping were performed on the MSAs, and support for the ML trees computed by mapping the bootstrap values onto the ML trees. The R script /local/cluster_aggregated.R was then used to filter and cluster the trees. Trees with low confidence might have introduced error into the analyses, so trees with bootstrap values of less than 70% for internal edges were rejected. Although this approach would filter out datasets with low numbers of informative sites that were not contradictory, it provided a conservative means for retaining high-quality trees. Following this step, the retained ML trees had edge lengths removed and were grouped according to their unrooted topology (Figure 2.5, A). This was done based on the clustering of taxa in the Newick file outputs from RAxML-NG. Branch length removal was necessary because in preliminary analyses using ward.D2 clustering (Ward Jr 1963) using tree-to-tree distances that incorporated branch lengths (Kuhner and Felsenstein 1994), trees with different taxonomic relationships were assigned to the same cluster. In contrast, clustering of unweighted trees, based on parsing each gene Newick file, gave the same result as ward.D2 clustering of trees after calculation of tree distances without branch lengths (Robinson and Foulds 1981; Penny and Hendy 1985); a method which produced correct clusters.

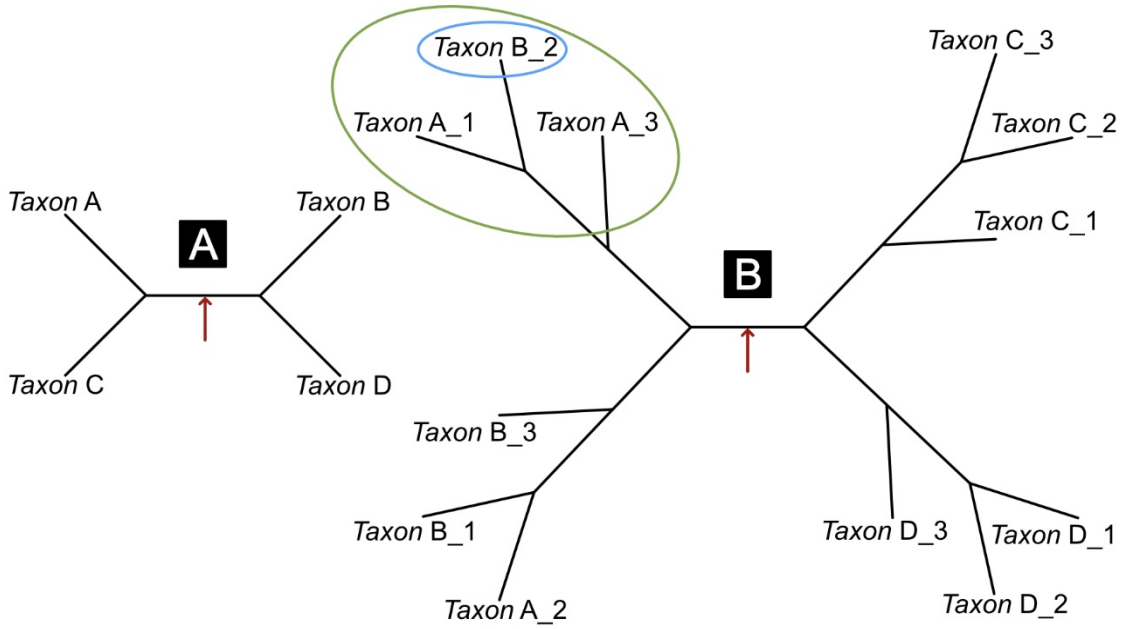


Figure 2.5 Unrooted and unweighted phylogenetic gene trees of nucleotide variation. A, four-taxon (aggregated assembly) tree. B, 12-taxon (individual assembly) tree. Both trees are indicative of maximum likelihood trees generated by RAxML-NG after removal of edge weights. Red arrows illustrate position of the central edge for each tree. Tree A shows a gene tree in which Taxon A and Taxon C comprise a taxonomic grouping or split. Tree B is indicative of a tree with clear taxonomic relationships. All samples of Taxon A and Taxon B are positioned on one side of the central edge, and as such comprise a taxonomic split deemed useful in this study. However, the taxa in Tree B are not monophyletic. Coloured circles indicate a situation in which a sample from Taxon B clusters with samples of Taxon A. In these analyses, such a situation would not be detected.

2.6.6 Phylogenetic analyses of individual assembly SCO nucleotide variation

Taxonomic relationships between all samples were investigated by generating ML trees for SCOs of the 12 individual assemblies. Maximum likelihood trees were built in RAxML-NG with the same parameters used for analysing SCOs from the aggregated assemblies. Again, trees with bootstrap values below 70% were excluded. Utilising 12 samples per gene complicated the analyses. For 12 taxa, the possible number of unrooted per trees per gene = (2n – 5)!! = 654,729,075 (Penny *et al.* 1982). The size of the problem required alterations to the approach used for the aggregated assemblies SCOs. This was performed by interrogating the Newick file for each tree to deduce whether the unweighted tree contained groups of two taxa at either side of the tree as shown in Figure 2.5 (Tree B). Only sample position either side of a tree's central

edge was considered in the filtering. For example, in Tree B (Figure 2.5) all samples of Taxon A and Taxon B occur to one side of the central edge and would be considered a clear taxonomic split. However, intermixing (or non-monophyly) of taxa at one end of the tree, as shown in Tree B (Figure 2.5), was not tested for. Because the convention in Newick notation is to arbitrarily root an unrooted tree (Felsenstein n.d.), it was necessary to root each tree output from RAxML-NG at the midpoint of the unweighted tree for identification of the central edge. This required resetting all edge lengths to one, mid-point rooting the tree, removing the edge lengths, then writing out the unweighted tree. Subsequently, the Newick notation of each unweighted tree was examined for having the correct topology: a bifurcating tree with six samples either side of the central edge. For trees with only two taxa on one side of the central edge, the tree was classified according to taxa groupings (or splits), while remaining trees were rejected (/local/cluster_individual.R).

2.6.7 Gene ontology term annotation

Functional annotation of genes allows phenotypic and physiological inference of genetic data. The Gene Ontology resource (Gene Ontology Consortium 2004, 2019) provides a means to assign function and relationships of gene classes. To perform GO term enrichment analyses, genes were annotated with associated GO terms. Araport 11 gene isoform identifiers annotated to the TransDecoder predicted peptide sequences of both the *R. lobulatus* and *R. lobulatus*_1 representative assemblies were uploaded to the UniProt website (UniProt Consortium 2018) on 3 April 2021. Gene ontology terms associated with the *A. thaliana* genes were then retrieved via the 'Retrieve/ID Mapping' function. The Araport 11 identifiers, annotated with GO terms, were downloaded in tab separated text documents.

For GO term annotation of genes from the reference assembly DGE (both pairwise and ranking) analyses, lists of genes for each were first annotated with Araport 11 isoform identifiers (/local/araport_ref.R), before these lists were used to produce further lists of genes annotated with GO terms (/local/go_ref.R). The background (see Section 2.5.6) generated for both the differentially expressed genes output from DESeq2 as well as genes clustered using the expression ranking system comprised genes mapped with reads from all 12 samples. Genes of interest, i.e. the three groups of genes that constituted the different taxa splits, were annotated in the same way (/local/araport_ref.R; /local/go_ref.R).

Annotating genes of the SCO analyses with GO terms was carried out similarly to genes from the reference assembles. However, in these cases, the GO terms were annotated to the OrthoFinder gene/cluster IDs. The background genes for these analyses were either all SCOs

output by OrthoFinder for expression analyses (/local/araport_sco_exp.R) or SCOs that passed the filtering steps and were used for phylogenetic tree building (/local/araport_sco_nuc.R).

2.6.8 Gene ontology term enrichment analyses

Gene ontology term enrichment analyses were carried out for each type of gene expression and nucleotide variation analysis to statistically test if gene sets could potentially contribute to physiologies important in habitat adaptation. Lists of background genes and test sets were uploaded to agriGO v2 (Tian *et al.* 2017). Singular Enrichment Analysis (SEA) was carried out with a minimum number of mapping entries set to five. Using the Fisher statistical method (Fisher 1992) with Yekutieli (Benjamini and Yekutieli 2001) false discovery rate correction, the significance level was set at $p < 0.05$.

2.6.9 Function of overlapping gene sets

Single-copy orthologues were identified whose pattern of nucleotide variation supported gene trees with splits also corroborated by patterns of differential expression. The function of these genes was then characterised. Genes occurring in the Mount Hutt *R. monroi–R. crithmifolius* grouping for both the SCO expression analyses and the aggregated assembly SCO phylogenetic tree analyses were investigated. First, for all three splits, genes occurring in output lists from both analyses were extracted (/local/bash.html). Subsequently, sequences of genes from the Mount Hutt *R. monroi–R. crithmifolius* grouping, common to both analyses, were uploaded to the National Center for Biotechnology Information (NCBI) (Sayers *et al.* 2019a) website and the BLASTp function used to search the non-redundant protein sequences database (Pruitt *et al.* 2007). The protein giving the top hit (Max Score) was recorded for each gene. A graphical overview of the bioinformatic pipeline leading to the identification of SCOs common to both gene expression and nucleotide variation analyses is provided in Supplementary Figure 1.

2.6.10 Analyses of heterozygosity in the individual assemblies

To better understand the amount of allelic variation in the samples, an examination of heterozygotic SNPs within each individual assembly was made. Based on the recommendation of Yao *et al*. (2020) for variant calling in hexaploid wheat (*Triticium aestivum* L.), a pipeline using BWA-MEM v0.7.17 (Li 2013) and the mpileup function of BCFtools v1.9 (Li 2011) was

used. For each assembly, its RNA-seq reads were aligned against the representative assemblies with BWA-MEM using default parameters, and a sorted binary sequence alignment/map format (BAM) file reporting read alignments was output with SAMtools v1.10 (Li *et al.* 2009) (/nesi/bwa.sl). Lists of filtered SCOs used for ML tree building, for each assembly, were then generated (/local/bash.html). The /local/snp.sh script was then called to perform read filtering, variant calling, and variant filtering. In this script, SAMtools v1.11 was used to extract from the BAM file only concordantly mapped read pairs. Then, mpileup provided a summary of the coverage of mapped reads for each nucleotide of the filtered SCO contigs in a binary variant call format (BCF) file. The BCFtools 'call' function was then used to perform variant calling with '-m' flag set to include invariant sites in the output variant call format (VCF) file. Stringent criteria for the inclusion of SNPs were implemented as recommended by Telfer *et al*. (2018). Initial filtering of variant sites was carried out with BCFtools vcfutils.pl 'varFilter' set to retain only variant sites with a minimum coverage of 10 reads and a maximum coverage of 100 reads. A final round of filtering used the programme VCFtools v0.1.16 (Danecek *et al.* 2011) set to retain variant sites with two alleles and a threshold for the frequency of the minor allele of 0.25. Variant site statistics, including heterozygous SNP site frequency, were generated with 'vcfstats' function of RTG Tools v3.12 (Real Time Genomics, Hamilton, New Zealand).

Chapter 3: Results

3.1 Habitat characterisation

3.1.1 Physical and vegetation factors

The physical characteristics of the sites were, at first glance, not dissimilar (Table 3.1). Porters Pass was at the lowest altitude, at approximately 350 m below the elevation of the two Mount Hutt sites. All sites were on the faces of slopes with aspects from approximately south-southwest (Lake Tennyson) to southwest by west (Mount Hutt *R. crithmifolius* site). The slope at Porters Pass was much less severe than at the other three sites. All sites had some degree of rock or gravel present, and all sites appeared well drained. Notably, at Mount Hutt, the open scree of the *R. crithmifolius* site was much less stable than the face of the rock spur upon which the *R. monroi* was growing.

Table 3.1 Physical and vegetation characteristics of sampling sites

	Site			
	Mount Hutt *R. crithmifolius*	Mount Hutt *R. monroi*	Porters Pass	Lake Tennyson
Altitude (m)	1348	1344	998	1250
Physiography[a]	Face	Face	Face	Face
Aspect (0 - 359 °)	240	230	210	195
Slope (°)	37	40	20	35
General material description[b]	Scree	Rocky	Rocky and silty	Scree
Observed drainage[c]	Good	Good	Good	Good
Surface characteristics				
Bedrock (%)	5	60	0	0
Broken rock (%)	95	40	30	100
Size of broken rock	Mainly 1 - 20 mm with scattered larger fragments	Up to 10 cm	2 - 10 cm	1 - 10 cm
Ground cover				
Vascular plants (%)	5	10	50	55
Non-vascular plants (%)	< 1	15	40	5
Litter (%)	0	0	0	0
Bare soil (%)	0	20	0	0
Rock (%)	94	55	10	40
Average top height (cm)[d]	< 1	< 1	70	2 - 3

[a] Ridge, Face, Gully, Terrace

[b] Scree, Rocky, Silty, Other:

[c] Good, Moderate, Poor

[d] Average height of the dominant vegetation

It was in vegetation that the sites appeared more obviously variable. There was very little vegetation at the Mount Hutt *R. crithmifolius* site which was mainly comprised of loose rock (Table 3.1; Figure 3.1). The Mount Hutt *R. monroi* site had a greater proportion of vegetation covering the ground. Plant coverage at Porters Pass was much more extensive, with increased vegetation height, compared with other sites, although Lake Tennyson ground cover included a similar proportion of vascular plants (Table 3.1; Figure 3.1). Interestingly, the *R. monroi* at Porters Pass was observed to typically occur on the vegetation fringes, where shelter was abundant, but competition potentially reduced.

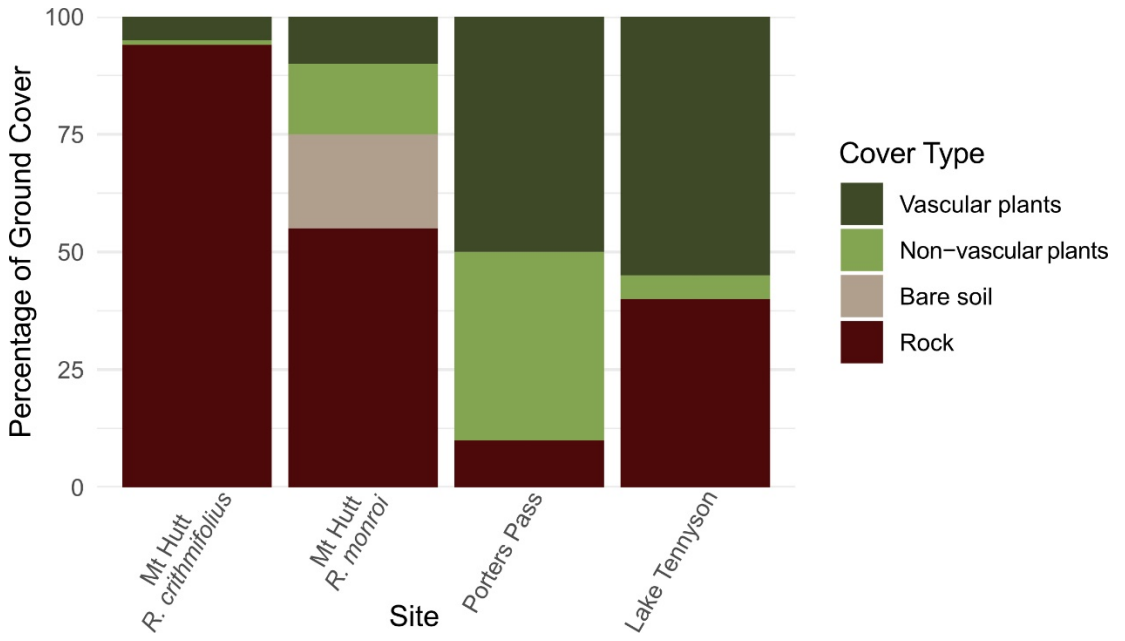


Figure 3.1 Ground cover at sampling sites. Estimated percentages of ground cover types at each of the sampling sites is indicated by coloured blocks within the stacked bar chart. No vegetation litter was recorded at any site.

3.1.2 Soil moisture

Soil moisture profiles varied at each of the sample sites. The profiles at Porters Pass and Lake Tennyson showed increasing soil moisture levels from late January 2020 until approximately the middle of the following November (Figure 3.2). Soil moisture levels at Mount Hutt remained relatively consistent throughout the main monitoring period. Notably, moisture readings typically varied widely between probes at each site, with large standard deviations from the mean (Figure 3.2; /supplementary/MH.csv; /supplementary/PP.csv; /supplementary/LT.csv). However, even considering the variation of readings between probes, each site displayed distinct profiles (Figure 3.2). Only for short periods of monitoring were the means within one standard deviation (1 x *SD*) from each other. Almost no overlap in 1 x *SD* from the mean moisture levels occurred between Mount Hutt and Porters Pass.

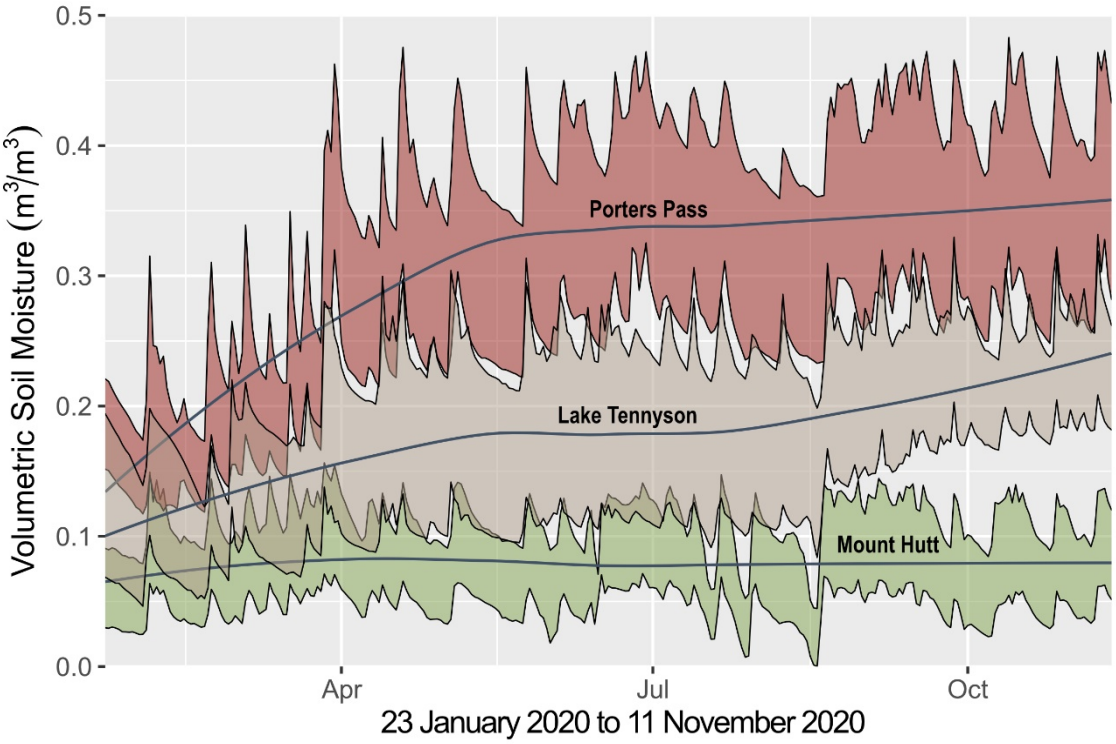


Figure 3.2 Soil moisture at three sample sites. One standard deviation either side of mean daily soil moisture, measured using four electrical conductivity probes, is indicated by coloured bands. Porters Pass, Lake Tennyson, and Mount Hutt *R. crithmifolius* sites are shown in red, brown, and pale green respectively. Dark grey lines represent mean daily soil moisture smoothed using locally weighted scatterplot smoothing (LOESS).

To test for difference between the daily means of Porters Pass and Mount Hutt, a two-tailed Wilcoxon signed rank test was performed. The test indicated that mean daily soil moisture levels were significantly different between Porters Pass and Mount Hutt (V = 43,660, $p < 0.001$). Additionally, at Porters Pass, measured soil moisture significantly increased from the beginning of the main sampling period (late January 2020) until approximately May 2020 (Figure 3.2). Increase, of a lesser magnitude, was observed to occur contemporaneously at Lake Tennyson, and a small increase was noted at Mount Hutt at approximately the same time. The data logger at Porters Pass malfunctioned three days after installation. This issue was discovered on 23 January 2020 during a routine download, at which point a replacement logger was installed. Consequently, only three days of data were available before this time point (Supplementary Figure 2). Unfortunately, the data logger at Lake Tennyson failed after data were downloaded on 14 December 2020, so no data from this site were subsequently available.

The moisture profile at the Mount Hutt *R. monroi* sampling site was more similar to the Mount Hutt *R. crithmifolius* site than it was to the Porters Pass site. Data logger and moisture probe installation was delayed for the Mount Hutt *R. monroi* site, so the monitoring period

encompassed 12 November 2020 to 8 March 2021. Through this period, mean daily soil moisture was higher at the Porters Pass site than it was at either of the Mount Hutt sites (Figure 3.3; /supplementary/MHC.csv; /supplementary/MHM.csv; /supplementary/PP1.csv). Even though sampling sites at Mount Hutt were within 10 m of one another, a two-tailed Wilcoxon signed rank sum test indicated that the mean daily soil moisture level was significantly higher where the *R. monroi* were growing compared to where the *R. crithmifolius* were growing (V = 6,902, $p < 0.001$). Figure 3.3 visually represents the 1x *SD* for mean daily soil moisture and displays the obvious difference in soil moisture levels between the two Mount Hutt sites and Porters Pass.

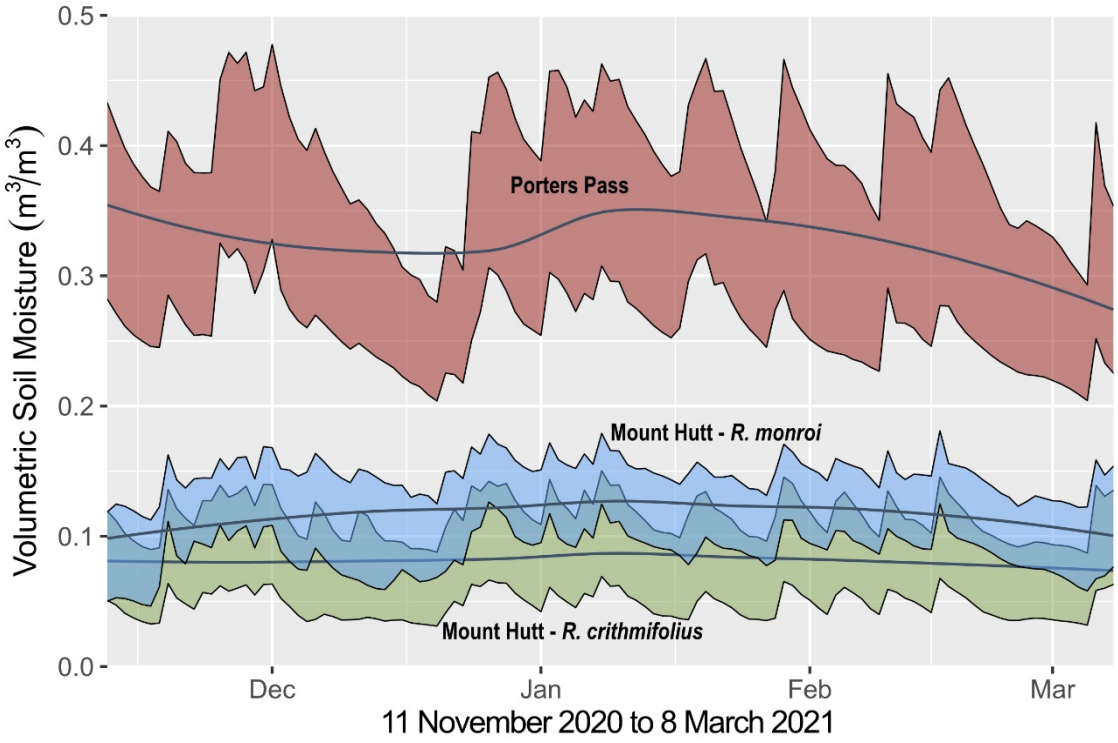


Figure 3.3 Soil moisture Mount Hutt and Porters Pass sample sites. One standard deviation either side of mean daily soil moisture, measured using four electrical conductivity probes, is indicated by coloured bands. Porters Pass, Mount Hutt *R. monroi*, and Mount Hutt *R. crithmifolius* sites are shown in red, blue, and pale green respectively. Dark grey lines represent mean daily soil moisture smoothed using locally weighted scatterplot smoothing (LOESS).

3.1.3 Soil physical characteristics

Replicate soil samples at each site were highly variable but there was a comprehensive distinction between particle size distributions at Porters Pass when compared with the other sites (Supplementary Table 2; /supplementary/soil_raw.csv). Without measuring an upper size limit,

the largest fractions (> 16.00 mm) were composed of individual rocks from 3.98 g (*R. crithmifolius* site, 5–10 cm depth) to 97.84 g (Lake Tennyson site, 0–5 cm depth). The potential for a single rock to comprise a substantial proportion of sample weight meant these > 16.00 mm fractions were excluded from analyses. Plots A and B (Figure 3.4), showing mean proportions of soil particle size composition for the sampling sites at depths of 0–5 cm and 5–10 cm respectively, indicate that samples from Porters Pass appear primarily composed of particles smaller than 0.25 mm (M = 0.86, 95% CI [0.80, 0.93] for 0–5 cm and M = 0.72, 95% CI [0.50, 0.95] at 5–10 cm depths). In comparison, the mean proportion of soil particles smaller than 0.25 mm from the Mount Hutt *R. monroi* site for the 0–5 cm soil depth was 0.20 (95% CI [0.15, 0.26]), while the mean proportion for the 5–10 cm depth was 0.15 (95% CI [0.08, 0.22]). Proportions for the soil samples taken from the *R. crithmifolius* site were similar to those from the Mount Hutt *R. monroi* site, with M = 0.21 (95% CI [0.15, 0.27]) for 0–5 cm soil depth and M = 0.16 (95% CI [0.08, 0.23]) for 5–10 cm. It is noteworthy that of the < 0.25 mm fractions, 95% CIs overlapped for all sites and depths excepting the Porters Pass fractions and the *R. lobulatus* 5–10 cm fraction (Supplementary Table 2). The lower bound for Porters Pass 5–10 cm at 0.46 was larger than the highest upper bound for all other sites at a proportion of 0.25 (*R. crithmifolius* 0–5 cm). Figure 3.4 also gives an indication of soil gravel content at each of the sites. Indeed, for all sites and depths, excepting Porters Pass (both depths) and Mount Hutt *R. monroi* 0–5 cm, more than half the soil sampled appears to be composed of particles larger than 2.00 mm.

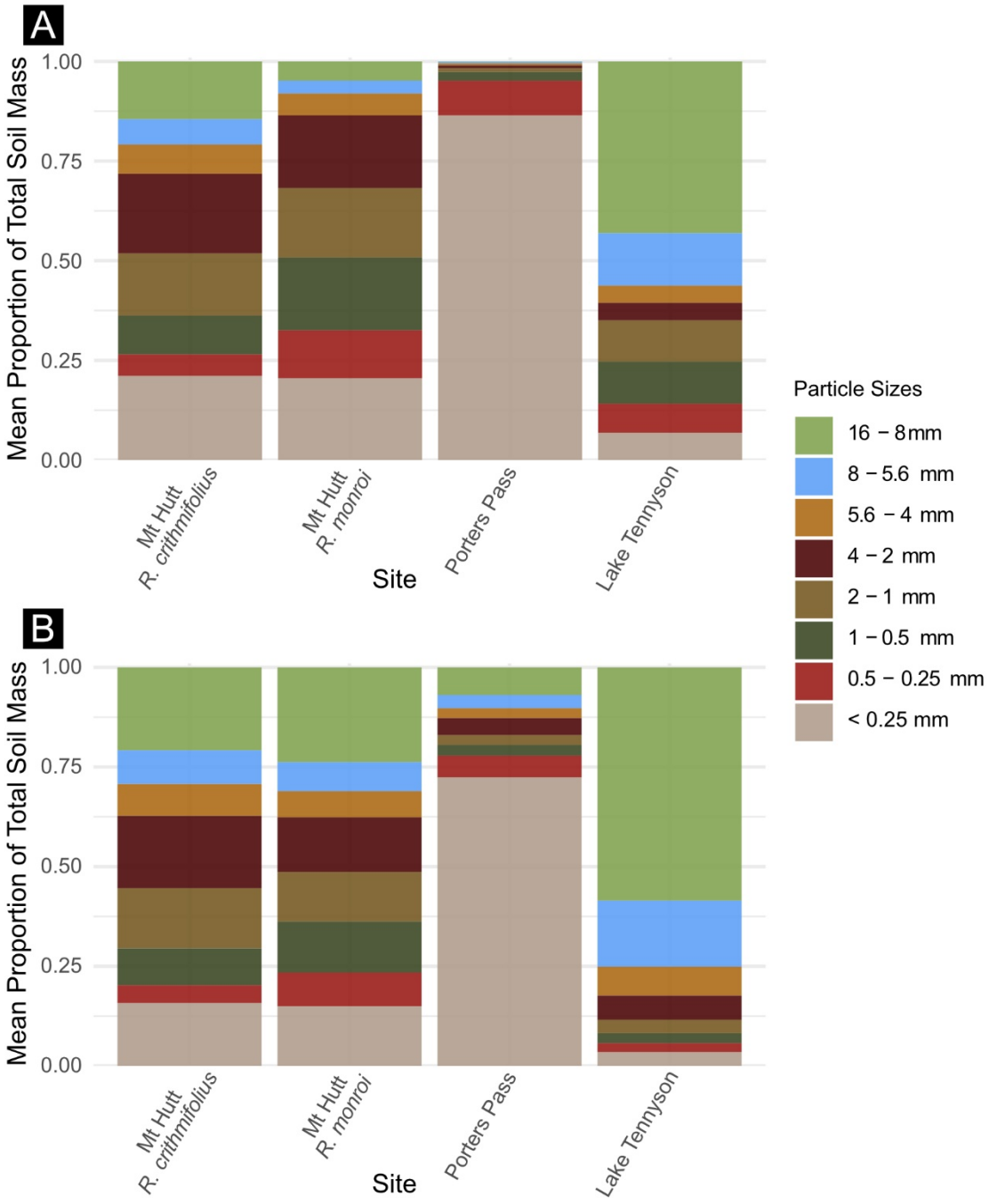


Figure 3.4 Particle size distributions of soil from all four sample sites. A, particle sizes extracted from 0–5 cm soil depth. B, particle sizes extracted from 5–10 cm soil depth. Size fractions are shown as proportions of total sample mass. Gravel larger than 16.00 mm has been excluded from the analyses. Error bars have been omitted for clarity (see Supplementary Table 2).

3.1.4 Soil chemical composition

The two *R. monroi* sites demonstrated conspicuous differences in soil chemistry. According to initial PCA analysis (Supplementary Figure 3; Supplementary Table 3), most of the variation between sites (81.5%) was explained by the first two principal compoments. Notably, the samples from each site (particularly *R. crithmifolius* and Mount Hutt *R. monroi* sites) generally formed discrete clusters, and these clusters grouped closely together relative to the Porters Pass samples. The two *R. lobulatus* site samples potentially formed a loose cluster intermediate between the Porters Pass and Mount Hutt *R. monroi* samples. However, this was difficult to confirm with only two data points. Overall, the spread of data was greatest for the Porters Pass samples, and one Porters Pass data point was significantly outlying.

A further PCA was carried out to test the effect of removing the *R. lobulatus* site samples and the outlying Porters Pass point (Figure 3.5; Supplementary Table 4). This increased the percentage of variance explained by PC1 from 50.6% to 64.8%. The percentage of variance explained by PC2 decreased by a similar magnitude from 31.6% to 18.6%. This indicated the effect of the three removed data points in skewing the loading scores for each PC. In Figure 3.5, the contribution each variable makes to segregation of sampling sites along the PC1 axis is evident. Porters Pass strongly correlated with high levels of organic matter, Na, P, and overall CEC. In contrast, the *R. crithmifolius* and Mount Hutt *R. monroi* sites were most strongly correlated with soil alkalinity and high P levels. Interestingly, the Mount Hutt *R. monroi site* appeared intermediate between the *R. crithmifolius* site and Porters Pass but remained more closely grouped with the former.

Examination of the individual components showed that the soil was acidic at all sites. Total P was relatively low at Porters Pass. Olsen P was higher at the Mount Hutt *R. crithmifolius* and Lake Tennyson sites. Mean levels of Ca were highest at the *R. monroi* sites, but standard error of the mean for Porters Pass and Lake Tennyson were large for these measurements. Magnesium concentration appeared similar at Porters Pass and Lake Tennyson but again, standard error was relatively large. Total N, C:N, K, Na, CEC, and OM were lowest at the Mount Hutt sites, while TBS was highest for these. The mean values of soil chemistry measurements at each site have been provided in Supplementary Table 5.

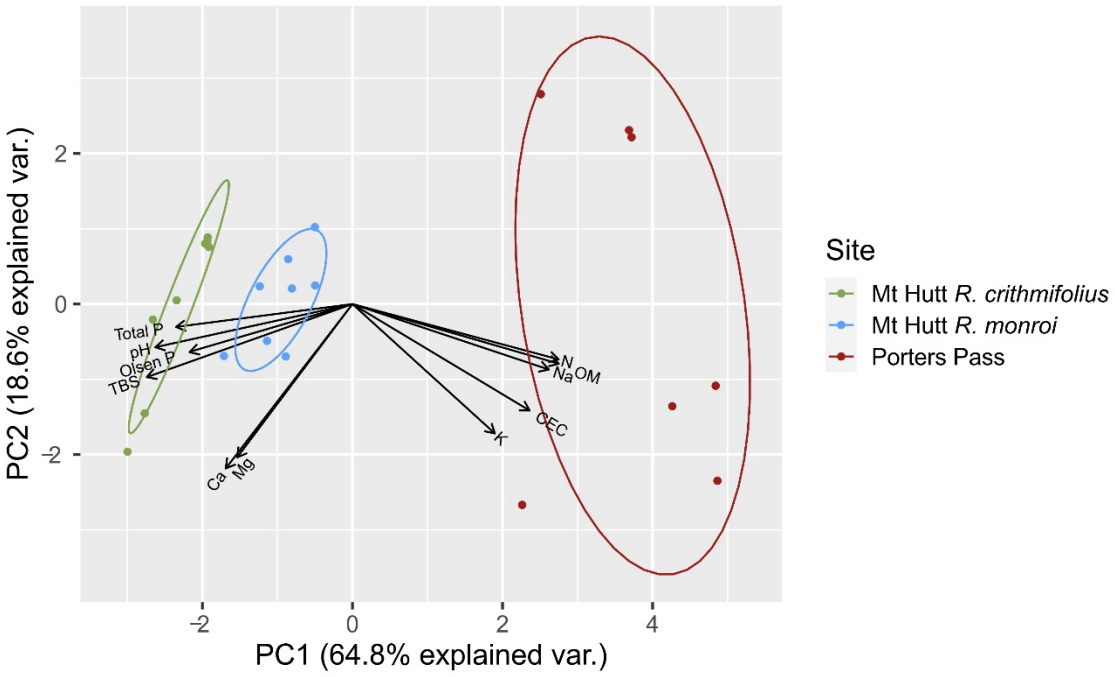


Figure 3.5 Principal coordinate analysis of sample site soil chemistry. Dots represent soil samples. Vector directions and lengths indicate variable loadings for each axis. Each ellipse represents the 68% confidence interval of the group core.

3.2 Experimental plants

3.2.1 DNA barcoding

Sequencing and alignment of the ITS2 PCR products produced a MSA that was 602 nt in length with 8 SNPs and 24 ambiguous sites. Sequencing and alignment of the partial *ycf1* PCR products produced a MSA that was 478 nt long with a total of 10 SNPs.

The ITS2 sequences for these samples contained significant numbers of ambiguous nucleotide sites (/supplementary/ITS2.fasta; Supplementary Figure 4). Manual examination of chromatograms in Geneious (example as Supplementary Figure 5) indicated the site patterns were polymorphic characters and not artifacts arising from sequencing error. Site pattern incompatibilities in these ITS2 sequences supported alternative relationships among the *R. lobulatus* and the two *R. monroi* forms when represented as a phylogenetic network (Figure 3.6). However, regardless of incompatibilities, the *R. crithmifolius* samples formed a distinct cluster. The edge lengths separating the *R. crithmifolius* cluster from the *R. insignis* forms, indicated these taxa were relatively phylogenetically distant. Interestingly, when all sites with ambiguous nucleotides were removed from the MSA, sequences for all *R. crithmifolius* samples

collapsed to a single node, as did sequences for all *R. insignis* form samples. This produced a graph with two nodes connected by a single edge. In other words, after filtering, there were no sequence differences among the *R. insignis* forms.

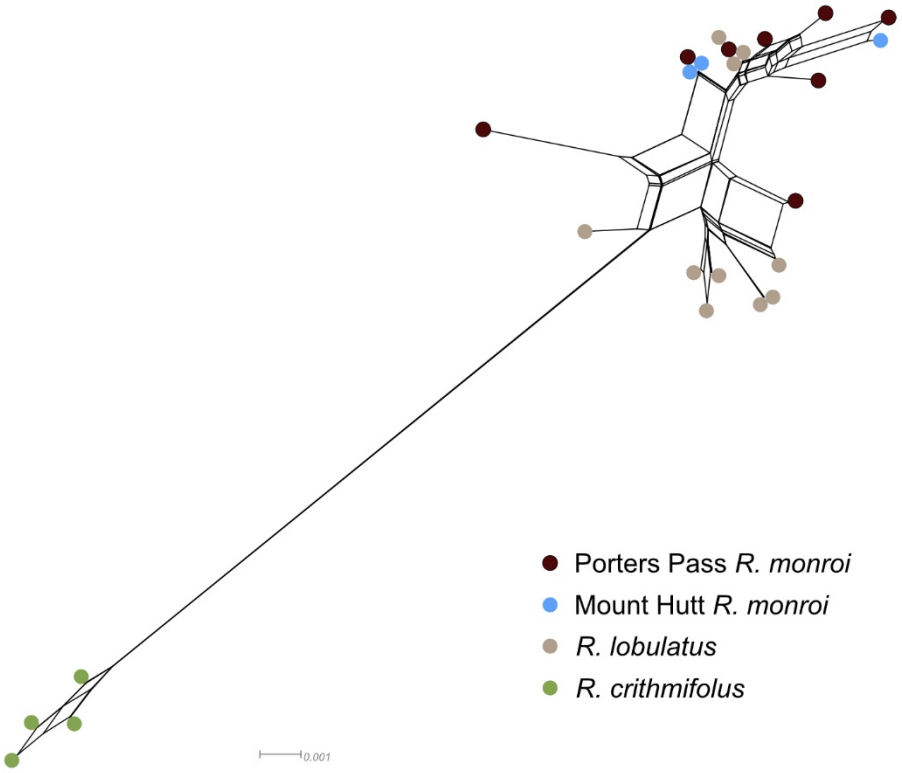


Figure 3.6 Neighbor-Net of internal transcribed spacer 2 (ITS2) DNA sequences. Coloured dots represent plants sampled. *Ranunculus crithmifolius* samples cluster together strongly. The remaining three taxa (Mount Hutt *R. monroi*, Porters Pass *R. monroi*, and *R. lobulatus*) form a loose cluster with no taxon-specific groups.

The *ycf1* MSA (/supplementary/ycf1.fasta) contained few parsimony-informative sites. Accordingly, each sample cluster differed by at most seven SNPs from any other cluster (Figure 3.7). Three distinct haplotypes were detected for *R. monroi* at Porters Pass but only a single haplotype for each other taxon. The phylogenetic network (Figure 3.7) showed each of the taxa joined distinct internal nodes but the pattern of sequence variation did not resolve the relationships at the centre of the network.

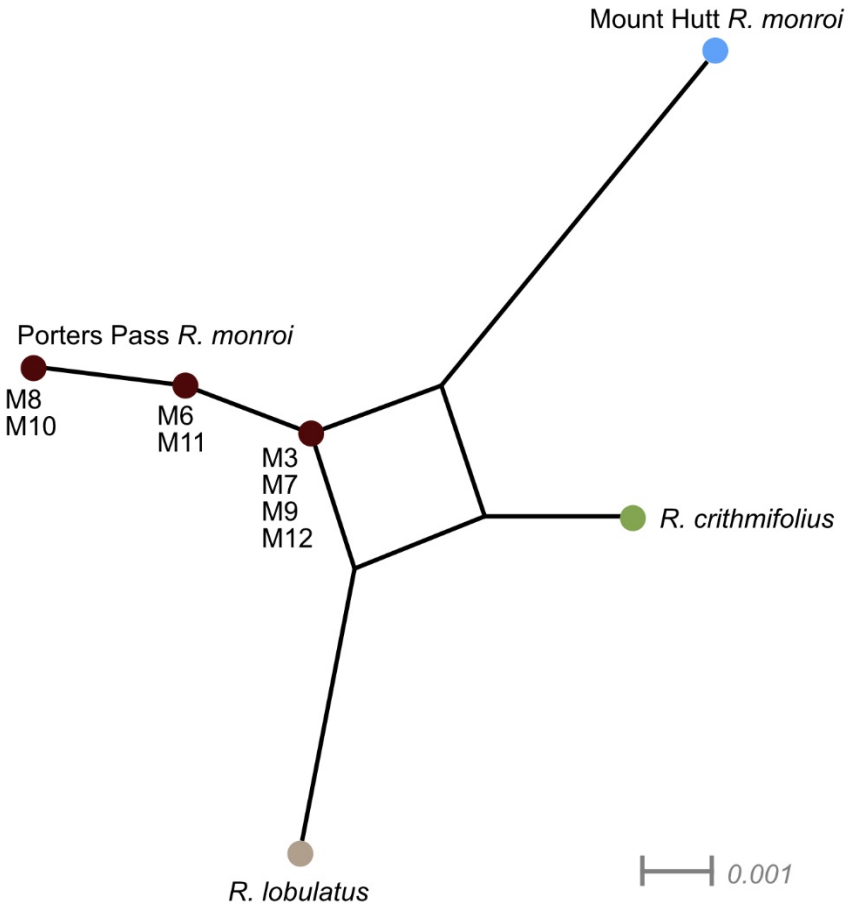


Figure 3.7 Neighbor-Net of partial *ycf1* cpDNA sequences. Coloured dots represent plants sampled. Mount Hutt *R. monroi*, *R. crithmifolius*, and *R. lobulatus* sequences are identical for each taxa so collapse to single nodes. Three haplotypes are present for Porters Pass *R. monroi*. Plant IDs belonging to each of these haplotypes are noted under the nodes. How taxa relate to one another cannot be resolved.

3.2.2 Plant Growth

The standardised growth chamber conditions, and watering regimen used for plant growth were generally successful in maintaining healthy plants of all four taxa.

Plant weight varied considerably at the time of potting (Supplementary Table 6). The two lightest plants (both Porters Pass *R. monroi*) weighed 5 g each, and the heaviest plant (*R. crithmifolius*) weighed 27 g. A Kruskal-Wallis test detected no significant differences ($\chi^2 = 3.752$, $df = 3$, $p = 0.290$) in weight between the four taxa (*R. monroi* from Mount Hutt and Porters Pass, *R. crithmifolius, R. lobulatus*).

Of the 30 plants initially sampled from the environment and grown, five did not survive to harvest. Four *R. crithmifolius* plants died due to unknown causes and one *R. lobulatus* was removed from the experiment due to severe aphid infestation. Growth of the remaining plants

was positive. Although plant weights were not recorded at the time of harvest due to time constraints, photographs illustrate the health of plants at the time of potting (Figure 3.8, A & C; Supplementary Figure 6) and harvest potting (Figure 3.8, B & D; Supplementary Figure 7).

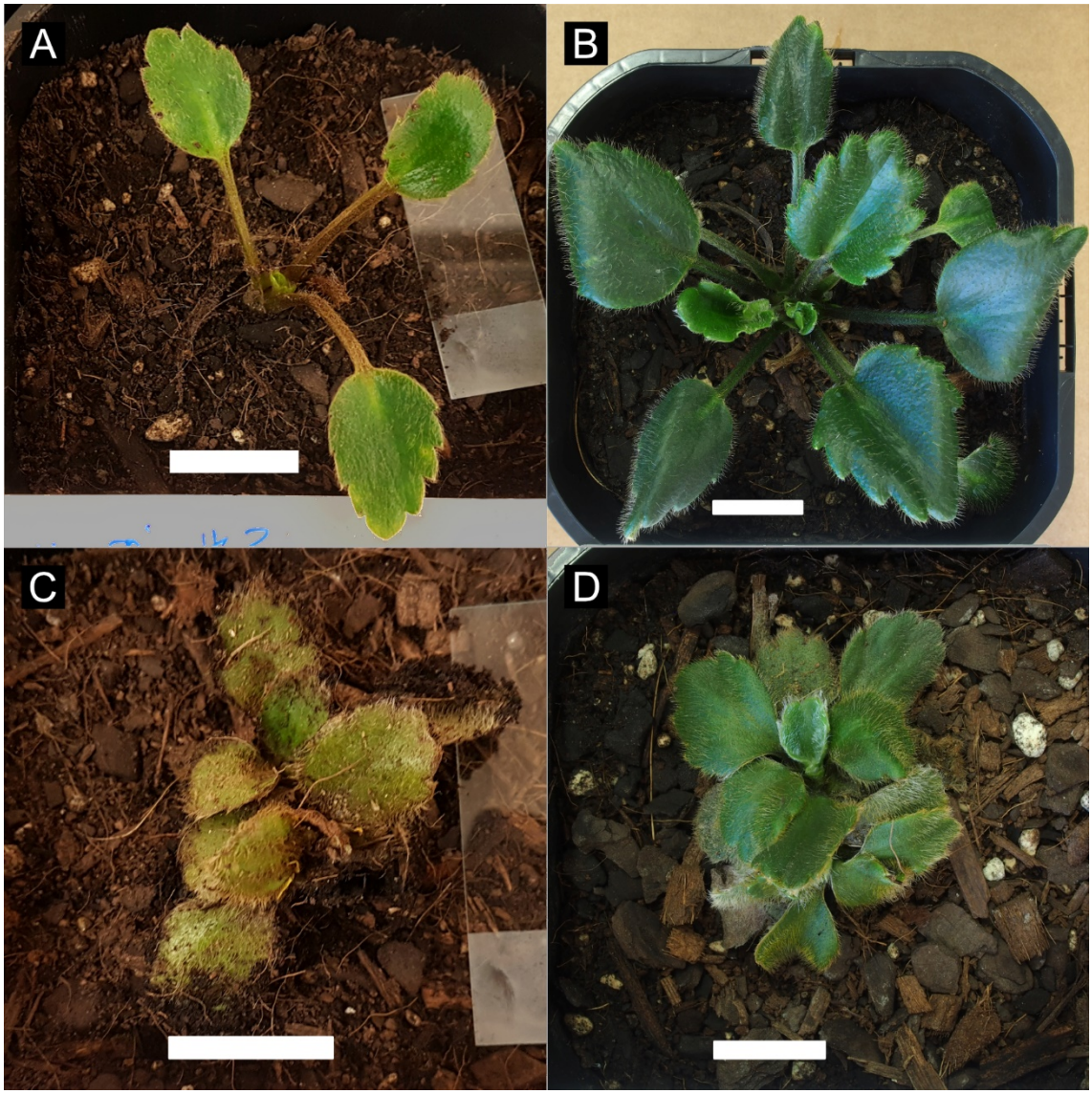


Figure 3.8 Plant health when potted and at harvest. A, Porters Pass *R. monroi* (ID: M3) after being potted (29 November 2019) and B, two days prior to harvest (25 February 2020). C and D, Mount Hutt *R. monroi* (ID: M2) at the same time points. Scale bars = 21 mm.

Mean water usage by the plants increased over the course of the experiment (Supplementary Figure 8; /supplementary/weights_total.csv). Analysis using Kendall's rank correlation supported a significant association between mean water usage and increasing duration of the experiment (tau-b = 0.739, $p < 0.001$). Interestingly, a correlation (tau-b = 0.593, $p < 0.001$) was also detected between cabinet position and water loss (Supplementary Figure 9).

Physiologically, the plants harvested did not appear to be undergoing any unusual stresses. However, taxon-specific responses were noted. *Ranunculus lobulatus* and Porters Pass *R. monroi* showed higher dark-adapted photosynthetic efficiency (Fv/Fm) readings than Mount Hutt *R. monroi* and *R. crithmifolius* (Figure 3.9; /supplementary/Fv-Fm.csv). The highest Fv/Fm values were recorded for *R. lobulatus* and Porters Pass *R. monroi* with means of 0.833 (SE = 0.002) and 0.832 (SE = 0.003) respectively. Mount Hutt *R. monroi* had the lowest mean Fv/Fm (M = 0.796, SE = 0.004) which was not dissimilar to *R. crithmifolius* (M = 0.803, SE = 0.005). To test for differences between taxa in the unwatered state, a Kruskal-Wallis test was carried out and a significant difference was detected ($\chi^2 = 40.05$, $df = 3$, $p < 0.001$). Following this, a Dunn test detected no significant difference between mean Fv/Fm values of Porters Pass *R. monroi* and *R. lobulatus* (p = 0.401), nor between Mount Hutt *R. monroi* and *R. crithmifolius* (p = 0.160). All other comparisons were found to be significant ($p < 0.001$).

Values for Fv/Fm were observed to be higher prior to watering (Figure 3.9). This finding supported harvesting plants on the fourth day after watering. However, a two-tailed Wilcoxon signed rank test showed the apparent reductions in Fv/Fm after watering were only significant in *R. lobulatus* and Porters Pass *R. monroi* (both $p < 0.001$). They were not significantly different for Mount Hutt *R. monroi* (p = 0.850) or for *R. crithmifolius* (p = 0.081).

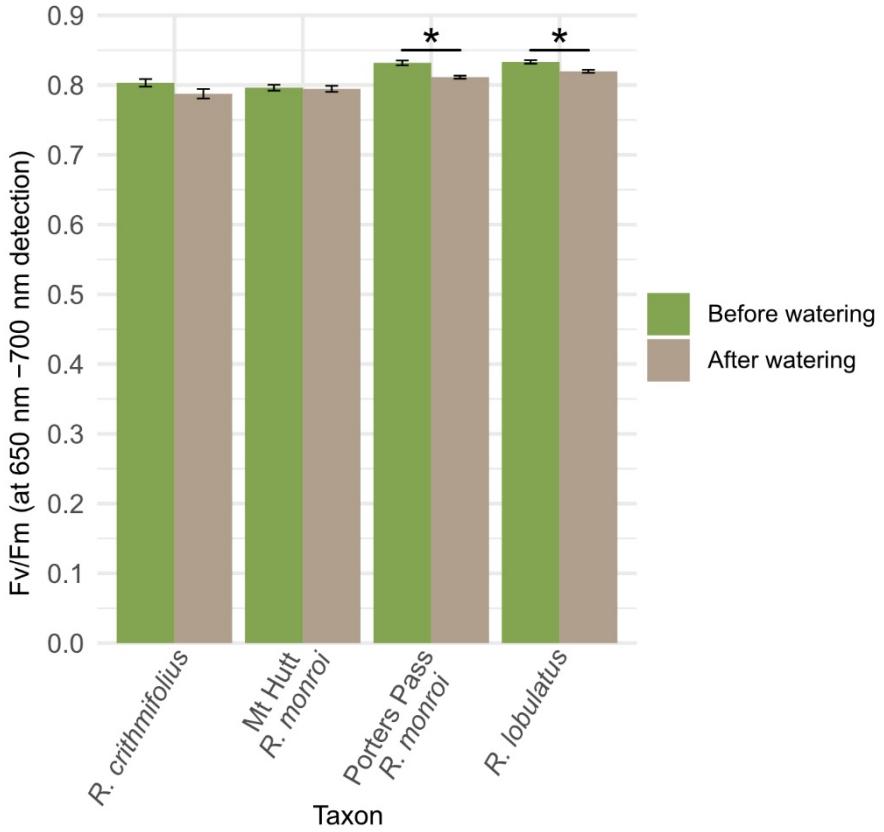


Figure 3.9 Plant stress measured using PAM. Dark-adapted photosynthetic efficiency (Fv/Fm) of each plant was measured immediately prior to watering (five days after last watering) and subsequently 2.5 hours after watering using pulse amplitude modulation fluorometry (PAM). For all taxa, a reduction in Fv/Fm after watering occurred. Significant ($p < 0.001$) reductions in mean Fv/FM after watering are indicated by a star. Significant differences between taxa are not shown. Error bars indicate standard error (*SE*).

The plants were provided similar water availability at harvest. Measurement of soil moisture at the time of harvest showed little difference between taxa (Figure 3.10; Supplementary Table 7). Mean soil moisture ranged from 0.229 m^3/m^3 (*SD* = 0.006 m^3/m^3) for Mount Hutt *R. monroi* to 0.252 m^3/m^3 (*SD* = 0.027 m^3/m^3) for *R. lobulatus.* As suggested by Figure 3.10, the differences in final soil moisture among taxa were not found to be significant (Kruskal-Wallis $\chi^2 = 3.8762$, $df = 3$, $p = 0.2752$).

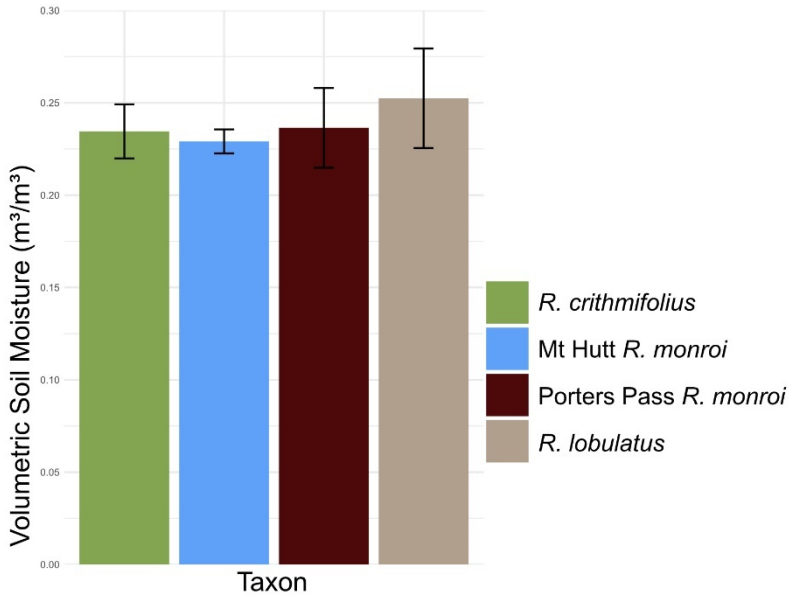


Figure 3.10 Soil moisture in pots after plant harvest. Volumetric water content of potting mix at the time of plant harvest was calculated. Error bars indicate 1x standard deviation (*SD*) either side of the mean.

3.3 RNA

3.3.1 Extraction of RNA, sequencing, and quality control

A method of extraction that yielded quality RNA for high-throughput sequencing was developed. This was necessary because techniques and methods initially used for *R. lobulatus* and *R. monroi* failed to successfully extract RNA from *R. crithmifolius.* It was the addition of PSS Solution (Qiagen) to the RNA extraction buffer that allowed for a single method of extraction to be used for all samples (Supplementary Table 8). With each tissue sample, only eluates that yielded RNA with a concentration above 30 ng/µL, and NanoDrop 260/280 and 260/230 values above 1.80 and 1.75 respectively, were retained. All samples returned LabChip RNA Quality Scores (RQS) above 5.5.

The Illumina RNA sequence data returned from Novogene was high quality. This comprised 24 (12 forward and 12 reverse) gzipped FASTQ files totaling approximately 137 GB of compressed data. Read numbers ranged from 66.9 million to 97.7 million paired-end reads per sample (Supplementary Table 9). A small amount of Illumina adapter contamination was also noted in most samples (Supplementary Figure 10). After removing reads deemed 'unfixable' by Rcorrector, and trimming with quality filtering using TrimGalore!, at least 96% of reads were retained across all samples (Supplementary Table 9).

3.3.2 Transcriptome assemblies and representative contigs—protein prediction and annotation

Aggregated *de novo* transcriptomes representing the four taxa under investigation were generated. Each aggregated assembly was constructed using RNA-sequencing reads available from the three samples per taxon. Subsequently, *de novo* assemblies were also constructed separately for each sample. Assembly statistics for all assemblies are provided in Supplementary Table 10.

The Trinity software outputs assembled contigs, or transcripts, as 'isoforms' connected to clusters of 'genes'. The greatest number of contigs assembled and clusters identified occurred for combined biological replicates in Mount Hutt *R. monroi* (359,976 contigs and 170,719 clusters). Conversely, assembly of Porters Pass *R. monroi* RNA-seq data produced the smallest number of contigs and clusters of the four taxa (310,862 contigs belonging to 131,148 clusters). Interestingly, assembly of RNA-seq data for *R. crithmifolius* and Mount Hutt *R. monroi* produced greater numbers of clusters than did assembly of RNA-seq data for *R. lobulatus* and Porters Pass *R. monroi* (162,975 and 170,719 compared to 140,165 and 131,148). The Mt Hutt-*R. monroi* transcriptome had four contigs (of the same gene cluster) longer than 22,000 nt, with the longest contig being 22,812 nt. This was noticeably longer than any contig in the other assemblies. Furthermore, of the individual assemblies constructed with single-sample data, the *R. crithmifolius*_2 and Mt Hutt-*R. monroi*_3 assemblies exhibited larger cluster sizes than the other *R. crithmifolius* and Mount Hutt *R. monroi* assemblies. Additionally, the longest contig in the Mt Hutt-*R. monroi*_1 assembly was 22,238 nt long, which corresponded well with the longest contig in the Mt Hutt-*R. monroi* aggregated assembly. Detailed assembly data are provided in Supplementary Table 10.

Despite Mt Hutt-*R. monroi* having the longest contigs, Mt Hutt-*R. monroi* and *R. crithmifolius* had lower N50 and median contig length values compared to the other taxa (Supplementary Table 10). Mt Hutt-*R. monroi* had the lowest median contig length (677 nt), lowest average contig length (994 nt), and lowest N50 value (1,377 nt) Conversely, the highest figures for these same metrics (833 nt, 1,124 nt, and 1,546 nt respectively) were found in *R. lobulatus*_3. GC content ranged from 40.62% (*R. lobulatus*) to 41.93% (Porters Pass-*R. monroi*_1 and *R. lobulatus*_3). Assembled bases differed between the aggregated and individual assemblies (Supplementary Table 10). The aggregated assemblies incorporated a greater number of bases (M = 3.46 x 10^8, SD = 1.44 x 10^7 nt) than the individual assemblies (M = 2.11 x 10^8, SD = 1.77 x 10^7 nt).

The BUSCO results indicated the aggregated transcriptomes were nearly complete representations of genes for each taxon. Between 91.4% and 92.5% complete single-copy genes from the ODB embryophyta dataset were present in each of the aggregated assemblies, and

there were high rates (77%–82.8%) of duplicate orthologues detected (Table 3.2). A small number of BUSCOs were recovered in a fragmented state, and across the aggregated transcriptomes only 2.7% to 3.6% of BUSCOs were found to be missing (Table 3.2).

Table 3.2 BUSCO recovery percentages

Assembly	Total Complete	Single Copy	Duplicated	Fragmented	Missing
R. crithmifolius	91.4	14.4	77.0	5.1	3.5
Mt Hutt-*R. monroi*	92.1	13.9	78.2	4.3	3.6
Porters Pass-*R. monroi*	92.5	13.4	79.1	4.8	2.7
R. lobulatus	92.1	9.3	82.8	5.0	2.9

Quality assessment of the aggregated transcriptomes indicated the assemblies utilised much of the input read data. Mapping of sequencing reads against the assembled contigs, using Bowtie2 (Table 3.3), resulted in overall alignment rates between 97.53% and 98.08% across all four taxa. Most reads (93.85%–94.98%) also aligned concordantly, such that each read from a pair aligned in the expected orientation and within the expected distance relative to one another. Notably, there was much multi-mapping of reads due to the presence of multiple isoforms in each cluster. Between 81.79% and 84.84% of reads mapped to more than one locus in this case.

Table 3.3 Bowtie2 read alignment percentages

Assembly	Initial Assemblies				Representative Assemblies			
	Total[a]	C = 1[b]	C > 1[c]	C Total[d]	Total	C = 1	C > 1	C Total
R. crithmifolius	97.53	12.06	81.79	93.85	70.68	61.17	1.56	62.73
Mt Hutt-*R. monroi*	97.82	10.61	83.97	94.58	73.59	64.02	2.40	66.42
Porters Pass-*R. monroi*	97.86	11.18	82.89	94.07	72.97	62.54	2.37	64.91
R. lobulatus	98.08	10.14	84.84	94.98	73.69	64.11	2.82	66.93

[a] Overall reads aligning

[b] Reads aligning concordantly to a single locus

[c] Reads aligning concordantly to more than one locus

[d] Total concordantly aligning reads

However, aligning reads to the most highly expressed contig for each gene cluster drastically changed the alignment statistics. While the alignments for reads mapping concordantly to the representative assemblies ranged between 62.73% and 66.93%, the multi-mapping that was observed in the initial assemblies was drastically reduced (Table 3.3). Of reads mapped to the representative transcriptomes, the highest percentage of multi-mapping observed was only 2.82% concordant alignment to more than one locus of the *R. lobulatus* assembly (Table 3.3).

The violin plots shown in Figure 3.11 for the aggregated assemblies and Supplementary Figure 11 for the individual assembles (with Supplementary Tables 10 and 11 giving detailed statistics) highlight the similarities and differences between initial assemblies and the representative assemblies. Notably, summary statistics for all initial assemblies are similar, and there is a general shift to smaller values for mean, median, and interquartile range of contig lengths when the representative assemblies are examined. But again, all four representative assembly statistics appear comparable to one another. Generally, the longest transcript for an initial assembly and its subsequent representative transcriptome were similar length, But, interestingly, the longest contig of the Mt Hutt-*R. monroi* representative transcriptome was only 14,620 nt, a reduction in length of over 8,000 nt compared with the longest transcript in the Mt Hutt-*R. monroi*_1 initial assembly.

Typically, TransDecoder outputs reflected the number of clusters in the assemblies. In other words, the greater the number of gene clusters, the greater the number of predicted proteins (Supplementary Table 11). Also, more predicted proteins were found in the aggregated assembly for a given taxon than in any assembly made from its individual samples. More proteins (51,866 and 50,082) were predicted in the *R. crithmifolius* and Mt Hutt-*R. monroi* transcriptomes than in the *R. lobulatus* or Porters Pass-*R. monroi* transcriptomes which had 29,749 and 28,581 proteins respectively (Supplementary Table 11). The *R. crithmifolius*_2 and Mt Hutt-*R. monroi*_3 individual assemblies had more proteins than any of the other individual transcriptomes for their respective taxa. The fewest predicted proteins (23,243) were discovered in Porters Pass-*R. monroi*_1 (Supplementary Table 11).

For the *R. lobulatus* aggregated assembly, 18,816 predicted proteins had BLASTp hits against the Araport11 gene dataset (/supplementary/lobulatus_arabidopsis.outfmt6). This means that 63% of the predicted proteins in the *R. lobulatus* transcriptome had successful hits against annotated *A. thaliana* proteins (/supplementary/lobulatus_arabidopsis.txt). The predicted proteins for *R. lobulatus*_1 successfully matched against the Araport11 gene dataset 17,893 (71%) times (/supplementary/lobulatus_1_arabidopsis.outfmt6; /supplementary/ lobulatus_1_arabidopsis.txt).

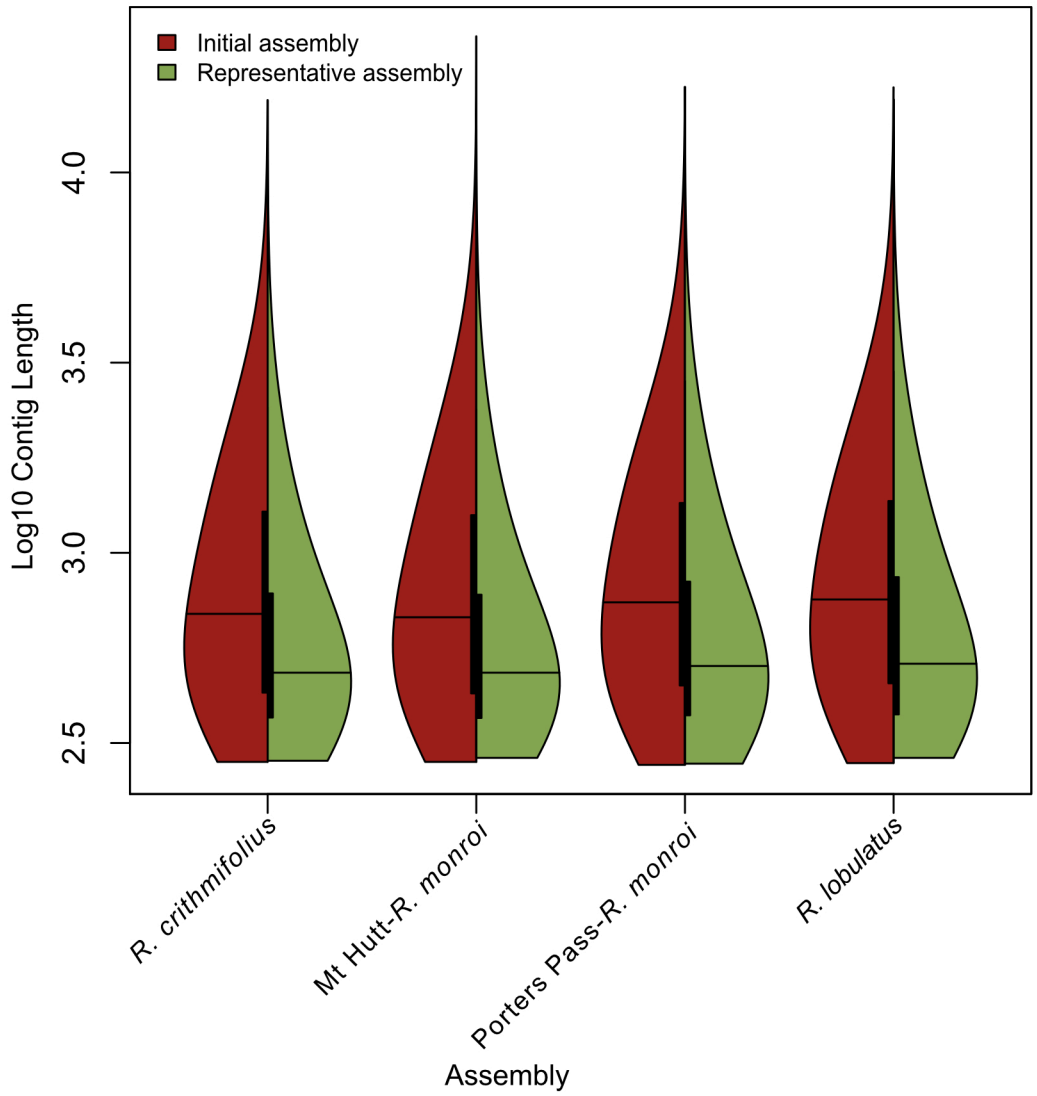


Figure 3.11 Aggregated transcriptome assembly statistics. The violin plots show the distribution of log10-normalised contig lengths for initial and representative aggregated assemblies. Median contig length is represented by horizontal bars, and interquartile range by thick vertical bands. Widths of the plots indicate the proportion of contigs at that length.

While investigating the large amount of predicted proteins in some assemblies it was found that a greater number of unique BLAST hits occurred for the samples with more TransDecoder predicted proteins. The TransDecoder peptide sequences for *R. crithmifolius*_2 and Mt Hutt-*R. monroi*_3 had increased numbers of unique BLASTp hits against the *A. thaliana* protein sequences when compared with other samples of the same taxa (Table 3.4).

Table 3.4 Predicted protein BLASTp matches against unique *A. thaliana* proteins		
Representative Assembly	*A. thaliana* Gene Matches	*A. thaliana* Isoform Matches
R. crithmifolius _1	12051	12911
R. crithmifolius _2	12879	14195
R. crithmifolius _3	11539	12223
Mt Hutt-*R. monroi* _1	11817	12428
Mt Hutt-*R. monroi* _2	11594	12264
Mt Hutt-*R. monroi* _3	12743	13950

3.3.3 Chloroplastic sequences

The longest contig of the Mt Hutt-*R. monroi* transcriptome (TRINITY_DN12941_c0_g1_i8) was derived from cpDNA. The top hit returned from searching the non-redundant nucleotide sequences database was the complete genome of *R. sceleratus* GenBank (Sayers *et al.* 2019b) accession MK253452.1 with 100% coverage of the Mt Hutt-*R. monroi* contig, E-value of zero, 96% identities, and 1% gaps. Additionally, the TransDecoder predicted protein (397 amino acids) from this gene cluster returned a (1,853 amino acids long) hypothetical chloroplast RF1 (*ycf1* gene) protein from *R. sceleratus* (GenBank: QFV17859.1) as the top hit after searching the non-redundant protein sequences database. Alignment of reads to the *R. sceleratus* genome produced much higher alignment rates for Mt Hutt-*R. monroi*_1 and Mt Hutt-*R. monroi*_2 compared with the other samples (Table 3.5).

Table 3.5 Bowtie2 read alignments to the *R. sceleratus* chloroplast	
Sample	Alignment Percentage[a]
R. crithmifolius _1	0.12
R. crithmifolius _2	0.10
R. crithmifolius _3	0.12
Mt Hutt-*R. monroi* _1	2.29
Mt Hutt-*R. monroi* _2	3.46
Mt Hutt-*R. monroi* _3	0.09
Porters Pass-*R. monroi* _1	0.20
Porters Pass-*R. monroi* _2	0.11
Porters Pass-*R. monroi* _3	0.16
R. lobulatus _1	0.52
R. lobulatus _2	0.27
R. lobulatus _3	0.25

[a] Overall alignments

3.4 Gene expression and phylogenetics

3.4.1 Standard pairwise comparisons of gene expression

The choice of reference assembly for mapping had an impact on read alignment and this affected counting of mapped reads. However, regardless of which aggregated assembly was chosen as the reference against which reads were aligned, the expected taxonomic signal remained consistent. Figure 3.12 (and Supplementary Figures 12 and 13) shows that when transcriptome profiles of individual biological replicates were compared, all replicates from a given taxon clustered together and the two *R. monroi* populations (Mount Hutt and Porters Pass) clustered together as would be expected under the currently accepted *Ranunculus* taxonomy (Fisher 1965; Webb *et al.* 1988). The clustering algorithm also produced a weighted tree in which *R. crithmifolius* samples are more distant from the other three taxa. Again, this was expected according to our current knowledge of species relationships. However, the heatmaps (Figure 3.12; Supplementary Figures 12 and 13) also graphically illustrate that when the reference assembly is derived from different taxa being mapped, sequencing reads from the same taxon map with greater efficiency than samples from the other taxa. Across all three assemblies tested, levels of gene expression are much greater for all self-mapped samples. Read mapping statistics reinforce this. The Bowtie2 overall alignment rates output from transcript quantification with RSEM were higher for each sample that was self-mapped to a reference transcriptome (Table 3.6).

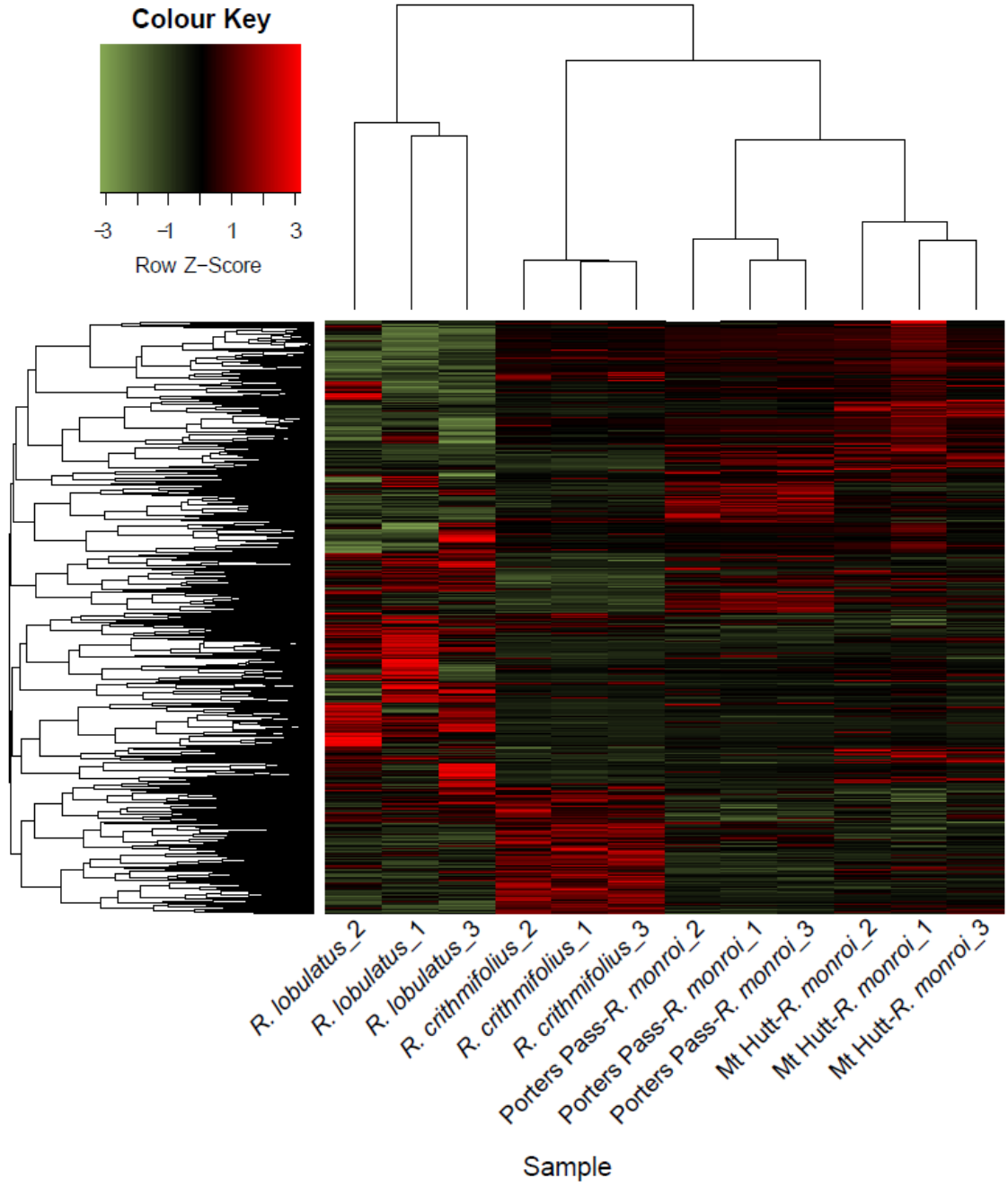


Figure 3.12 Gene expression analysis by alignment of reads against the *R. lobulatus* aggregated assembly. The heatmap shows row-scaled log10-transformed transcripts per million (TPM) normalised values for 16,427 genes in which at least one taxon was significantly differentially expressed. Red represents upregulation of genes relative to the other samples, and green represents downregulation. Each row represents one gene. Columns represent sample RNA-seq reads, and the weighted dendrogram generated using hierarchical clustering with Spearman's rho.

Table 3.6 Bowtie2 alignments to initial aggregated assemblies

Sample	Overall Read Alignment percentage for each assembly		
	R. crithmifolius	*R. lobulatus*	Porters Pass-*R. monroi*
R. crithmifolius _1	90.65	79.91	81.35
R. crithmifolius _2	90.74	77.60	79.44
R. crithmifolius _3	89.65	74.87	76.65
Mt Hutt-*R. monroi* _1	80.52	85.91	84.95
Mt Hutt-*R. monroi* _2	81.64	85.88	84.32
Mt Hutt-*R. monroi* _3	79.98	82.41	82.29
Porters Pass-*R. monroi* _1	81.24	82.10	91.03
Porters Pass-*R. monroi* _2	81.35	82.95	90.74
Porters Pass-*R. monroi* _3	81.65	82.88	91.13
R. lobulatus _1	80.40	90.87	83.86
R. lobulatus _2	80.58	90.27	84.29
R. lobulatus _3	79.03	90.28	83.28

This bias was also evident in Volcano plots which visualise differential expression in pairwise comparisons. Plots A and B of Figure 3.13 show downregulation of many genes when *R. lobulatus* is a member of the group serving as the reference level (or control). However, when *R. lobulatus* is a member of the group being compared against the reference level, the number of upregulated genes increases dramatically (Figure 3.13: Plot C). Additionally, all three plots appear to contain similar clusters of genes with expression levels below or above a log2 fold change of 20. Analyses of TPM values showed that of all genes with log2 fold change over 20, 80 genes were shared across all three comparisons. Again, this appears to be an artifact of using the *R. lobulatus* aggregated assembly to map reads against, as well as an issue with combining taxa into groups arbitrarily.

Investigation of these 80 shared genes with high log2 fold change revealed most of them to be genes with read counts for the *R. lobulatus* samples but no counts for the other taxa. This means that comparisons are not being made across all taxa when reads from only one taxon map against the assembly. For example, in comparing Mount Hutt *R. monroi* and Porters Pass *R. monroi* against *R. lobulatus* and *R. crithmifolius*, Trinity gene TRINITY_DN14978_c0_g1 had a log2 fold change of -27.826 and adjusted *p* value of 4.16E-17. However, examination of TPM values across all samples for this gene showed that *R. lobulatus*_1 = 40.21, *R. lobulatus*_2 = 186.18, and *R. lobulatus*_3 = 19.49. No reads for any other sample successfully aligned against the assembly. Therefore, although TRINITY_DN14978_c0_g1 was determined by DESeq2 to be downregulated, this finding obscures the fact that *R. crithmifolius* shares a null expression level of this gene with Mount Hutt *R. monroi* and Porters Pass *R. monroi*. There is

no phylogenetic signal for this locus which justifies separating the four taxa into groups of two taxa.

As with read alignments to the aggregated assemblies, more closely-related taxa had increased number of reads mapping when aligned to the individual assemblies (Table 3.7). However, both the Mt Hutt-*R. monroi*_1 and Porters Pass-*R. monroi*_1 datasets had noticeably higher percentages when self-mapping compared with mapping of reads from other samples of the same taxa (Table 3.7). Reads from Mt Hutt-*R. monroi*_1 had 91.86% concordant read alignments when mapping to the Mt Hutt-*R. monroi*_1 transcriptome. In contrast, 85.97% and 86.69% read alignments were recorded for the other two Mount Hutt *R. monroi* samples against the Mt Hutt-*R. monroi*_1 assembly. Alignments for the Porters Pass-*R. monroi*_1 sample reads against the Porters Pass-*R. monroi*_1 assembly were 90.73%, compared with 85.48% and 88.54% for Porters Pass-*R. monroi*_2 and Porters Pass-*R. monroi*_3 respectively.

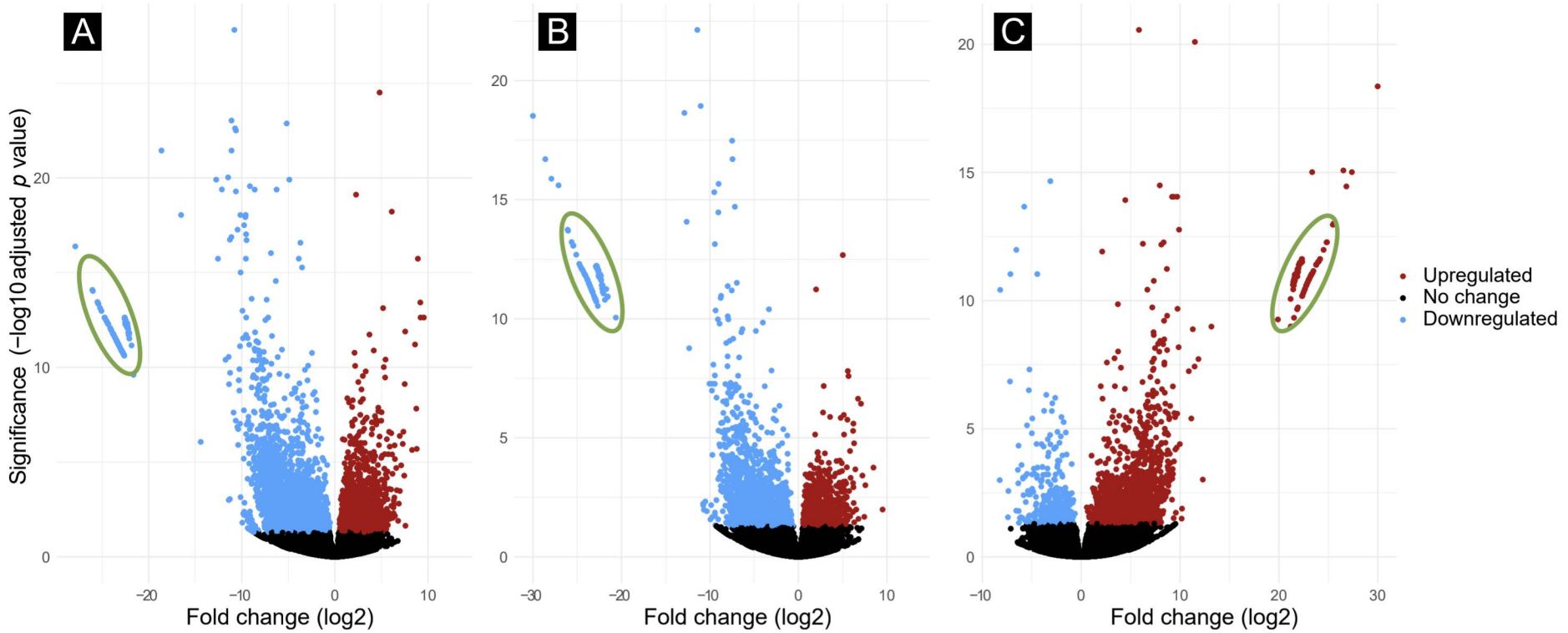


Figure 3.13 Standard pairwise differential gene expression analyses with DESeq2. A, comparison of Mount Hutt *R. monroi*–Porters Pass *R. monroi* against *R. crithmifolius–R. lobulatus* (reference level). B, comparison of Mount Hutt *R. monroi–R. crithmifolius* against Porters Pass *R. monroi–R. lobulatus* (reference level). C, comparison of Mount Hutt *R. monroi–R. lobulatus* against Porters Pass *R. monroi–R. crithmifolius* (reference level). Green ellipses indicate genes, with similar expression patterns in each analysis, that appear downregulated when *R. lobulatus* is included in the reference level and upregulated when *R. lobulatus* is not part of the reference level. Overall numbers of upregulated and downregulated genes also follow this pattern.

Table 3.7 Bowtie2 alignments to individual assemblies		
Sample Reads	Overall Read Alignment Percentage for Each Assembly	
	Mt Hutt-*R. monroi* _1	Porters Pass-*R. monroi* _1
R. crithmifolius _1	78.41	77.67
R. crithmifolius _2	76.16	74.64
R. crithmifolius _3	73.37	72.26
Mt Hutt-*R. monroi* _1	91.86	83.88
Mt Hutt-*R. monroi* _2	85.97	83.72
Mt Hutt-*R. monroi* _3	86.69	79.63
Porters Pass-*R. monroi* _1	83.03	90.73
Porters Pass-*R. monroi* _2	84.13	85.48
Porters Pass-*R. monroi* _3	82.2	88.54
R. lobulatus _1	80.39	80.16
R. lobulatus _2	81.55	80.98
R. lobulatus _3	81.47	80.88

Regardless of the issues noted when attempting to explore gene expression using a traditional pairwise approach, it is evident that signals of phylogenetic relationship dominate the similarity and differences between the transcription profiles of different plant populations. For example, more than double the DEGs were detected when Mount Hutt *R. monroi* and Porters Pass *R. monroi* were treated as part of the same group compared with groupings of Mount Hutt *R. monroi* with *R. crithmifolius* or Mount Hutt *R. monroi* with *R. lobulatus* (Supplementary Table 12). Grouping Mount Hutt *R. monroi* with *R. crithmifolius* resulted in a 28% increase in DEGs compared to grouping Mount Hutt *R. monroi* with *R. lobulatus*.

To overcome the more obvious effects of mapping biases, the data were re-analysed using a more conservative approach. In these comparisons, DESeq2 was run only on genes that had reads mapping from all 12 samples. In performing this filtering, the number of genes analysed for differential expression was reduced to 35,692. The extreme outlying values seen in the previous analyses were no longer evident in the volcano plots (Figure 3.14). Interestingly, these new analyses greatly weakened the Mount Hutt *R. monroi–R. lobulatus* grouping. The Mount Hutt *R. monroi*–Porters Pass *R. monroi* group had approximately 111% more DEGs compared with Mount Hutt *R. monroi–R. crithmifolius* (Figure 3.15; Supplementary Table 12), and this was similar to the 106% increase in the initial differential expression analysis (Supplementary Table 12). However, in these new analyses, the difference in the number of DEGs in Mount Hutt *R. monroi–R. crithmifolius* over the Mount Hutt *R. monroi–R. lobulatus* grouping increased to 169% (Figure 3.15; Supplementary Table 12).

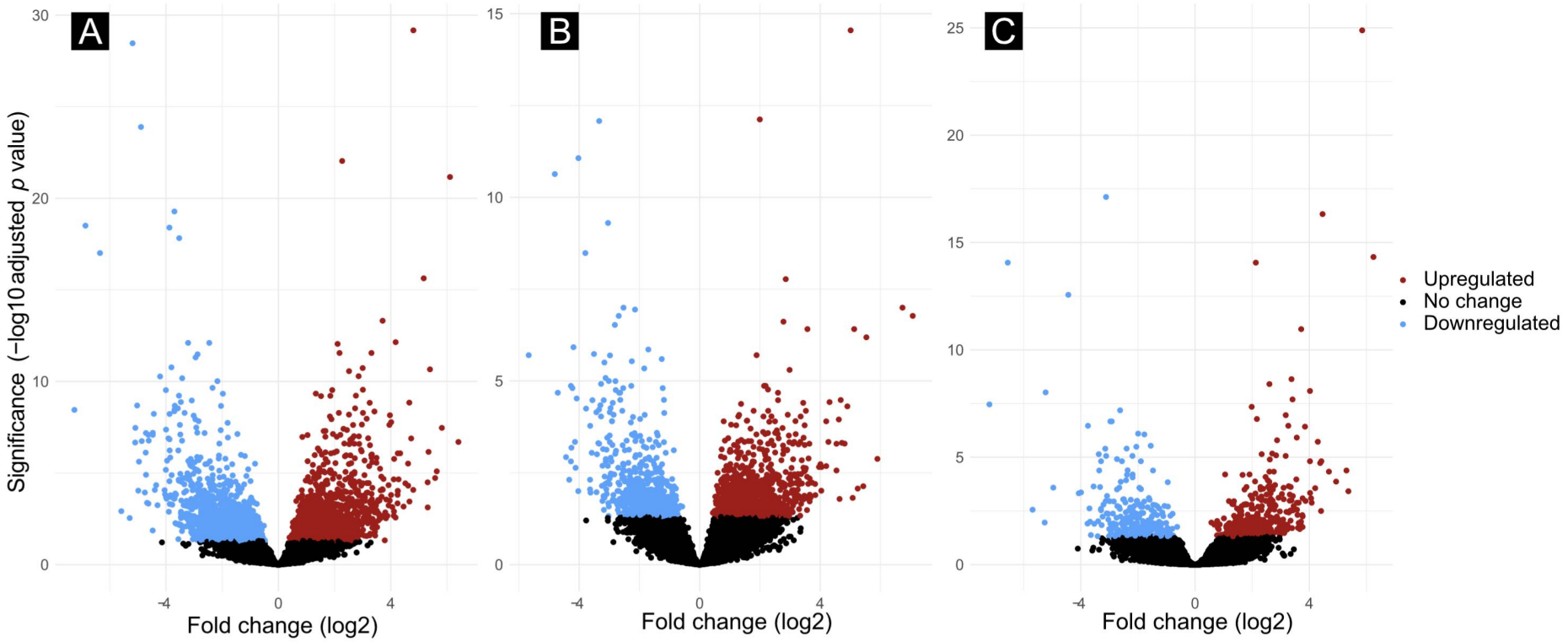


Figure 3.14 Pairwise differential gene expression analyses with DESeq2 after retaining only genes with all samples aligning reads. A, comparison of Mount Hutt *R. monroi*–Porters Pass *R. monroi* against *R. crithmifolius–R. lobulatus* (reference level). B, comparison of Mount Hutt *R. monroi–R. crithmifolius* against Porters Pass *R. monroi–R. lobulatus* (reference level). C, comparison of Mount Hutt *R. monroi–R. lobulatus* against Porters Pass *R. monroi–R. crithmifolius* (reference level). The more obvious data patterns observed in Figure 3.13 are no longer evident.

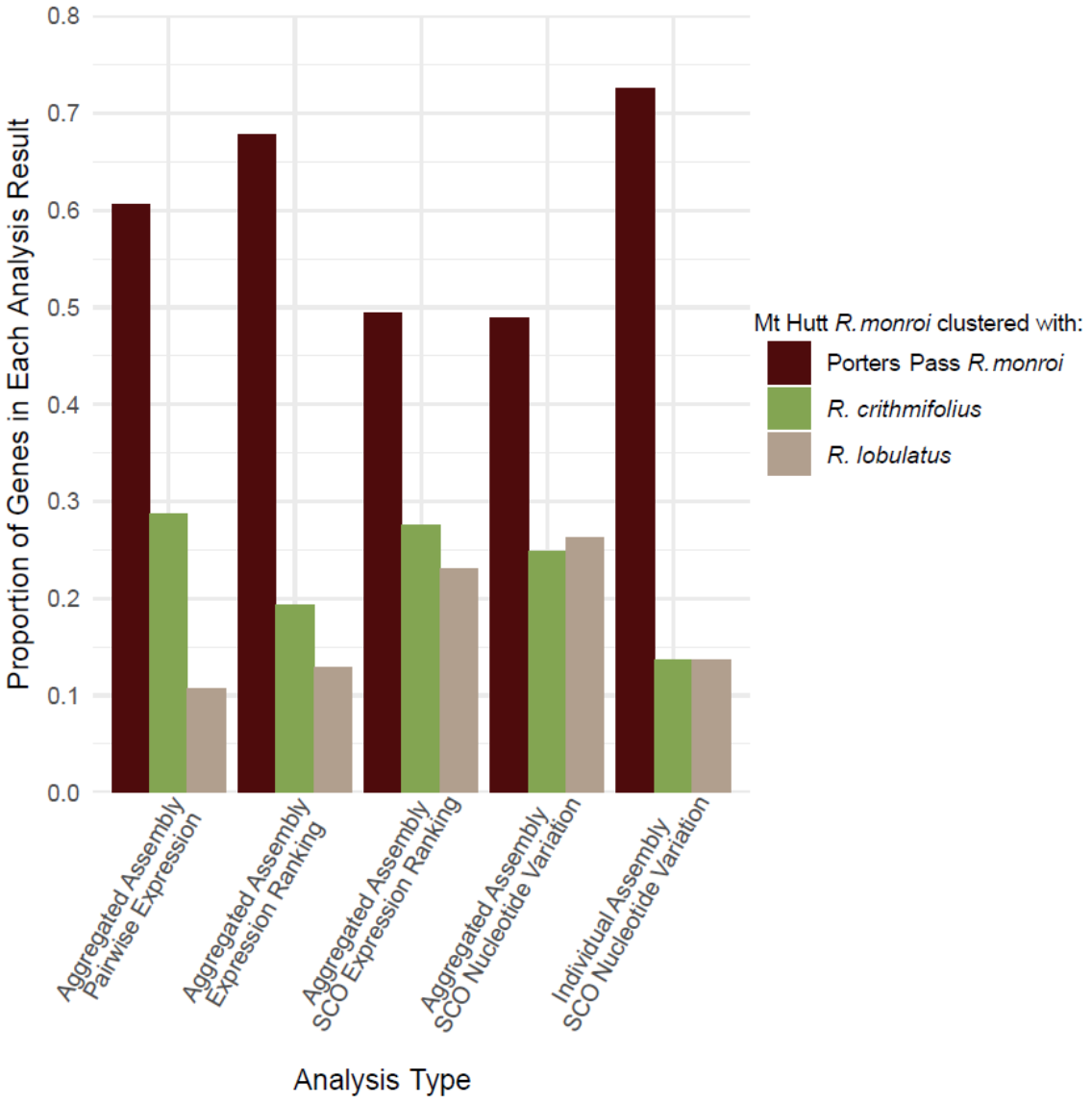


Figure 3.15 Summary plot of gene expression and nucleotide variation analyses. Aggregated Assembly Pairwise Expression: Section 3.4.1 (only genes with all samples mapping retained). Aggregated Assembly Expression Ranking: Section 3.4.2. Aggregated Assembly single-copy orthologue (SCO) Expression Ranking: Section 3.4.4. Aggregated Assembly SCO Nucleotide Variation: Section 3.4.5. Individual Assembly SCO Nucleotide Variation: Section 3.4.7. For each analysis, the number of positive results (DEGs in the case of Aggregated Assembly Pairwise Expression. Groupings or splits that clearly segregate two taxa from the other taxa in analyses using gene expression rankings and nucleotide variation phylogenetic trees) were totalled. Subsequently, the proportion of genes in each split that clustered Mount Hutt *R. monroi* with another taxon was calculated and plotted.

Gene ontology enrichment analysis of the Mount Hutt *R. monroi–R. crithmifolius* grouped DEGs was not informative. Of the 1,704 DEGs, 701 (41%) had GO term annotations (Supplementary Table 12), and testing these against the 12,401 background genes, for which GO term annotations were available, resulted in no significant enrichment of GO terms. For the

Mount Hutt *R. monroi*–Porters Pass *R. monroi* grouping, 12 GO terms were found to be significantly enriched but were not further investigated. The Mount Hutt *R. monroi–R. lobulatus* split produced no enrichment of GO terms.

3.4.2 Phylogenetic analyses of reference assembly expression levels

The similarity gene expression patterns were compared with reference to unrooted taxonomic relationships. After filtering out genes with any sample TPM values of zero, 36,769 genes were retained for use in these analyses. Again, three groupings or splits were considered in comparisons of DEGs: Mount Hutt *R. monroi*–Porters Pass *R. monroi* | *R. lobulatus*–*R. crithmifolius*; Mount Hutt *R. monroi–R. crithmifolius* | Porters Pass *R. monroi–R. lobulatus*; Mount Hutt *R. monroi–R. lobulatus* | *R. crithmifolius*–Porters Pass *R. monroi*. The majority of DEGs (1,036) grouped Mount Hutt *R. monroi* with Porters Pass *R. monroi* while 295 DEGs grouped Mount Hutt *R. monroi* with *R. crithmifolius.* Finally, Mount Hutt *R. monroi* was grouped with *R. lobulatus* 198 times (Figure 3.15; Supplementary Table 12).

Of the 36,769 genes used for expression-level ranking, 13,886 were annotated with *A. thaliana* gene function (Supplementary Table 12). Of these, 12,473 genes were annotated with GO terms (Supplementary Table 12). These 12,473 genes were used as the background for gene ontology enrichment analyses. There was no enrichment of GO terms for genes whose expression levels were most similar between Mt Hutt *R. monroi* and Porters Pass *R. monroi*, nor any enrichment for the Mount Hutt *R. monroi–R. lobulatus* grouping. However, there was enrichment for six GO terms when testing the gene set whose expression levels were most similar between Mount Hutt *R. monroi* and *R. crithmifolius* (Supplementary Table 13).

3.4.3 Single copy orthologue detection

OrthoFinder was successful in finding orthologue sets common to all four taxa. Using the peptide sequences identified by TransDecoder and the Trinity assemblies, OrthoFinder also inferred ‘species’ trees, for both the aggregated assemblies (Figure 3.16, A) and the individual assemblies (Figure 3.16, B) using the STAG algorithm (Emms and Kelly 2018). These consensus trees reconstructed Porters Pass-*R. monroi* and Mt Hutt-*R. monroi* as adjacent. This indicated that patterns of amino acid variation in most of the coding sequences identified by TransDecoder supported a split between *R. monroi* (Mount Hutt and Porters Pass) and the other *Ranunculus* taxa.

Aggregated assemblies produced more SCO sets (orthogroups) than individual assemblies. For the four aggregated assemblies, 5,669 single-copy orthogroups were found. Using the 12 individual assemblies, 4,438 single-copy orthogroups were found.

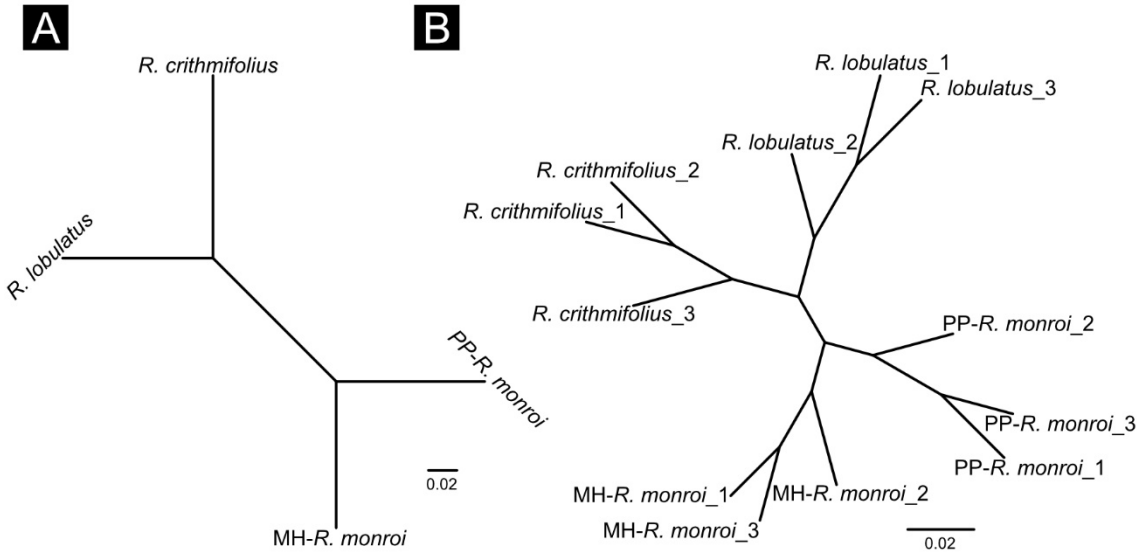


Figure 3.16 Consensus 'species' trees recovered from OrthoFinder using predicted peptide sequences. A, aggregated assemblies. B, individual assemblies. MH: Mt Hutt, PP: Porters Pass.

3.4.4 Phylogenetic analyses of aggregated assembly SCO expression levels

The Figure 3.17 heatmap shows that while expression levels for biological replicates are similar to each other, all SCOs exhibit a high level of gene expression variation. And, like the heatmaps generated by mapping reads to a reference assembly, *R. crithmifolius* remains distant from the other samples. In contrast to the reference assembly read mapping, hierarchical clustering of expression profiles for all SCOs grouped the Porters Pass *R. monroi* samples most closely with the *R. lobulatus* samples (Figure 3.17). However, as indicated by the relative length of internal and external branches, support for this grouping was weak.

When only SCOs in which at least one taxon had a significantly different expression level were examined, the patterns of clustering became more strongly supported (Figure 3.18). Of the 5,669 SCOs discovered, a Kruskal-Wallis test on TPM values, found 932 genes in which at least one taxon was significantly differentially expressed (at $p < 0.05$). In comparing the expression profiles of these genes, it is noticeable that with this filtering of data there was a significant reduction of noise in the data. The expression profiles of biological replicates became more similar to each other, and the relative lengths of internal to external branches increased. The topology most favoured with these data again grouped Porters Pass *R. monroi* and *R. lobulatus* most closely.

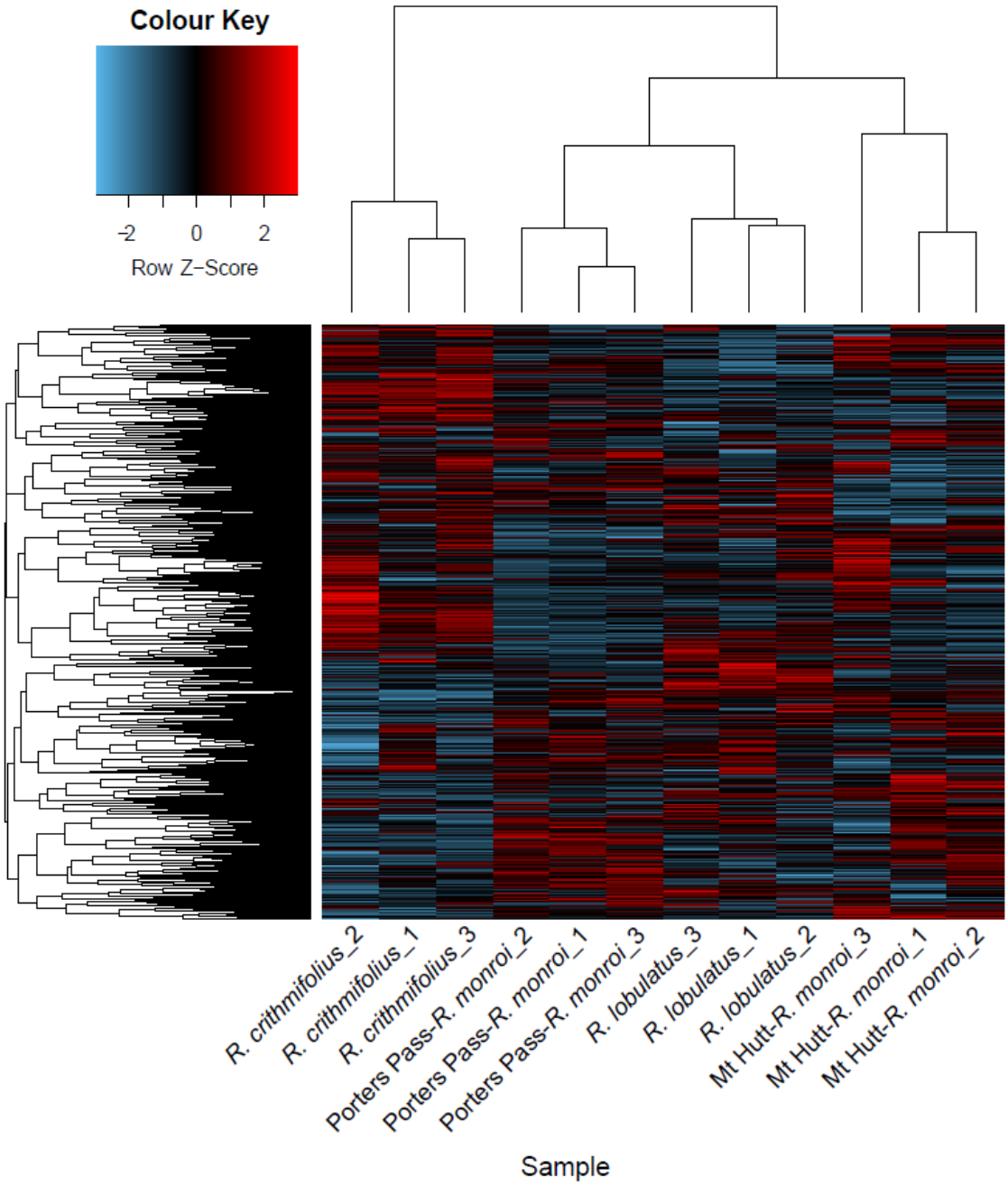


Figure 3.17 Analysis of all SCO gene expression levels. The heatmap shows row-scaled log10-transformed transcripts per million (TPM) normalised values for all (5,669) single-copy orthologous genes (SCOs). Red represents upregulation of genes relative to the other samples, and blue represents downregulation. Each row represents one gene. Columns represent sample RNA-seq reads, and the weighted dendrogram generated using hierarchical clustering with Spearman's rho.

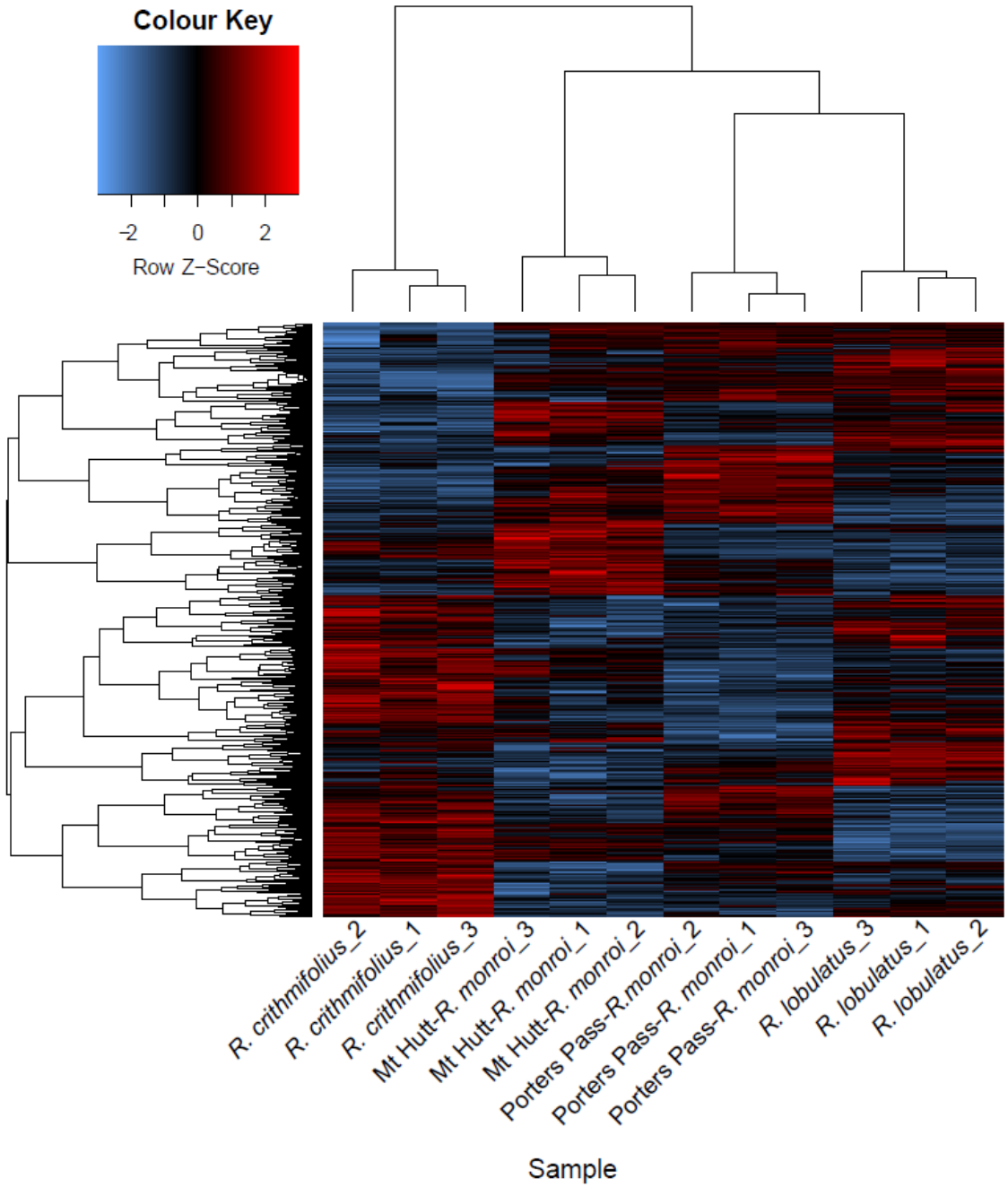


Figure 3.18 Analysis of significantly differentially expressed SCO gene expression levels. The heatmap shows row-scaled log10-transformed transcripts per million (TPM) normalised values for 932 single-copy orthologous genes (SCOs) in which at least one taxon was significantly differentially expressed. Red represents upregulation of genes relative to the other samples, and blue represents downregulation. Each row represents one gene. Columns represent sample RNA-seq reads, and the weighted dendrogram generated using hierarchical clustering with Spearman's rho.

Expression rankings of the 5,669 SCOs indicated that 176 genes had similar expression levels between Mount Hutt *R. monroi* and Porters Pass *R. monroi*. For the Mount Hutt *R. monroi–R. crithmifolius* group, 98 genes showed similar expression levels. Lastly, 82 genes had most similar expression levels between Mount Hutt *R. monroi* and *R. lobulatus* (Figure 3.15; Supplementary Table 12).

Of the 5,669 SCO clusters identified by OrthoFinder, 4,822 (85%) were functionally annotated with *A. thaliana* gene identifiers (Supplementary Table 12). Of these, 4,194 genes were annotated with GO terms (Supplementary Table 12). However, no significantly enriched GO terms were discovered for any of the three taxa groupings.

3.4.5 Phylogenetic analyses of aggregated assembly SCO nucleotide variation

After filtering the SCO MSAs from the aggregated assemblies, 2,669 clusters of high-confidence contigs were generated. After removal of trees with low bootstrap values (< 70%), 1,828 trees were retained. Of the retained trees, 893 trees grouped Mount Hutt *R. monroi* with Porters Pass *R. monroi*, whereas 454 trees supported Mount Hutt *R. monroi* and *R. crithmifolius* as sister taxa, and 481 trees grouped Mount Hutt *R. monroi* with *R. lobulatus* (Figure 3.15; Supplementary Table 12).

Of the high-confidence contigs, 2,159 (81%) could be functionally annotated (Supplementary Table 12). Of the taxa groupings, there were 746 functional annotations for the Mount Hutt *R. monroi*–Porters Pass *R. monroi* split, 362 for Mount Hutt *R. monroi–R. crithmifolius* split, and 391 for the Mount Hutt *R. monroi–R. lobulatus* split.

The 1,835 high-confidence contigs, that were annotated with GO terms, were used as background for GO enrichment analyses. However, in these analyses there was no enrichment of GO terms detected for any of the potential splits.

Of particular interest was whether the expression levels of SCOs were correlated with patterns of nucleotide variation in the same orthologues. And, if so, did these genes encode interesting biological functions. However, few correlations were detected. Thirty genes exhibited similar expression levels and patterns of nucleotide variation in Mount Hutt *R. monroi* and Porters Pass *R. monroi*. Five genes showed similar expression levels and patterns of nucleotide variation in the Mount Hutt *R. monroi* and *R. lobulatus* populations. While 12 genes showed similar expression levels and patterns of nucleotide variation in Mount Hutt *R. monroi* and *R. crithmifolius*. In the latter case, the top protein BLAST hits from the NCBI website are shown in Table 3.8.

Table 3.8 Top BLASTp[a] matches for aggregated assembly SCOs clustering *R. crithmifolius* with Mt Hutt *R. monroi* in expression and nucleotide analyses

OrthoFinder Orthogroup	Protein Name	Protein ID	Organism	GenBank Accession	Query Cover (%)	E-value	Percent Identity
OG0009643.fa	Hypothetical protein	AQUCO_03700003v1	*Aquilegia coerulea*	PIA34435.1	83	2e-136	56.76
OG0010657.fa	Inter alpha-trypsin inhibitor, heavy chain	NA	*Thalictrum thalictroides*	KAF5176287.1	99	0.0	69.64
OG0010696.fa	Hypothetical protein	IFM89_009976	*Coptis chinensis*	KAF9600527.1	98	5e-77	97.44
OG0010969.fa	Hypothetical protein	AQUCO_01700644v1	*Aquilegia coerulea*	PIA45249.1	99	2e-37	69.78
OG0012116.fa	Hypothetical protein	NA	*Saccharolobus solfataricus P2*	CAB57660.1	75	6e-06	39.24
OG0012263.fa	Inositol 2-dehydrogenase	NA	*Thalictrum thalictroides*	KAF5193941.1	99	1e-144	86.52
OG0013128.fa	Hypothetical protein	AQUCO_03400320v1	*Aquilegia coerulea*	PIA36342.1	99	2e-131	87.26
OG0013162.fa	Ferredoxin c 1 protein	NA	*Thalictrum thalictroides*	KAF5198671.1	97	7e-73	73.79
OG0013303.fa	Trafficking protein particle complex subunit 6b	NA	*Thalictrum thalictroides*	KAF5193617.1	98	2e-115	89.47
OG0013520.fa	Dnaj-like protein	NA	*Thalictrum thalictroides*	KAF5197044.1	83	3e-130	68.44
OG0014742.fa	At-rich interactive domain-containing protein	NA	*Thalictrum thalictroides*	KAF5190668.1	99	1e-84	41.72
OG0014995.fa	Hypothetical protein	IFM89_019510	*Coptis chinensis*	KAF9609954.1	94	2e-144	73.27

[a] Via the NCBI web interface

3.4.6 Analyses of heterozygosity in the individual assemblies

The frequency of heterozygous SNPs in orthologues was found to be similar in all biological replicates of all populations. Filtering SCO MSAs of the individual assemblies resulted in 2,541 clusters of high-confidence contigs. From these clusters, the filtered outputs from bcftools (Table 3.9) demonstrated that all 12 assemblies (three assemblies x four populations) returned comparable statistics. Even with strict filtering, across all 12 assemblies there was an average SNP frequency of 0.26% (*SD* = 0.04%). This corresponded to more than two heterozygous SNPs every 1,000 nt. Superficially, it appeared that the percentage of SNPs might be lower in the *R. crithmifolius* assemblies, but a Kruskal-Wallis test failed to detect any significant differences ($\chi^2 = 11$, $df = 11$, $p = 0.443$) between taxa.

Table 3.9 Variant site statistics for individual assembly high-confidence contigs

Assembly	Passed Filters	SNPs	% SNPs	Insertions	Deletions	SNP Transitions/Transversions	Insertion/Deletion ratio	Indel/SNP+MNP ratio
R. crithmifolius _1	8236	7724	0.23	245	267	1.27 (4322/3402)	0.92 (245/267)	0.07 (512/7724)
R. crithmifolius _2	7473	6991	0.20	244	238	1.25 (3882/3109)	1.03 (244/238)	0.07 (482/6991)
R. crithmifolius _3	7025	6575	0.19	230	220	1.26 (3670/2905)	1.05 (230/220)	0.07 (450/6575)
Mt Hutt-*R. monroi* _1	10031	9340	0.27	326	365	1.23 (5148/4192)	0.89 (326/365)	0.07 (691/9340)
Mt Hutt-*R. monroi* _2	11580	10816	0.32	367	397	1.25 (6002/4814)	0.92 (367/397)	0.07 (764/10816)
Mt Hutt-*R. monroi* _3	10451	9731	0.28	344	376	1.30 (5491/4240)	0.91 (344/376)	0.07 (720/9731)
Porters Pass-*R. monroi* _1	7238	6692	0.20	276	270	1.23 (3686/3006)	1.02 (276/270)	0.08 (546/6692)
Porters Pass-*R. monroi* _2	10012	9283	0.27	383	346	1.23 (5129/4154)	1.11 (383/346)	0.08 (729/9283)
Porters Pass-*R. monroi* _3	10321	9583	0.28	352	386	1.24 (5302/4281)	0.91 (352/386)	0.08 (738/9583)
R. lobulatus _1	11711	10980	0.32	351	380	1.21 (6007/4973)	0.92 (351/380)	0.07 (731/10980)
R. lobulatus _2	9954	9185	0.26	388	381	1.21 (5037/4148)	1.02 (388/381)	0.08 (769/9185)
R. lobulatus _3	9596	8970	0.26	298	328	1.23 (4953/4017)	0.91 (298/328)	0.07 (626/8970)

3.4.7 Phylogenetic analyses of individual assembly SCO nucleotide variation

Nucleotide variation was also studied in MSAs of SCOs identified by OrthoFinder analysis of the 12 individual assembles. The taxonomic pattern remained clear in these analyses, but much of the data was not useful. The conservative criterion applied meant that 1,896 orthologue sets were discarded because of low bootstrap values. Of the remaining trees, 543 were rejected for displaying topologies where there was not a strongly supported split separating individuals from two of the four populations. Of trees with the correct topologies, 74 genes supported a Mount Hutt *R. monroi*–Porters Pass *R. monroi* split. Fourteen genes supported the Mount Hutt *R. monroi–R. crithmifolius* split, and the same number supporting the Mount Hutt *R. monroi–R. lobulatus* split (Figure 3.15; Supplementary Table 12).

Of the 2,541 clusters of high-confidence contigs, 2,418 (95%) were annotated with *A. thaliana* gene function. This meant all genes in the groupings, bar one from the Mount Hutt *R. monroi*–Porters Pass *R. monroi* split, were annotated (Supplementary Table 12). Gene ontology enrichment analyses failed to detect significant enrichment, but this was not surprising given the small numbers of genes tested.

Trees conforming with the expected bipartition of populations did not mean individuals were fixed in taxon-specific clades (i.e. monophyletic). Examination of the 14 trees supporting the Mount Hutt *R. monroi* and *R. crithmifolius* split showed that, in two of these trees, individuals of these taxa were not monophyletic (as per Figure 3.19). In seven trees there was non-monophyly of Porters Pass *R. monroi* and *R. lobulatus*. Within the Mount Hutt *R. monroi–R. lobulatus* grouping, there was non-monophyly of the Mount Hutt *R. monroi* and *R. lobulatus* samples four times and the Porters Pass *R. monroi* and R. *crithmifolius* samples once. Similar results were obtained for the Mount Hutt *R. monroi*–Porters Pass *R. monroi* split. Thirty-five trees showed non-monophyly of Mount Hutt *R. monroi* and Porters Pass *R. monroi* individuals, while 15 trees showed non-monophyly of *R. crithmifolius* and *R. lobulatus*.

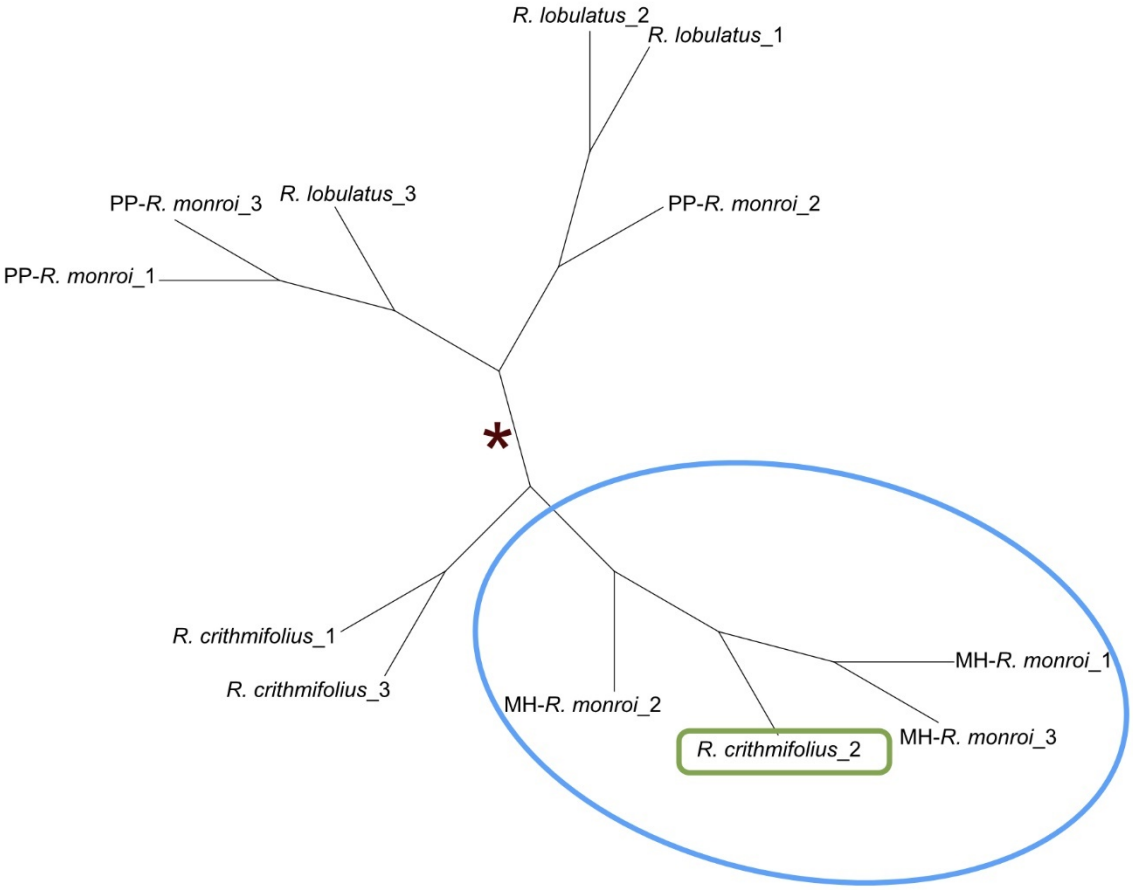


Figure 3.19 Example of non-monophyly in a phylogenetic tree with correct topology. The unweighted maximum likelihood tree displays a topology supporting a Mount Hutt *R. monroi–R. crithmifolius* split. All Mount Hutt *R. monroi* and all *R. crithmifolius* are segregated from the other two taxa by the central edge of the tree. However, the *R. crithmifolius*_2 sample (highlighted in green) is clustered within the Mount Hutt *R. monroi* samples (circled in blue). The dark red star indicates the root position of the tree. PP: Porters Pass, MH: Mount Hutt.

Chapter 4: Discussion

4.1 Habitats

4.1.1 Physical and vegetation factors

Of the physical factors investigated, slope and altitude appear to constitute the greatest differences between Porters Pass and the two Mount Hutt sites. Slope angle at Porters Pass is gentler compared with the three other sites. It is important to highlight there is an inverse correlation between slope and retention of soil (Quansah 1981), and differences in soil composition between sample sites possibly reflect this relationship. Furthermore, slope angle has been shown to influence plant community composition (Zhang *et al.* 2018). Altitudinal differences, on the other hand, contribute to changes in temperature and ultraviolet (UV) (280–400 nm) light. Diurnal temperatures are likely to be significantly warmer at Porters Pass than at Mount Hutt. The environmental lapse rate, or the rate at which air temperature declines with increasing altitude, is—without adjustment for topographical, climatic, and temporal factors (Minder *et al.* 2010; Kirchner *et al.* 2013)—commonly assumed to be a linear decrease of 6.5 °C per 1000 m (Minder *et al.* 2010; Lute and Abatzoglou 2021). In New Zealand, the mean maximum lapse rate for sites above 300 m elevation has been calculated at 6.7 °C per 1,000 m; a value little impacted by changes in latitude (Norton 1985). However, this value typically reflects the day-time lapse rate, while night-time conditions generally result in lower lapse rates that are more variable with regard to seasonality (Norton 1985). More difficult to infer than temperature shifts are differences in solar radiation. However, both UVA (315–400 nm) and UVB (280–315 nm) intensity increase with altitude (Blumthaler *et al.* 1997). And, while plastic responses allow many plant species to tolerate high UV conditions (Barnes *et al.* 2017), decreased fitness—attributed to UV exposure—has been observed in species not altitudinally adapted (Watermann *et al.* 2020).

Conspicuous differences in vegetation cover occur between each of the sites. Although immediately adjoining the *R. crithmifolius* site, the *R. monroi* site at Mount Hutt has a substantially greater amount of vegetation. This likely reflects the differences in stability between the scree upon which the *R. crithmifolius* grows and the face of the rock outcrop where the *R. monroi* is located (Fisher 1952). Ground cover of vascular plants at Lake Tennyson is similar to Porters Pass but there is little non-vascular vegetation and the plant height is much lower. At Porters Pass both vascular and non-vascular plants are in abundance. Most notably,

the dominant vegetation is approximately 70 cm high which contrasts comprehensively with top heights of 0–3 cm at the other sampling sites.

4.1.2 Soil moisture

While measurable variation exists between the two Mount Hutt sampling sites, the Porters Pass site displayed very different soil moisture characteristics over the sampling period. The two Mount Hutt sites showed similar soil moisture readings with the *R. monroi* site being somewhat higher in moisture than the area in which *R. crithmifolius* grows. Soil moisture at Lake Tennyson was intermediate between Mount Hutt and Porters Pass. The lack of observed overlap in soil moisture (1 x *SD*) indicates that water availability for flora at the three main locations (Porters Pass, Mount Hutt, Lake Tennyson) is dissimilar over the longer term. At Mount Hutt, water availability within the two *Ranunculus* niches also appeared to differ but this difference was not as apparent. Spatial separation of plant species according to available moisture is not a surprising finding (Craine *et al.* 2013; Silvertown *et al.* 2015; Camarero *et al.* 2018; Taseski *et al.* 2021), although the mechanism underlying distributions can differ among species. Often, range limits are attributed to morphological and physiological characteristics of mature plants (Talbi *et al.* 2015; Anderegg and Hille Ris Lambers 2016; Ferris and Willis 2018), but Yang *et al*. (2017a) found soil moisture to be a key distributional determinant for pioneering plants in desert areas due to its effect on seed germination rates. The response of seedlings to environmental variables is another important component of species' habitat preferences (Johnson *et al.* 2017). Importantly, Taseski *et al*. (2021) discovered intra-specific trait differences allowed some species to broadly occupy habitats with different water availability, while species without adaptive or plastic capacity were excluded from niches.

In contrast to Porters Pass, soil moisture at both the Mount Hutt *R. monroi* and *R. crithmifolius* sites was relatively stable over the sampling period. However, extended monitoring is needed to determine longer term trends. One interesting aspect of soil moisture at each of the sites (particularly Porters Pass and Lake Tennyson) was the wide standard deviations observed. Since relatively small areas were sampled, this illustrates the high degree of spatial heterogeneity in soil moisture levels that can occur in natural systems (Choi *et al.* 2007; Chen *et al.* 2010). Interestingly, the data logger at Porters Pass recorded relatively low soil moisture at the beginning of the uninterrupted monitoring period, with the moisture substantially increasing over approximately three months. This behaviour raised the question as to whether this was a monitoring artifact of the soil around the sensors settling. However, the three days of data collected before data logger failure indicate soil moisture was much higher at logger installation. Furthermore, the data reflect weather conditions at that time. There was

lower than average rainfall, and subsequent lower soil moisture, in the eastern South Island (NIWA 2021b) during that period. Moreover, November 2019 was the hottest November on record for New Zealand (NIWA 2021a), and a similar, albeit less aggressive, soil moisture trend was observed for Lake Tennyson during part of the same period.

Attempts to generate a longer time series of soil moisture using computer-calculated data were unsuccessful. The Virtual Climate Station Network (VCSN) is a grid of 11,491 virtual climate stations maintained by NIWA, positioned across the New Zealand landscape, with daily estimates of climatic variables. Rainfall data extends back to 1960 and data for other variables goes back to 1972. Measurements of the climatic variables are calculated by interpolation of data measured at nearby physical weather stations. Error in rainfall data for some sites in the VCSN network is a known issue (Cichota *et al.* 2008), particularly for high-altitude areas where physical sampling is sparse (Tait *et al.* 2012). For this reason, Tait *et al*. (2012) recommend using detailed on-site measurements to calibrate the VCSN rainfall data. Unfortunately, at Mount Hutt, there was poor correlation between rainfall events indicated by the VCSN data and data logger measurements of increases in soil moisture. Conversely, often no increase in soil moisture was measured during, or soon after, large rainfall events recorded by the VCSN.

4.1.3 Soil physical characteristics

Soil particle size distributions are very different at Porters Pass, and correlate with the soil moisture observations. Soil particle size analyses indicate clear differences between Porters Pass and all other sites. The smallest sieve size used for soil analyses was 0.25 mm. Below this size, soil particles are classed—in decreasing particle sizes—as fine sand, very fine sand, and various types of silt and clay (Soil Science Division Staff 2017). Compared with the soils from Mount Hutt and Lake Tennyson, which are composed of large amounts of coarse sands and gravel, the Porters Pass soils are primarily made up of much finer materials. A major contributor to a soil's water holding capacity (WHC) is soil particle size. Decreasing particle sizes means smaller pore sizes and increasing surface area (McLaren and Cameron 1996). The cohesive and adhesive nature of water molecules means water movement under these conditions is minimised and WHC increased, although excess clay content might limit plant water availability (McLaren and Cameron 1996). Certainly, physically altering soil particle size distribution is an aid to increasing WHC in arable soils (Suzuki *et al.* 2007), and changes in porosity are an important component in improving productive pasture through the addition of biochar (Liu *et al.* 2017; Verheijen *et al.* 2019).

Robust comparison of soil physical characteristics between sites and sampling depths was difficult because, excepting Porters Pass < 0.25 mm data, the heterogeneity of soil

contributed to large 95% CIs for each fraction. In many cases, the CI for a given fraction was larger than the mean. Superficially however, at both sampling depths, the soils at the two Mount Hutt sites appear to be composed of similar particle sizes. Interestingly, the samples from Lake Tennyson appear anomalous with respect to the relationship between soil particle sizes and WHC. The mean proportions for the < 0.25 mm fractions are the lowest at Lake Tennyson, yet, data indicate this site has a soil moisture profile intermediate between Mount Hutt and Porters Pass. This suggests a factor (or factors) in addition to particle size that is influencing soil moisture.

4.1.4 Soil chemical composition

The chemical composition of soil at Porters Pass is different to that at the two Mount Hutt sites. The soil that Mount Hutt *R. monroi* grows in is chemically similar to the soil in which *R. crithmifolius* (at Mount Hutt) is found. There is also greater heterogeneity at Porters Pass than at either of the Mount Hutt sites. At Mount Hutt, PCA implies *R. crithmifolius* and *R. monroi* have slightly different requirements for, or tolerances to, soil chemistry. Although, this variability between these two sites might have arisen due to differences in physical stability since *R. crithmifolius* inhabits what appears to be a more mobile slope. However, the somewhat intermediate nature of chemistry (between Porters Pass and the *R. crithmifolius* site) for the Mount Hutt *R. monroi* soil, suggests soil chemistry might be physiologically important.

Although PCA segregated the sites according to soil chemistry, the physiological implications of chemical differences are not immediately apparent. Particularly, as not all chemical forms of nutrients are available for plant uptake (Kirkman *et al.* 1994; Webb 1995). Mean pH for each site is within a range expected for New Zealand alpine soils which are typically acidic (Manaaki Whenua – Landcare Research 2020). Total P was low at Porters Pass but plant-available P at this site was similar in concentration to the Mount Hutt *R. monroi* site. New Zealand soils are typically considered to be low in N (Webb 1995), and while Porters Pass had higher N than either of the Mount Hutt sites, high C:N of soil at Porters Pass means plant availability of this N might be limited (McLaren and Cameron 1996). Porters Pass is very rich in OM, and this high OM likely explains the high CEC which is a measure of a soil's ability to bind and make available positively charged ions (Ketterings *et al.* 2007). The Mount Hutt sites have much higher TBS than Porters Pass. This high TBS reflects a higher concentration of base cations (K^+, Na^+, Ca^{2+}, Mg^{2+}) and low CEC values, and this is important because it represents a plant-available source of these nutrients (McLaren and Cameron 1996). Interestingly, the low TBS measurements at Porters Pass might suggest higher levels of acid cations (H^+, Al^{3+}) at this site (McLaren and Cameron 1996).

Organic matter might explain the surprising WHC of soil at Lake Tennyson. Certainly, rainfall is a major contributor to soil moisture (Li *et al.* 2016), but a lack of crossover in daily mean soil moisture between Lake Tennyson and the other sites, implies that physical factors explain the soil moisture profile better than weather events. One element potentially underlying WHC of soil at Lake Tennyson is the relatively large amount of OM. Although the < 0.25 mm fractions at Lake Tennyson comprise only a small proportion of soil, they are made up of 23.5% and 15.4% OM for 0–5 cm and 5–10 cm depths, respectively. Contrastingly, the mean percentage of OM (both depths) at the two Mount Hutt sites was 0.8% for *R. crithmifolius* and 2.7% for *R. monroi*. Furthermore, Porters Pass, high in soil moisture, had large amounts of OM. While exactly quantifying the effect of OM on soil moisture is difficult (Libohova *et al.* 2018), increased OM produces soils with greater WHC (Saxton and Rawls 2006; Acín-Carrera *et al.* 2013), and these increases in WHC appear to be most pronounced in sandy soils (Minasny and McBratney 2018).

4.2 Experimental plants

4.2.1 DNA barcoding

The *Ranunculus* phylogenies constructed from DNA barcoding data displayed discordant relationships among taxa. However, despite a putative lack of reproductive barriers, there is no evidence in these data to support contemporary gene flow between *R. crithmifolius* and *R. monroi* at Mount Hutt. Sequencing of the ITS2 region returned a phylogenetic network that conforms to the currently accepted alpine *Ranunculus* taxonomy (Webb *et al.* 1988), This finding is strengthened by analysis of sequences after removal of sites with ambiguous characters. Compared with the ITS2 data, *ycf1* sequences indicate a more complex pattern of relationships among taxa.

While the Neighbor-Net of ITS2 sequences segregates the *R. crithmifolius* samples into a discrete cluster remote from the *R. insignis* forms, no clear clusters of *R. monroi* or *R. lobulatus* emerge. This appears consistent with the findings of Lehnebach (2008) who positioned *R. monroi* samples either side of *R. lobulatus* in a phylogenetic tree obtained by parsimony analysis. Interestingly, Lehnebach (2008) detected only a single *R. lobulatus* genotype that was shared with North Island *R. insignis*. Contrastingly, in this study, several *R. lobulatus* genotypes were found when ambiguous sites were retained in the ITS2 MSA. These might have resulted from the larger number of samples used in these analyses but potentially reflect the difficulty in manually validating DNA sequences by examining Sanger sequencing electropherograms. More work is needed to identify polymorphisms among these

sequences; a process which can be technically challenging. When researching *Cycas* (Cycadaceae), Xiao *et al*. (2010) resorted to cloning multiple ITS fragments from each individual to identify ITS paralogues, although polymorphisms can sometimes be inferred by assembly of sequences generated through whole genome sequencing reads (Ganley and Kobayashi 2007).

The mechanisms underlying the polymorphic nature of the alpine *Ranunculus* ITS2 regions are difficult to determine, and it is unknown if the polymorphisms are ancestral or more recently arisen. Certainly, analysis of the ITS2 region provides no evidence of introgression between Mount Hutt *R. monroi* and *R. crithmifolius*. After a hybridisation event, rapid concerted evolution typically results in all copies of a species' ITS region sharing DNA sequence similarity, primarily through a process of unequal recombination (Elder Jr and Turner 1995; Xiao *et al.* 2010). Sequence similarity of the ITS region in a lineage might be maintained either through further concerted evolution (Ganley and Kobayashi 2007; Naidoo *et al.* 2013) or via selection against deleterious alleles (Nei and Rooney 2005; Rooney and Ward 2005). Concerted evolution can be incomplete which means polymorphic ITS sequences can be retained or arise within an individual's genome (Xu *et al.* 2017), particularly if loss of function occurs in an ITS array (Xiao *et al.* 2010). However, factors underlying incomplete concerted evolution can be many (Vollmer and Palumbi 2004) and determination of the mechanisms often inconclusive (Vollmer and Palumbi 2004; Xu *et al.* 2017). With these alpine *Ranunculus* samples, it is currently possible to infer only the existence, and not the underlying cause, of intra-genomic variation in the amplified ITS2 regions. And furthermore, due to PCR amplification biases (Polz and Cavanaugh 1998) or differences in intragenomic variant ITS copy number (Ganley and Kobayashi 2007), the complete ITS2 profile for each taxa cannot be stated with certainty.

Chloroplast haplotype data suggest more complex evolutionary histories than can be inferred from the ITS2 analyses. Incongruence in phylogenetic trees constructed with chloroplast data (Butcher *et al.* 1995), or incongruence between trees constructed with chloroplast and nuclear data (Pirie *et al.* 2010), has been used to infer instances of hybridisation. In the alpine *Ranunculus*, phylogeographic patterns of chloroplast haplotype sharing between taxa have been hypothesised to reflect past introgression (Lockhart *et al.* 2014). However, the molecular signature of incomplete lineage sorting—the persistence of ancestral alleles which are later retained or lost in different populations of an extant species (Maddison 1997)—can be similar to that of introgression (Pease and Hahn 2015). This means that robust statistical methodologies are required to support hypotheses of introgression (Joly *et al.* 2009b; Joly 2012; Lockhart *et al.* 2014). The results of the present study tentatively support previous findings which inferred non-monophyly of *R. monroi*, and clustered Mount Hutt *R. monroi* more closely with accessions of *R. crithmifolius* (Carter 2006; Lehnebach 2008). In the current analyses, the close relationship between the three *R. monroi* haplotypes detected at Porters Pass is reflected

by their position on the same external branch of the phylogenetic network. Conversely, Mount Hutt *R. monroi* is positioned on a different branch. Unfortunately, even given the relatively rapid mutation rate of the *ycf1* gene (Dong *et al.* 2015), the *ycf1* fragments used in these analyses are too short, with too few phylogenetically informative sites, to resolve the relationships any further.

4.2.2 Plant growth

The environmental conditions set for the plant growth experiment were conducive to active plant growth. Water usage and pulse amplitude modulated fluorometry (PAM) measurements indicate that plants harvested for analyses were healthy and not undergoing excess stresses.

Providing controlled conditions for all individuals of different plant taxa is not a trivial exercise. Random replacement of plants after each watering was used to reduce positional effects on plant responses. The limited number of plants and relatively small cabinet space made this strategy preferable to the random block design recommended by Hartung *et al*. (2019) for greenhouse experiments. Importantly, cabinet position appears to contribute to water usage but no environmental factor (light, temperature, etc.) underlying this finding can be determined. No *R. monroi* plants were lost during this experiment, and only one *R. lobulatus* individual was disposed of due to insect infestation. Unfortunately, four *R. crithmifolius* plants were also lost. This poor survival rate of *R. crithmifolius* plants, during both this study and the pilot study, illustrates the difficulty in providing hospitable conditions for plant taxa which naturally occur in different environments. However, photographs taken at watering, water usage, and PAM measurements, indicate the plants harvested were in a healthy state and subject to similar water availability.

Water usage over the course of the growth experiment increased. This would seem to illustrate rising transpiration rates as plant growth increased. There is a strong correlation between plant dry matter and transpiration rate (Arkley 1963). There was one outlying value at the second measurement which is unexplained. However, given it occurs near the beginning of the experiment, it is potentially the result of plant acclimation to the experimental conditions.

Measured dark-adapted photosynthetic efficiency levels among the plants were within expected values for unstressed plants. Both before and after watering, mean Fv/Fm for all taxa was approximately 0.8. The slight reductions in Fv/Fm after watering suggest the 4-day watering regimen is satisfactory for the growth of these plants. Interestingly, the reductions were only significant for Porters Pass *R. monroi* and *R. lobulatus*. Additionally, in the unwatered state, Mount Hutt *R. monroi* and *R. crithmifolius* displayed similar Fv/Fm values, as

did Porters Pass *R. monroi* and *R. lobulatus*. Taken together, these findings suggest similarities in response to the growing conditions and water availability. However, small sample sizes caution against overinterpreting these results. Although Fv/Fm is not a universal indicator of plant stress, in that only one tissue type is evaluated (Murchie and Lawson 2013), abiotic stresses commonly affect photosynthesis (Ashraf and Harris 2013; Lu *et al.* 2019) and are often first apparent in damage to PSII (Maxwell and Johnson 2000). In *A. thaliana*, Barbagallo *et al*. (2003) determined an Fv/Fm value of approximately 0.8 was indicative of plant health. Of the 44 plant species surveyed by Björkman and Demmig (1987), Fv/Fm in healthy plants ranged from 0.778 to 0.860 with a mean among C3 plants of 0.832. Values lower than these are generally considered suggestive of stress (Maxwell and Johnson 2000; Murchie and Lawson 2013).

Plants had similar available water at harvest. This is important because soil moisture variability is well understood to drive physiological changes (Puértolas *et al.* 2017) and alterations in gene expression (Lawlor 2012; Fracasso *et al.* 2016). This can be particularly problematic, for experiments in which plants are grown in containers of a limited size, when morphological and physiological differences between plants, rapidly drive dissimilar soil moisture profiles (Lawlor 2012). In this experiment, final mean soil moisture differed by only 0.023 m^3/m^3 among all taxa. Furthermore, no significant differences in final soil moisture were detected. Although, it must be noted, once again, that small sample sizes result in decreased statistical power.

4.3 RNA

4.3.1 Extraction of RNA, and sequence read quality control

Extraction of RNA from plants can be problematic. In this study, *R. crithmifolius* was particularly recalcitrant. A number of RNA extraction methods failed to produce high-quality RNA from this plant. These methods included using manufacturers' instructions for different RNA extraction kits: E.Z.N.A.® Plant RNA Kit, RNeasy PowerPlant Kit (Qiagen), Sigma-Aldritch® Solutions Spectrum™ Plant Total RNA Kit (Merck KGaA, Darmstadt, Germany), Direct-zol RNA Miniprep Kits (Zymo Research, Irvine, USA), as well as an in-house protocol using TRIzol Reagent (Thermo Fisher Scientific). The published protocols tested (Kalinowska *et al.* 2012; Ouyang *et al.* 2014) also failed to successfully extract RNA. New Zealand alpine scree and rock plants, such as *R. crithmifolius*, commonly exhibit red or grey colouration (Strauss *et al.* 2015), which is attributed to increased concentrations of anthocyanins in the leaves (Niu *et al.* 2014). Being phenolic compounds (Khoo *et al.* 2017), anthocyanins are

thought to become oxidised during the RNA/DNA extraction process and irreversibly bind nucleic acids (Salzman *et al.* 1999; Vennapusa *et al.* 2020). Unfortunately, the recommended addition of PVP (polyvinylpyrrolidone) to extraction buffers (Salzman *et al.* 1999; Sánchez *et al.* 2016) failed to improve results with *R. crithmifolius*. Surprisingly, the addition of Qiagen's PSS Solution to the E.Z.N.A.® Plant RNA Kit resulted in successful RNA extraction, despite the failure of the RNeasy PowerPlant Kit of which PSS solution is a component. Understanding the mechanism by which the PSS solution improved RNA extraction would potentially aid in the future study of many alpine plants.

RNA-seq read error correction, trimming and quality filtering was successful. Less than 4% of reads per sample were lost through the quality control process. It has been argued that gentle trimming of reads, by setting a low (PHRED ≤ 5) nucleotide quality threshold, or not trimming at all, improves transcriptome *de novo* assembly (MacManes 2014). Conversely, it has been contended that uncorrected reads align back to assemblies with reduced efficiency (Payá-Milans *et al.* 2018). Regardless, MacManes (2014) believed a reduction in coverage, resulting from vigorous trimming of reads, led to poor assemblies. Whereas, in this study, the high proportion of retained reads for each sample appears to alleviate that concern.

4.3.2 Transcriptome assemblies and representative contigs—protein prediction and annotation

The quantities of contigs and gene clusters generated were as expected for Trinity *de novo* assemblies. Typically, Trinity reconstructs large numbers of transcripts and genes (Liu *et al.* 2018; Moreno-Santillán *et al.* 2019). And, it has been shown that these numbers are increased in polyploid plants when compared to diploids (Madritsch *et al.* 2021). Notably, the aggregated alpine *Ranunculus* assemblies consist of more contigs and clusters than the individual assemblies. Li *et al*. (2019a) found that increasing input data for *de novo* assembly of the tea (*Camellia sinensis*) transcriptome increased contigs even after estimates of transcriptome completeness had stabilised. While it has been noted that excessive read numbers can result in misassembles (Conesa *et al.* 2016), deeper sequencing is expected to detect more (rare) isoforms (Haas *et al.* 2013). Additionally, Trinity performs *in silico* k-mer coverage normalisation to decrease complexity of the assembly process (Haas *et al.* 2013).

The assembly statistics for aggregated assemblies are generally similar to one another, as are the statistics for the individual assemblies. Median and average contig lengths, %GC, and N50 are similar for all assemblies. Average contig lengths for each assembly compare favourably against results from studies using similar assembly methods (Moreno-Santillán *et al.* 2019; Wang *et al.* 2019b) although much longer average contig lengths are typically reported

for studies using long-read sequencing technologies (Yang *et al.* 2018a; Luo *et al.* 2019a; Qiao *et al.* 2019). Although higher N50 measurements have been reported for *de novo* transcriptome assemblies (Zhang *et al.* 2020b), the N50 values in this study are similar to (Liu *et al.* 2018; Li *et al.* 2019b) or higher than (Bhardwaj *et al.* 2015; Zhao *et al.* 2016; Mahmood *et al.* 2020) many reported. Surprisingly, given the common reporting of median length, average length, and N50 for *de novo* transcriptome assemblies, these metrics are not considered reliable indicators of assembly quality (O'Neil and Emrich 2013; Li *et al.* 2014).

BUSCO scores and remapping of reads better indicate quality of *de novo* transcriptome assemblies. Completeness for all the aggregated assemblies was assessed to be over 91% when assessed against the OrthoDB Embryophyta database. That is, most of the single-copy orthologous genes expected to be present in the transcriptomes were recovered. The high rate of duplicate detection is partially due to the alpine *Ranunculus* being polyploid species. Madritsch *et al*. (2021) found that *de novo*-assembled transcriptomes of autotetraploid plant species consistently contained greater numbers of duplicated BUSCOs than transcriptomes of the diploid relatives. However, duplication rates over 50% have been reported for transcriptomes of diploid plant species (Qiao *et al.* 2019), and *de novo* assemblies comprised of short-read data might contain a greater number of redundant contigs than transcriptomes generated with long-read data (Hoang *et al.* 2017). For all aggregated assemblies, read remapping is above 93%, and although only two individual assemblies had reads self-mapped, both had an alignment rate over 90%. These data indicate most reads are incorporated in the assemblies and are suggestive of high-quality assemblies (Haas *et al.* 2013; Moreno-Santillán *et al.* 2019). The high rate of multimapping reads can be attributed to exon sharing of different transcript isoforms (Conesa *et al.* 2016) and/or the number of redundant transcripts in a Trinity assembly (Hoang *et al.* 2017). Although, mapping to different homeologous (or ohnologous) genes cannot be discounted as another factor influencing multimapping of reads.

Selecting the most highly expressed contig of each Trinity cluster as representative isoform of each gene, is an effective way to reduce redundancy in transcriptome assemblies. Removal of redundant and alternatively spliced transcripts is typically undertaken in studies utilising *de novo* transcriptome assemblies (Bhardwaj *et al.* 2015; Xu and Huang 2018; Neji *et al.* 2019; Zhou *et al.* 2019; Madritsch *et al.* 2021). Commonly, CD-HIT-EST (Li and Godzik 2006) is used to re-cluster contigs based on a sequence similarity threshold, and the longest contig of each new cluster is retained as representative (Xu and Huang 2018; Neji *et al.* 2019; Madritsch *et al.* 2021). Also available, are alternative transcript clustering software such as Corset and Grouper (Malik *et al.* 2018) which use similar approaches to each other in that they both utilise read alignment data for clustering (Davidson and Oshlack 2014; Malik *et al.* 2018; Chen *et al.* 2019). Surprisingly, given its common usage, CD-HIT-EST performed poorly compared with Grouper, Corset, and the internal Trinity algorithm for clustering transcripts of

polyploid plant species (Chen *et al.* 2019). And interestingly, Chen *et al*. (2019) found the latter three methods performed transcript clustering comparably, although they caution that all clustering efficiency is reduced in neo-polyploids. In the present study, the very low percentages of reads aligning more than once against transcripts of the representative assemblies indicates Trinity has not under-clustered isoforms. Contrastingly, concordant multi-mapping against the rnaSPAdes-assembled Mt Hutt-*R. monroi*_1 representative contigs, clustered using Corset, occurred with 46.89% of reads.

An interesting question concerns the larger number of TransDecoder predicted proteins in some assemblies. Logically, the increased number of TransDecoder hits for the *R. crithmifolius* and Mt Hutt-*R. monroi* initial assemblies is driven by their large numbers of gene clusters. And, it is also apparent these aggregated assembly results are driven by the TransDecoder predicted proteins and number of gene clusters in the *R. crithmifolius*_2 and Mt Hutt-*R. monroi*_3 individual assemblies. It was initially thought the increased number of clusters might reflect chloroplast contamination. However, the low alignment rate of Mt Hutt-*R. monroi*_3 reads (0.09%) and the *R. crithmifolius* reads (0.10%–0.12%) against the *R. sceleratus* chloroplast genome does not support that hypothesis. Furthermore, no RNA extraction data such as LabChip® RQS, nor assembly summary statistics imply excessive RNA fragmentation in the *R. crithmifolius*_2 and Mt Hutt-*R. monroi*_3 samples. This raises the question as to whether these two samples contain biologically real differences in gene expression when compared against samples of the same taxa. The results of searching *A. thaliana* proteins with BLASTp indicate the increased numbers reflect different gene expression profiles. Although the number of unique BLAST hits for *R. crithmifolius*_2 and Mt Hutt-*R. monroi*_3 are not as inflated as the many thousands of TransDecoder protein sequences predicted for these assemblies, both have hits against a greater number of *A. thaliana* genes and isoforms compared with remaining individual assemblies of the same taxa.

4.3.3 Chloroplastic sequences

Nucleic acids derived from cpDNA were detected. It is worth noting that, even after treatment of the total RNA using DNase 1, and mRNA enrichment during sequencing library preparation, chloroplast sequences were detected in the Mt Hutt-*R. monroi* assembly. DNase 1 is expected to efficiently digest cpDNA (Shi *et al.* 2012). However, the possibility exists that cpDNA remained in the final total RNA eluate after RNA extraction, and this was retained during mRNA enrichment due to the presence of adenine-rich repeat regions (Choi *et al.* 2016). Alternatively, RNA transcripts derived from cpDNA may have remained after mRNA enrichment, regardless of chloroplast mRNA tails being merely adenine-rich regions rather than

the ribohomopolymers characteristic of mRNA transcribed from the nuclear genome. Transcription of multiple genes as polycistronic transcripts (Sugita and Sugiura 1996; Shi *et al.* 2016) could account for the length of contigs detected. Interestingly, the Mt Hutt-*R. monroi* representative assembly does not include the very long contig of the Mt Hutt-*R. monroi* aggregated assembly, so this indicates at least one other Mount Hutt *R. monroi* sample contains reads derived from cpDNA which supported a chloroplast contig of a shorter length. Alignment of reads against the *R. sceleratus* genome supports this hypothesis. The overall alignment percentages of samples Mt Hutt-*R. monroi*_1 and Mt Hutt-*R. monroi*_2 are much higher (2.29% and 3.46% respectively) than for any other sample, the highest of which is 0.52% for *R. lobulatus*_1. These data suggest a small amount of chloroplast contamination in samples. However, the low level of contamination present, and later filtering steps such as orthologue finding, mean gene expression and nucleotide variation analyses should be unaffected.

4.4 Gene expression and phylogenetics

What has been considered a confounding variable in gene expression studies provides an opportunity in this research. Inter-specific and multi-specific comparisons of gene expression have been criticised when phylogeny is not accounted for (Rohlfs and Nielsen 2015; Dunn *et al.* 2018). This is because trait sharing is more likely between taxa with a shared evolutionary history (Dunn *et al.* 2018). If the above concerns are correct, then gene expression data should contain phylogenetic signals. This hypothesis is supported by the results of this study, and phylogenetic patterns of gene expression are compared with phylogenetic inferences from nucleotide variation.

4.4.1 Standard pairwise comparisons of gene expression

Performing gene expression analyses by aligning reads to a single reference is problematic if the reference is not adequate. Concerns of biased read mapping arising through use of a single reference during heterospecific gene expression analysis, have been previously raised (Voelckel *et al.* 2012, 2017). Importantly, three of the four taxa under investigation in this study are considered the same species but biased read alignment is still evident. The degree to which these biases are influenced by phylogenetic relatedness versus technical artifacts requires more in-depth analysis. Reads align more successfully to a reference of the same taxa, but remapped reads which were used to construct the assemblies align with greater success than reads of the same taxa not used in assembly construction. It would seem the ideal reference is well annotated and phylogenetically equidistant from the study taxa. Unfortunately, in testing,

less than 26% of *R. lobulatus*_1 reads successfully aligned to the *Aquilegia coerulea* transcriptome even when stringency was reduced. Therefore, careful selection of a reference and optimisation of read alignment parameters are needed. Of course, this approach might allow for investigation of the genes and processes underlying physiological responses but does not immediately convey inference of allelic variation.

Using a single reference does allow for gene expression analysis to be performed on many genes, but results caution against arbitrarily grouping taxa together. In this study, a greater number of DEGs were detected by aligning reads to the *R. lobulatus* assembly than were found by performing DGE analyses on SCOs. Additionally, aligning to a single reference means software such as DESeq2 and edgeR, which do not use normalised transcript counts as input (Love *et al.* 2014; Conesa *et al.* 2016), can be used to successfully execute the analyses. However, importantly, this study demonstrates that, when grouping taxa together to perform the analysis, false inferences of differential expression can arise from inflated gene expression levels of one taxon.

Standard pairwise comparisons with DESeq2 give results consistent with hybridisation and introgression between the two Mount Hutt taxa. Of DEGs (after filtering of genes with all samples aligning reads) detected when the two *R. monroi* taxa are clustered together, shared expression occurs in genes more frequently than is the case for either of the other possible groupings. But interestingly, grouping Mount Hutt *R. monroi* with *R. crithmifolius* more than doubles the DEGs detected compared with the final and most geographically/biologically unlikely grouping (Mount Hutt *R. monroi* and *R. lobulatus*). An alternative explanation for this result is that, rather than expression similarities between Mount Hutt *R. monroi* and *R. crithmifolius*, there is shared expression between Porters Pass *R. monroi* and *R. lobulatus*. However, the ecological and geographic distributions of these taxa make this scenario less probable. Therefore, this result is consistent with the hypothesis of ancestral introgression of *R. crithmifolius* genetic material into *R. monroi* at Mount Hutt.

However, introgression is not the only explanation for greater shared expression between Mount Hutt *R. monroi* and *R. crithmifolius* because incomplete lineage sorting or epigenetic regulation might also explain these data. Often, putative instances of introgression are not strongly supported because researchers have difficulty ruling out incomplete lineage sorting being responsible for allele sharing between taxa (Morgan-Richards *et al.* 2009; Lockhart *et al.* 2014; Jay *et al.* 2018; Suarez-Gonzalez *et al.* 2018). Although, when the necessary molecular data is available, statistical testing can robustly assess hypotheses of introgression (Joly 2012; Hibbins and Hahn 2019; Lee-Yaw *et al.* 2019). Epigenetics might also have influenced gene expression in the test plants. Although plants were acclimated to homogeneous conditions, epigenetic regulation can continue to influence gene expression in organisms long after a change in conditions (Baulcombe and Dean 2014). Epigenetic regulation

is the chemical modification of DNA and histone proteins with resultant changes in gene expression (Kumar and Mohapatra 2021). Importantly, epigenetic changes, which often result from biotic (Zhang *et al.* 2013) and abiotic (Hewezi *et al.* 2017) stimuli, can be relatively stable through a plant's lifecycle and even be heritable (Chang *et al.* 2020). This leaves open the possibility that, any observed gene expression similarities between the two Mount Hutt *R. monroi* taxa might be due to, and not responsible for, occupation of comparable environmental niches. Nevertheless, regardless of mechanism, the lack of GO term enrichment means no physiological inferences could be made based on pairwise comparisons with DESeq2.

4.4.2 Phylogenetic analyses of reference assembly expression levels

More stringent assessment of gene expression gives similar, although somewhat less emphatic, results to standard pairwise comparisons. Rather than relying on statistical inference of differential expression of genes, explicit testing of gene expression similarities among taxa was also undertaken. This is because arbitrary grouping of taxa for analyses by DESeq2 is observed to result in a gene being called differentially expressed when only one taxon displays a different expression profile. Although, as noted previously, ranking of gene expression levels to infer phylogenetic signals remains susceptible to biased read alignment. Ranking each gene according to expression level and then assessing taxon positions in the ranking is a stricter method than making pairwise comparisons with DESeq2 because one anomalous sample can cause unclear groupings. This strictness reflects the lower total DEGs detected using this method. Surprisingly, considering the failure of GO term enrichment analyses for the pairwise comparisons, six GO terms were enriched for the Mount Hutt *R. monroi–R. crithmifolius* split. However, no physiological inference beyond an enrichment in genes involved in protein synthesis could be made from these results. A weakness, due to the use of a single reference, common to both the standard pairwise comparisons and the phylogenetic analyses of expression levels, is the inability to infer physiological differences arising from gene variants of interest. Even when variation in the protein-coding region of a gene does not give rise to expression differences, important regulatory elements might be in linkage with the protein-coding region (Hubner *et al.* 2005; GuhaThakurta *et al.* 2006). To investigate the overlap between gene expression and allelic variation, analyses of gene sequence data from the taxa of interest is required.

4.4.3 Phylogenetic analyses of aggregated assembly SCO expression levels

Phylogenetic analyses of SCO expression levels, once again, supports a greater level of shared gene expression between *R. crithmifolius* and Mount Hutt *R. monroi* than between a grouping of *R. lobulatus* and Mount Hutt *R. monroi*. Surprisingly, hierarchical clustering of normalised gene expression levels results in *R. crithmifolius* and Mount Hutt *R. monroi* clustering together. Unfortunately, the filtering imposed by OrthoFinder makes inferring physiological differences between the taxa difficult because of the relatively small number of orthologue sets retained for analyses. These results indicate that using SCOs for analyses, reduces resolution of data. In particular, expression data for any genes other than SCOs is ignored. Furthermore, each filtering step, while increasing confidence in the data, reduces the power to detect differences between the taxa. The strict ranking of expression levels means many data are filtered out because gene expression in one anomalous sample can cause the entire orthogroup to be discarded. This means few clear gene groupings can be distinguished. Consequently, this results in reduced statistical power for GO term enrichment analyses.

4.4.4 Phylogenetic analyses of aggregated assembly SCO nucleotide variation

Aggregated assembly gene trees most often group the two *R. monroi* together, but the alternative groupings occur with approximately equal frequency. Constructing individual phylogenetic trees for each gene allows for species tree inference (Degnan *et al.* 2009). The gene trees also provide a means for identifying genes displaying relationships anomalous with respect to the species tree, which can imply introgression or incomplete lineage sorting in the evolutionary history of the species (Joly *et al.* 2009b; Suarez-Gonzalez *et al.* 2018). For closely-related taxa, phylogenetic reconstruction using nucleotide sequences provides greater resolution than the use of amino acid sequences (Simmons *et al.* 2002). In the present study, there is less than six percent difference in the number of gene trees grouping Mount Hutt *R. monroi* with *R. lobulatus*, compared with the Mount Hutt *R. monroi–R. crithmifolius* split. Notwithstanding the clustering of the two *R. monroi* taxa, this difference does not provide sufficient support to make inferences with respect to relationships between taxa. Additionally, no GO term enrichment was found for any of the three clusters.

Genes found to be common to both the SCO expression-level groupings and the SCO nucleotide variation groupings do not immediately explain the co-occurrence of *R. monroi* and *R. crithmifolius* at Mount Hutt. Aside from a single hypothetical protein belonging to the Archaean organism, *Saccharolobus solfataricus* P2, the top BLASTp hits for the twelve overlapping genes of this grouping were against proteins of species belonging to

Ranunculaceae. Mostly, the importance of these genes in niche specialisation appears unlikely or unclear. For example, one protein is uncharacterised while a second protein, inter alpha-trypsin inhibitor heavy chain, has no functional information available because it is part of a gene family relatively well-studied in humans (Hamm *et al.* 2008; Lord *et al.* 2020) and other mammals (Pineiro *et al.* 2004), but only predicted through sequence analyses in plants. Additionally, the hypothetical proteins IFM89_009976 (*Coptis chinensis*) and AQUCO_01700644v1 (*A. coerulea*) are part of larger protein complexes required for autophagy (Tang and Bassham 2018) and RNA polymerase II activity (Soutourina *et al.* 2011) respectively. No other proteins forming part of these complexes are found in this overlapping gene set. Two genes might potentially contribute to acclimation to conditions at Mount Hutt. The DNAJ-like protein (*Thalictrum thalictroides*) is likely to act as a co-chaperone with HSP70 to regulate cellular protein homeostasis (Cyr *et al.* 1994; Pulido and Leister 2018). DNAJ proteins are upregulated in response to abiotic stresses (Luo *et al.* 2019b), and overexpression of DNAJ proteins has been shown to confer resistance to biotic and abiotic stresses (Wang *et al.* 2014, 2019a). Furthermore, DNAJ loss-of-function mutants display increased stress sensitivity (Salas-Muñoz *et al.* 2016). Lastly, the AT-rich interactive domain-containing protein (*T. thalictroides*) is part of a class of proteins that interact with DNA (Roy *et al.* 2016) and are reported to mediate gene expression through changes in chromatin topology (Euskirchen *et al.* 2012; Xu *et al.* 2015).

A greater number of gene trees, incongruent with respect to phylogenetic expectations, are evident in the aggregated assembly nucleotide analyses than expected. If not an artifact of incorrect transcriptome assembly or imprecise orthologue detection, then this result (based on the spatial distribution of populations) is consistent with incomplete lineage sorting being important in the dispersal of alleles. However, the quantity of gene trees discarded because of low bootstrap values suggests conflicting signals in the data (Kennedy *et al.* 2005). While low bootstrap values can be explained by few informative sites present in a MSA (Berry and Gascuel 1996), these might also be symptomatic of chimeric transcript construction (Yang and Smith 2013). Overall, lack of confidence in the aggregated transcriptome assemblies means phylogenetic inferences based on allelic variation must be treated with caution.

4.4.5 Phylogenetic analyses of individual assembly SCO nucleotide variation

Phylogenetic analyses of the individual assemblies, again, clusters the two *R. monroi* taxa together but noise in the data renders further inferences concerning relationships impossible. Certainly, no GO term enrichment was found for any of the taxa clusters. However, making Trinity assemblies for each sample does provide an avenue for inferring if the large

number of incongruent gene trees constructed from the aggregated assembly dataset are biologically feasible—a not uncommon result for taxa with recent divergence times (Cranston *et al.* 2009)—or merely artifacts reflecting difficulties in reconstruction of transcriptomes for polyploid organisms. Interestingly, the taxonomic signal grouping the two *R. monroi* becomes stronger when clustering is based on individual assembly nucleotide variation, compared with aggregated assembly nucleotide variation. Yet, few bifurcating trees resolved taxonomic relationships with a central edge that had high bootstrap support. And, even among trees which did, the two populations of plants on either side of this split were often intermixed. This observation is explained by incompatible site patterns in the data. Even if the transcriptomes are correctly assembled, Trinity collapses heterozygous SNPs and outputs a chimeric isoform sequence. Retained SNPs are those with higher read support (B. Haas, personal communication, April 20, 2021). Therefore, for a given gene, a combination of heterozygous genes and codominant expression of alleles might make Trinity struggle to reconstruct a biologically accurate transcript.

4.4.6 Analyses of heterozygosity in the individual assemblies

The finding of a substantial level of SNP variation, similar in frequency in the four different taxa, was surprising. At the outset of this project it was assumed that many allelic variants would be fixed in the different taxa given the small sizes and geographic disjunction of their populations (Turner *et al.* 1982; Ellstrand and Elam 1993; Yang *et al.* 2018b). Proper characterisation of genome-wide heterozygosity and SNP variation is a challenging undertaking (Schmidt *et al.* 2021). However, the seemingly appreciable levels of SNP variation (Foster *et al.* 2010; Telfer *et al.* 2018)—particularly given the strict thresholds set for variant calling—detected within individual samples, raises the possibility that multiple alleles occur for many of the assembled genes. If so, this could make interpretation of the tree building analyses based on nucleotide variation problematic. Depending on the expression levels of different alleles at the time of sampling, in biological replicates, different alleles might be represented in the SCO dataset. While this finding identifies an unrecognised complexity of the data, it raises the intriguing question of whether maintenance of high levels of allelic variation might provide a means for rapid plastic and adaptive (Zhang *et al.* 2013; Lai *et al.* 2019) responses of these plants which live in extreme and fluctuating environments (Fisher 1965).

Alternatively, the nucleotide variation detected in the individual assemblies might be due to misassembly of short-read data. Yang and Smith (2013) have outlined the prevalence of chimeric transcripts arising from misassemblies by a suite of *de novo* transcriptome assembly software. In polyploid plants, the problem is even more pronounced. Despite the reported

success of Trinity (and other *de novo* assemblers) in the reconstruction of polyploid transcriptomes (Ilut *et al.* 2012; Chopra *et al.* 2014; Payá-Milans *et al.* 2018; Wang *et al.* 2019b; Zhou *et al.* 2019), the presence of ohnologous or homeologous genes is thought to cause assembly difficulties through increased transcriptome complexity (Li *et al.* 2019a; Madritsch *et al.* 2021), particularly in the case of relatively newly-formed polyploids (Chen *et al.* 2019). The TransDecoder results in this study might reflect this scenario. Slightly less than one third (29%) of representative contigs for all assemblies contain predicted protein sequences. Potentially, misassembly of contigs has introduced 'mutations', such as stop codons, which interfere with TransDecoder prediction of likely coding sequences.

Polyploid species often display increased transcriptome complexity. Highly self-fertile, but also outcrossing via wind-mediated and insect-mediated pollen transfer (Klein *et al.* 2003), the allotetraploid *Coffea arabica* (Rubiaceae) putatively originates from a single allopolyploidy event (Scalabrin *et al.* 2020) that occurred within the last 1.1 million years (Bawin *et al.* 2020). This single event caused a genetic bottleneck which means *C. arabica* has little nucleotide diversity and high levels of homozygosity (Scalabrin *et al.* 2020). Given current knowledge of the life and evolutionary histories of the alpine *Ranunculus*, it might be assumed a similar situation would exist in this group. Homozygosity, gene silencing (Edger *et al.* 2019), and gene removal (Thomas *et al.* 2006) might have acted to reduce transcriptome complexity. However, while extensive genome fractionation has occurred in plant species evolved subsequent to polyploidy events occurring > 10 million years ago (Bowers *et al.* 2003; Cai *et al.* 2021), and rapid gene loss has been reported in neopolyploid *Tragopogon miscellus* (Asteraceae) (Tate *et al.* 2009), high rates of homeologue retention and heterozygosity are commonly reported in polyploid systems. Originating approximately 300 years ago, from the hybridisation of two octaploid progenitors which each formed within the last 2 million years (Njuguna *et al.* 2013), the octaploid cultivated garden strawberry displays high levels of heterozygosity and retains all homeologous chromosomes (Edger *et al.* 2019). Additionally, smaller genomic and transcriptomic alterations may occur in allopolyploid species with disomic inheritance (bivalent chromosome pairing) patterns (Kryvokhyzha *et al.* 2019).

4.4.7 Summary of gene expression and phylogenetic analyses

Gene expression analyses indicate a level of shared expression between *R. crithmifolius* and Mount Hutt *R. monroi* anomalous with respect to expectations based on taxonomic relationships. However, physiological responses potentially important for habitat acclimation are not yet understood, and nucleotide variation provides no phylogenetic inferences beyond the currently accepted alpine *Ranunculus* taxonomy. Regardless of how gene expression data are

analysed, more differentially expressed genes cluster Mount Hutt *R. monroi* with *R. crithmifolius* than with *R. lobulatus* (Figure 3.15). But, the genetic underpinnings of this are undetermined. As expected, allele sharing appears greatest between the two *R. monroi*, but there is no signal in the data to separate the other two taxonomic splits (Figure 3.15). Potential issues with the transcriptome assemblies, arising from excess heterozygosity and/or the presence of many homologous genes reduced the number of genes that could be functionally annotated, and make taxonomic findings less robust. Without robust molecular markers that can be comprehensively tested, it is not possible to infer if the shared expression between *R. crithmifolius* and Mount Hutt *R. monroi* results from epigenetic regulation, incomplete lineage sorting, or introgression.

4.5 Research limitations

4.5.1 Tissue selection

In this study, gene expression within mature leaf tissue is investigated. Yet, other tissue might be informative for understanding physiologies important for the ecological requirements of these taxa. Plant distributions can be heavily influenced by seed germination (Yang *et al.* 2017a) and seedling survival (Johnson *et al.* 2017). Furthermore, owing to the lack of new root tissue from the *R. crithmifolius* plants grown, physiological responses in the roots remain unexamined. The importance of this is unclear because environmental stimuli in one tissue type often triggers gene expression changes in distal tissue via signaling mechanisms (Li *et al.* 2021). And interestingly, studies of drought stress (Iovieno *et al.* 2016; Zhang *et al.* 2016; Ma *et al.* 2017; Liu *et al.* 2018; Ye *et al.* 2018; Gonçalves *et al.* 2019) and other soil factors (Peng *et al.* 2018; Zhang *et al.* 2020b) commonly survey gene expression in solely leaf tissue. However, soil factors such as water deficit can cause changes in root morphology (Kadam *et al.* 2015; Bloch *et al.* 2019), and Corso *et al*. (2015) found differential expression of approximately twice as many genes in root tissue compared with shoot tissue of two drought stressed grapevine (*Vitis*) (Vitaceae) rootstock cultivars. Yet, in contrast to the findings of Corso *et al.* (2015), Khadka *et al*. (2019) discovered a five-fold increase in abundance of DEGs within shoots of water-stressed *Vitis riparia Michx* compared with roots. Of these DEGs, only 836 were common to both tissue types, which illustrates how gene expression profiles under water stress differed between the tissue types. Gene expression responses to soil chemicals can also be more pronounced in either above-ground or below-ground tissue. Nitrogen stress induced a greater number of DEGs in shoots than in root tissue of potato (Tiwari *et al.* 2020), but the converse situation was observed

for Indian mustard (*Brassica juncea*) (Brassicaceae) responding to cadmium (Thakur *et al.* 2019).

4.5.2 Artificial growth environment

By growing plants under controlled conditions, noise in the data is minimised, but interesting physiological responses might not be detected. Under the growth parameters used in this study (plastic responses to the artificial growth conditions notwithstanding), only constitutive expression is investigated, and responses to environmental stimuli are unexamined.

Phenotypic plasticity has an important adaptive function in allowing long-lived species to tolerate rapid shifts in environmental conditions (Hoffmann *et al.* 2015). An impressive example of plasticity is seen within *Ranunculus*. In the semi-aquatic Korean species *Ranunculus trichophyllus* var. *kadzusensis*, leaf morphology is decisively altered in submerged leaves. This heterophylly and the underlying abscisic acid (ABA) and ethylene response, is not observed in the closely-related *R. sceleratus* (Kim *et al.* 2018). More importantly, for the New Zealand alpine *Ranunculus*, extinction is hypothesised to be less likely for species of restricted geographic range and limited dispersal capability, during environmental shifts, when a plastic responses allow them to adapt to local conditions (Chevin *et al.* 2010).

Unfortunately, conducting comparative gene expression analyses with field samples is not trivial. Expression variance increases in field samples, so a greater number of biological replicates are needed for equivalent statistical power (Todd *et al.* 2016). Controlling for confounding short-term variables is also problematic for field studies in this system. The spatial separation of study sites means weather conditions can vary appreciably. Samples can be collected at similar times of day to control for changes in circadian rhythm (Edger *et al.* 2017), but environmental factors are nearly impossible to account for without conducting a common garden or reciprocal transplant experiment (de Villemereuil *et al.* 2016). Although it is hypothesised that the population of *R. monroi* at Mount Hutt has greater fitness in this environment than Porters Pass *R. monroi* would have, explicit testing of this assumption requires a reciprocal transplant experiment to be conducted (de Villemereuil *et al.* 2016). Unfortunately, conducting such an experiment is logistically challenging, and regulatory approval is likely to be difficult to obtain.

Chapter 5: Conclusions

5.1 Habitat differences

Environmental differences between the habitats of the two *R. monroi* taxa are clear. In addition, measured and observed variables indicate the habitat being occupied by Mount Hutt *R. monroi* is more similar to *R. crithmifolius* habitat than it is to the typically-described *R. monroi* habitat that is evident at Porters Pass.

5.2 DNA barcoding

DNA molecular markers provided interesting results but were ultimately uninformative. The ITS2 data alone do not strongly support the partitioning of the *R. insignis* forms, investigated here, into separate species. The *ycf1* data only tentatively support a hypothesis of ancestral introgression from *R. crithmifolius* into *R. monroi* at Mount Hutt, and further work using chloroplast markers is needed to improve resolution of the taxonomic relationships.

5.3 Common garden experiment

Design of a common garden experiment allows for inferences of gene expression free from noise induced by environmental variability. However, this study illustrates the difficulty in establishing a set of conditions conducive to the active growth of multiple taxa from different habitats. Additionally, gene expression patterns responding to important environmental stimuli cannot be examined when environmental stimuli are fixed. Regardless of the above considerations, phylogenetic signals contained within the expression data could be detected and analysed when using the methodology adopted in this research.

5.4 Gene expression analyses

5.4.1 Read alignment biases must be accounted for

The biased read alignment detected in this study highlights issues that must be addressed when making interspecific gene expression comparisons. As noted by Voelckel *et al*. (2012, 2017), taxonomic relationships influence the efficiency with which RNA-seq reads align to a reference. Additionally, the present study indicates the set of RNA-seq reads used in construction of a *de novo* transcriptome also influence read alignment. This suggests reads used in the construction of an assembly should not be used for quantification purposes. Or alternatively, all reads to be used for quantification should be utilised in the transcriptome construction. Mitigating biased read mapping by quantification of self-mapped SCOs also has drawbacks. Expression data for all other genes is lost. This means some genes with important ecological functions are likely to remain unexamined.

5.4.2 Gene expression data can reflect evolutionary relationships

Concerns raised over evolutionary relationships confounding comparative gene expression analyses (Rohlfs and Nielsen 2015; Dunn *et al.* 2018) are shown to be well-founded in this research. The taxonomic signals contained within this study's gene expression data meant all methods of DGE analyses grouped the two *R. monroi* together at a greater frequency than the alternative splits. Consistent with the hypothesis of introgression of genetic material from *R. crithmifolius* to the *R. monroi* found at Mount Hutt, DGE analyses also grouped these taxa together more frequently than was the case for the Mount Hutt *R. monroi–R. lobulatus* split. Unfortunately, the data gathered are also consistent with competing hypotheses such as incomplete lineage sorting or epigenetic regulation. This means the evolutionary mechanism underlying the habitat tolerance of *R. monroi* at Mount Hutt remains unsolved.

5.4.3 Transcriptome assembly issues with polyploid species might be masked

Complexity within the RNA-seq data make it difficult to have confidence in the *de novo* assemblies. Interestingly, validation methods, such as BUSCO recovery rates, suggest the transcriptome assemblies are of high quality. However, discordant gene trees with intermixing of samples, and low numbers of predicted proteins in the assemblies indicate otherwise. These

results indicate issues arising from the transcriptome complexity of many polyploid species are not easy to detect using typical *de novo* transcriptome validation methods.

5.4.4 The alpine *Ranunculus* appear to contain substantial genetic resources

While a lack of confidence in the *de novo* assemblies makes phylogenetic inferences difficult, the underlying issues contributing to this problem might be beneficial for the alpine *Ranunculus*. Standing genetic variation is an important component in the evolutionary potential of an organism (Barrett and Schluter 2008; Lai *et al.* 2019), as is the ability to tolerate rapid environmental shifts through phenotypic plasticity (Snell-Rood *et al.* 2018; Fox *et al.* 2019). The difficulties encountered in this study suggest the plants assayed are actively expressing numbers of homologous genes and different alleles. While this is problematic for transcriptome assembly, it is indicative of rich allelic diversity, as well as retained function of homeologues (or ohnologues) which can be epigenetically regulated for swift plastic changes (Zhang *et al.* 2013).

5.5 Future work

Concern over *de novo* assembly of polyploid plant transcriptomes makes high-quality transcriptomes necessary for robust inferences of heterozygosity and allelic variation between taxa. The advent of long-read RNA sequencing technologies alleviates the potential for chimeric transcripts being reconstructed from short-read data (Hoang *et al.* 2017), and correction (Nielsen *et al.* 2019; Hu *et al.* 2020) or hybrid assembly (Puglia *et al.* 2020) using short reads can be conducted to reduce the relatively high error rates associated with both PacBio and ONT sequencing. Constructing high-quality reference transcriptomes will likely improve functional annotation rates and therefore provide greater statistical power to detect physiological differences. Moreover, utilising a reference species of equal evolutionary distance from the study taxa (i.e. a species from Fisher's many-petalled line) means non-biased alignment of reads from the four study taxa to this reference will allow validation of heterozygosity inference, and robust gene expression analyses to be carried out. In addition to investigations based on RNA sequencing, genome skimming followed by chloroplast assembly can resolve the incongruities in the chloroplast haplotype data (Liu *et al.* 2020; Wang *et al.* 2020) and be applied to a wider dataset.

5.6 Concluding statement

The current work illustrates both the potential and the pitfalls of using RNA sequencing to understand polyploid plants with unclear ancestry. Gene expression data can be successfully used to investigate evolutionary relationships, but careful planning is required to avoid introducing biases and to generate sufficient data for statistical analyses. RNA-seq data can also be used to construct phylogenies based on nucleotide variation. However, at least in a polyploid system, biological complexity makes generating robust molecular markers a challenging proposition.

At the outset of this study, it was assumed that responses to climate change by individual alpine *Ranunculus* species might be evolutionarily constrained by low genetic variation. And that, for this reason, introgression may be a necessary mechanism for increasing evolutionary potential through rapid changes to standing genetic variation. However, while further research is needed to establish the relative contribution of introgression to an apparent niche shift by *R. monroi*, this study suggests the polyploid genomes of the alpine *Ranunculus* might contain a genetic richness previously unanticipated.

Bibliography

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, *et al.*, 2013 Hybridization and speciation. Journal of Evolutionary Biology 26: 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Abrams, Z. B., T. S. Johnson, K. Huang, P. R. O. Payne, and K. Coombes, 2019 A protocol to evaluate RNA sequencing normalization methods. BMC Bioinformatics 20: 679. https://doi.org/10.1186/s12859-019-3247-x

Acín-Carrera, M., M. José Marques, P. Carral, A. M. Álvarez, C. López, *et al.*, 2013 Impacts of land-use intensity on soil organic carbon content, soil structure and water-holding capacity. Soil Use and Management 29: 547–556. https://doi.org/10.1111/sum.12064

Adler, D., and S. T. Kelly, 2019 *vioplot: violin plot*

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990 Basic local alignment search tool. Journal of Molecular Biology 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amaral, D. T., I. A. S. Bonatelli, R. Cerri, and V. R. Viviani, 2019 Phylogenomic analyses and divergence time estimation of Elateroidea (Coleoptera) based on RNA-Seq data. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 30: 283–289. https://doi.org/10.1016/j.cbd.2019.04.001

Anderegg, L. D. L., and J. Hille Ris Lambers, 2016 Drought stress limits the geographic ranges of two tree species via different physiological mechanisms. Global Change Biology 22: 1029–1045. https://doi.org/10.1111/gcb.13148

Anderson, E., and L. Hubricht, 1938 Hybridization in Tradescantia. III. The evidence for introgressive hybridization. American Journal of Botany 25: 396–402. https://doi.org/10.1002/j.1537-2197.1938.tb09237.x

Anderson, E., 1953 Introgressive Hybridisation. Biological Reviews 28: 280–307. https://doi.org/10.1111/j.1469-185X.1953.tb01379.x

Andrews, S., 2010 *FastQC: a quality control tool for high throughput sequence data*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arkley, R., 1963 Relationships between plant growth and transpiration. Hilgardia 34: 559–584. https://doi.org/DOI:10.3733/hilg.v34n13p559

Ashraf, M., and P. J. C. Harris, 2013 Photosynthesis under stressful environments: an overview. Photosynthetica 51: 163–190. https://doi.org/10.1007/s11099-013-0021-6

Baduel, P., S. Bray, M. Vallejo-Marin, F. Kolář, and L. Yant, 2018 The "Polyploid Hop": Shifting Challenges and Opportunities Over the Evolutionary Lifespan of Genome Duplications. Frontiers in Ecology and Evolution 6: 117. https://doi.org/10.3389/fevo.2018.00117

Baniaga, A. E., H. E. Marx, N. Arrigo, and M. S. Barker, 2020 Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. Ecology Letters 23: 68–78. https://doi.org/10.1111/ele.13402

Barbagallo, R. P., K. Oxborough, K. E. Pallett, and N. R. Baker, 2003 Rapid, noninvasive screening for perturbations of metabolism and plant growth using chlorophyll fluorescence imaging. Plant Physiology 132: 485–493. https://doi.org/10.1104/pp.102.018093

Barnes, P. W., R. J. Ryel, and S. D. Flint, 2017 UV Screening in Native and Non-Native Plant Species in the Tropical Alpine: Implications for Climate Change-Driven Migration of Species to Higher Elevations. Frontiers in Plant Science 8: 1451. https://doi.org/10.3389/fpls.2017.01451

Barrett, R. D. H., and D. Schluter, 2008 Adaptation from standing genetic variation. Trends in Ecology and Evolution 23: 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Baulcombe, D. C., and C. Dean, 2014 Epigenetic Regulation in Plant Responses to the Environment. Cold Spring Harbor Perspectives in Biology 6: a019471. https://doi.org/10.1101/cshperspect.a019471

Bawin, Y., T. Ruttink, A. Staelens, A. Haegeman, P. Stoffelen, *et al.*, 2020 Phylogenomic analysis clarifies the evolutionary origin of *Coffea arabica*. Journal of Systematics and Evolution. https://doi.org/10.1111/jse.12694

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. https://doi.org/10.1038/nclimate2027

Becker, M., 2020 *Drought resistance in alpine Ranunculus and extinction risk from climate change*. Manuscript in preparation.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57: 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Y., and D. Yekutieli, 2001 The Control of the False Discovery Rate in Multiple Testing Under Dependency. The Annals of Statistics 29: 1165–1188.

Berry, V., and O. Gascuel, 1996 On the Interpretation of Bootstrap Trees: Appropriate Threshold of Clade Selection and Induced Gain. Molecular Biology and Evolution 13: 999–1011. https://doi.org/10.1093/molbev/13.7.999

Bhardwaj, A. R., G. Joshi, B. Kukreja, V. Malik, P. Arora, *et al.*, 2015 Global insights into high temperature and drought stress regulated genes by RNA-Seq in economically important oilseed crop *Brassica juncea*. BMC Plant Biology 15: 9. https://doi.org/10.1186/s12870-014-0405-1

Bilandžija, H., B. Hollifield, M. Steck, G. Meng, M. Ng, *et al.*, 2020 Phenotypic plasticity as a mechanism of cave colonization and adaptation. Elife 9: e51830. https://doi.org/10.7554/eLife.51830

Björkman, O., and B. Demmig, 1987 Photon yield of O_2 evolution and chlorophyll fluorescence characteristics at 77 K among vascular plants of diverse origins. Planta 170: 489–504. https://doi.org/10.1007/BF00402983

Blanc, G., and K. H. Wolfe, 2004 Widespread Paleopolyploidy in Model Plant Species Inferred from Age Distributions of Duplicate Genes. The Plant Cell 16: 1667–1678. https://doi.org/10.1105/tpc.021345

Bloch, D., M. R. Puli, A. Mosquna, and S. Yalovsky, 2019 Abiotic stress modulates root patterning via ABA-regulated microRNA expression in the endodermis initials. Development 146. https://doi.org/10.1242/dev.177097

Blumthaler, M., W. Ambach, and R. Ellinger, 1997 Increase in solar UV radiation with altitude. Journal of Photochemistry and Photobiology B: Biology 39: 130–134. https://doi.org/10.1016/S1011-1344(96)00018-8

Bona, A., U. Kulesza, and K. A. Jadwiszczak, 2019 Clonal diversity, gene flow and seed production in endangered populations of *Betula humilis* Schrk. Tree Genetics & Genomes 15: 50. https://doi.org/10.1007/s11295-019-1357-2

Booker, T. R., B. C. Jackson, and P. D. Keightley, 2017 Detecting positive selection in the genome. BMC Biology 15: 98. https://doi.org/10.1186/s12915-017-0434-y

Bourke, P. M., R. E. Voorrips, R. G. F. Visser, and C. Maliepaard, 2018 Tools for Genetic Studies in Experimental Populations of Polyploids. Frontiers in Plant Science 9: 513. https://doi.org/10.3389/fpls.2018.00513

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. https://doi.org/10.1038/nature01521

Bozdogan, H., 1987 Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. Psychometrika 52: 345–370. https://doi.org/10.1007/BF02294361

Bryant, D., and V. Moulton, 2004 Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. Molecular Biology and Evolution 21: 255–265. https://doi.org/10.1093/molbev/msh018

Bushmanova, E., D. Antipov, A. Lapidus, and A. D. Prjibelski, 2019 rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. GigaScience 8: giz100. https://doi.org/10.1093/gigascience/giz100

Butcher, P. A., M. Byrne, and G. F. Moran, 1995 Variation within and among the chloroplast genomes of *Melaleuca alternifolia* and *M. linariifolia* (Myrtaceae). Plant Systematics and Evolution 194: 69–81. https://doi.org/10.1007/BF00983217

Cai, X., L. Chang, T. Zhang, H. Chen, L. Zhang, *et al.*, 2021 Impacts of allopolyploidization and structural variation on intraspecific diversification in *Brassica rapa*. Genome Biology 22: 166. https://doi.org/10.1186/s13059-021-02383-2

Camarero, J. J., R. Sánchez-Salguero, G. Sangüesa-Barreda, and L. Matías, 2018 Tree species from contrasting hydrological niches show divergent growth and water-use efficiency. Dendrochronologia 52: 87–95. https://doi.org/10.1016/j.dendro.2018.10.003

Campa, A., and J. J. Ferreira, 2018 Genetic diversity assessed by genotyping by sequencing (GBS) and for phenological traits in blueberry cultivars. PLoS ONE 13: e0206361. https://doi.org/10.1371/journal.pone.0206361

Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón, 2009 trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Carter, R. J., 2006 *An investigation into the evolutionary relationships of the North Island alpine* Ranunculus*.* Massey University, Palmerston North, New Zealand.

Carter, K. A., A. Liston, N. V Bassil, L. A. Alice, J. M. Bushakra, *et al.*, 2019 Target Capture Sequencing Unravels *Rubus* Evolution. Frontiers in Plant Science 10: 1615. https://doi.org/10.3389/fpls.2019.01615

Caruso, B. S., R. King, S. Newton, and C. Zammit, 2017 Simulation of Climate Change Effects on Hydropower Operations in Mountain Headwater Lakes, New Zealand. River Research and Applications 33: 147–161. https://doi.org/10.1002/rra.3056

Chang, Y., C. Zhu, J. Jiang, H. Zhang, J. Zhu, *et al.*, 2020 Epigenetic regulation in plant abiotic stress responses. Journal of Integrative Plant Biology 62: 563–580. https://doi.org/10.1111/jipb.12901

Chen, H., W. Zhang, K. Wang, and W. Fu, 2010 Soil moisture dynamics under different land uses on karst hillslope in northwest Guangxi, China. Environmental Earth Sciences 61: 1105–1111. https://doi.org/10.1007/s12665-009-0428-3

Chen, L.-Y., D. F. Morales-Briones, C. N. Passow, and Y. Yang, 2019 Performance of gene expression analyses using *de novo* assembled transcripts in polyploid species. Bioinformatics 35: 4314–4320. https://doi.org/10.1093/bioinformatics/btz620

Cheng, F., J. Wu, L. Fang, S. Sun, B. Liu, *et al.*, 2012 Biased Gene Fractionation and Dominant Gene Expression Among the Subgenomes of *Brassica rapa*. PLoS ONE 7: e36442. https://doi.org/10.1371/journal.pone.0036442

Cheng, C., V. Krishnakumar, A. P. Chan, F. Thibaud-Nissen, S. Schobel, *et al.*, 2017 Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. The Plant Journal 89: 789–804. https://doi.org/10.1111/tpj.13415

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. https://doi.org/10.1038/s41477-018-0136-7

Chevin, L. M., R. Lande, and G. M. Mace, 2010 Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory. PLoS Biology 8: e1000357. https://doi.org/10.1371/journal.pbio.1000357

Choi, M., J. M. Jacobs, and M. H. Cosh, 2007 Scaled spatial variability of soil moisture fields. Geophysical Research Letters 34. https://doi.org/10.1029/2006GL028247

Choi, K. S., M. G. Chung, and S. Park, 2016 The Complete Chloroplast Genome Sequences of Three Veroniceae Species (Plantaginaceae): Comparative Analysis and Highly Divergent Regions. Frontiers in Plant Science 7: 355. https://doi.org/10.3389/fpls.2016.00355

Chopra, R., G. Burow, A. Farmer, J. Mudge, C. E. Simpson, *et al.*, 2014 Comparisons of *De Novo* Transcriptome Assemblers in Diploid and Polyploid Species Using Peanut (*Arachis* spp.) RNA-Seq Data. PLoS ONE 9: e115055. https://doi.org/10.1371/journal.pone.0115055

Cichota, R., V. O. Snow, and A. B. Tait, 2008 A functional evaluation of virtual climate station rainfall data. New Zealand Journal of Agricultural Research 51: 317–329. https://doi.org/10.1080/00288230809510463

Claros, M. G., R. Bautista, D. Guerrero-Fernández, H. Benzerki, P. Seoane, *et al.*, 2012 Why Assembling Plant Genome Sequences Is So Challenging. Biology 1: 439–459. https://doi.org/10.3390/biology1020439

Clevenger, J. P., W. Korani, P. Ozias-Akins, and S. Jackson, 2018 Haplotype-Based Genotyping in Polyploids. Frontiers in Plant Science 9: 564. https://doi.org/10.3389/fpls.2018.00564

Cockayne, L., 1910 *New Zealand plants and their story.* Government Print., Wellington, New Zealand.

Cockayne, L., 1923 Hybridism in the New Zealand Flora. New Phytologist 22: 105–127.

Cockayne, L., and H. H. Allan, 1934 An Annotated List of Groups of Wild Hybrids in the New Zealand Flora. Annals of Botany 48: 1–55. https://doi.org/10.1093/oxfordjournals.aob.a090429

Comai, L., 2005 The advantages and disadvantages of being polyploid. Nature Reviews Genetics 6: 836–846. https://doi.org/10.1038/nrg1711

Conesa, A., P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, *et al.*, 2016 A survey of best practices for RNA-seq data analysis. Genome Biology 17: 13. https://doi.org/10.1186/s13059-016-0881-8

Connor, H. E., 1967 Interspecific hybrids in *Chionochloa* (Gramineae). New Zealand Journal of Botany 5: 3–16. https://doi.org/10.1080/0028825X.1967.10428731

Connor, H. E., 1985 Biosystematics of higher plants in New Zealand 1965–1984. New Zealand Journal of Botany 23: 613–643. https://doi.org/10.1080/0028825X.1985.10434233

Corneillie, S., N. De Storme, R. Van Acker, J. U. Fangel, M. De Bruyne, *et al.*, 2019 Polyploidy Affects Plant Growth and Alters Cell Wall Composition. Plant Physiology 179: 74–87. https://doi.org/10.1104/pp.18.00967

Cornish-Bowden, A., 1985 Nomenclature for incompletely specified bases in nucleic acid sequences: rcommendations 1984. Nucleic Acids Research 13: 3021–3030. https://doi.org/10.1093/nar/13.9.3021

Corso, M., A. Vannozzi, E. Maza, N. Vitulo, F. Meggio, *et al.*, 2015 Comprehensive transcript profiling of two grapevine rootstock genotypes contrasting in drought susceptibility links the phenylpropanoid pathway to enhanced tolerance. Journal of Experimental Botany 66: 5739–5752. https://doi.org/10.1093/jxb/erv274

Costa-Silva, J., D. Domingues, and F. M. Lopes, 2017 RNA-Seq differential expression analysis: An extended review and a software tool. PLoS ONE 12: e0190152. https://doi.org/10.1371/journal.pone.0190152

Couvreur, T. L. P., A. J. Helmstetter, E. J. M. Koenen, K. Bethune, R. D. Brandão, *et al.*, 2019 Phylogenomics of the Major Tropical Plant Family Annonaceae Using Targeted Enrichment of Nuclear Genes. Frontiers in Plant Science 9: 1941. https://doi.org/10.3389/fpls.2018.01941

Craine, J. M., T. W. Ocheltree, J. B. Nippert, E. G. Towne, A. M. Skibbe, *et al.*, 2013 Global diversity of drought tolerance and grassland climate-change resilience. Nature Climate Change 3: 63–67. https://doi.org/10.1038/nclimate1634

Cranston, K. A., B. Hurwitz, D. Ware, L. Stein, and R. A. Wing, 2009 Species Trees from Highly Incongruent Gene Trees in Rice. Systematic Biology 58: 489–500. https://doi.org/10.1093/sysbio/syp054

Cyr, D. M., T. Langer, and M. G. Douglas, 1994 DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. Trends in Biochemical Sciences 19: 176–181. https://doi.org/10.1016/0968-0004(94)90281-x

da Fonseca, R. R., A. Albrechtsen, G. E. Themudo, J. Ramos-Madrigal, J. A. Sibbesen, *et al.*, 2016 Next-generation biology: Sequencing and data analysis approaches for non-model organisms. Marine Genomics 30: 3–13. https://doi.org/10.1016/j.margen.2016.04.012

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Davidson, N. M., and A. Oshlack, 2014 Corset: enabling differential gene expression analysis for *de novo* assembled transcriptomes. Genome Biology 15: 410. https://doi.org/10.1186/s13059-014-0410-6

de Lange, P. J., J. R. Rolfe, J. W. Barkla, S. Courtney, P. D. Champion, *et al.*, 2018 *Conservation status of New Zealand indigenous vascular plants, 2017*. Department of Conservation, Wellington, New Zealand.

De, Villemereuil P., O. E. Gaggiotti, M. Mouterde, and I. Till-Bottraud, 2016 Common garden experiments in the genomic era: new perspectives and opportunities. Heredity 116: 249–254. https://doi.org/10.1038/hdy.2015.93

Degnan, J. H., M. DeGiorgio, D. Bryant, and N. A. Rosenberg, 2009 Properties of Consensus Methods for Inferring Species Trees from Gene Trees. Systematic Biology 58: 35–54. https://doi.org/10.1093/sysbio/syp008

Derbyshire, M. C., 2020 Bioinformatic Detection of Positive Selection Pressure in Plant Pathogens: the Neutral Theory of Molecular Sequence Evolution in Action. Frontiers in Microbiology 11: 644. https://doi.org/10.3389/fmicb.2020.00644

Soil Science Division Staff, 2017. *Soil survey manual*. C. Ditzler, K. Scheffe, and H.C. Monger (eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.

Dong, W., C. Xu, C. Li, J. Sun, Y. Zuo, *et al.*, 2015 *ycf1*, the most promising plastid DNA barcode of land plants. Scientific Reports 5: 8348. https://doi.org/10.1038/srep08348

Dorken, M. E., and C. G. Eckert, 2001 Severely reduced sexual reproduction in northern populations of a clonal plant, *Decodonverticillatus* (Lythraceae). Journal of Ecology 89: 339–350. https://doi.org/10.1046/j.1365-2745.2001.00558.x

Dunn, C. W., F. Zapata, C. Munro, S. Siebert, and A. Hejnol, 2018 Pairwise comparisons across species are problematic when analyzing functional genomic data. Proceedings of the National Academy of Sciences 115: E409–E417. https://doi.org/10.1073/pnas.1707515115

East, E. M., 1908 Inbreeding in corn. Report of the Connecticut Agricultural Experiment Station 1907: 419–428.

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792–1797. https://doi.org/10.1093/nar/gkh340

Edger, P. P., R. Smith, M. R. McKain, A. M. Cooley, M. Vallejo-Marin, *et al.*, 2017 Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. The Plant Cell 29: 2150–2167. https://doi.org/10.1105/tpc.17.00010

Edger, P. P., T. J. Poorten, R. VanBuren, M. A. Hardigan, M. Colle, *et al.*, 2019 Origin and evolution of the octoploid strawberry genome. Nature Genetics 51: 541–547. https://doi.org/10.1038/s41588-019-0356-4

Elder, Jr J. F., and B. J. Turner, 1995 Concerted evolution of repetitive DNA sequences in eukaryotes. The Quarterly Review of Biology 70: 297–320. https://doi.org/10.1086/419073

Ellstrand, N. C., and D. R. Elam, 1993 Population Genetic Consequences of Small Population Size: Implications for Plant Conservation. Annual Review of Ecology and Systematics 24: 217–242. https://doi.org/10.1146/annurev.es.24.110193.001245

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, *et al.*, 2011 A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6: e19379. https://doi.org/10.1371/journal.pone.0019379

Emadzade, K., C. Lehnebach, P. J. Lockhart, E. Hörandl, S. Taxon, *et al.*, 2010 A molecular phylogeny , morphology and classification of genera of Ranunculeae (Ranunculaceae). Taxon 59: 809–828. https://doi.org/10.1111/j.1365-2699.2010.02404.x

Emms, D. M., and S. Kelly, 2018 STAG: Species Tree Inference from All Genes. bioRxiv 267914. https://doi.org/10.1101/267914 (Preprint posted February 19, 2018).

Emms, D. M., and S. Kelly, 2019 OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology 20: 238. https://doi.org/10.1186/s13059-019-1832-y

Euskirchen, G., R. K. Auerbach, and M. Snyder, 2012 SWI/SNF chromatin-remodeling factors: multiscale analyses and diverse functions. Journal of Biological Chemistry 287: 30897–30905. https://doi.org/10.1074/jbc.R111.309302

Felsenstein, J., n.d. *The Newick tree format.* Available at https://evolution.genetics.washington.edu/phylip/newicktree.html (Accessed: 11 April 2021)

Ferreira de Carvalho, J., J. Lucas, G. Deniot, C. Falentin, O. Filangi, *et al.*, 2019 Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in *Brassica*. The Plant Journal 98: 434–447. https://doi.org/10.1111/tpj.14228

Ferris, K. G., and J. H. Willis, 2018 Differential adaptation to a harsh granite outcrop habitat between sympatric *Mimulus* species. Evolution 72: 1225–1241. https://doi.org/10.1111/evo.13476

Finotello, F., and B. Di Camillo, 2015 Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. Briefings in Functional Genomics 14: 130–142. https://doi.org/10.1093/bfgp/elu035

Fisher, F. J. F., 1952 Observations on the vegetation of screes in Canterbury, New Zealand. The Journal of Ecology 40: 156–167.

Fisher, F. J. F., 1965 *The alpine Ranunculi of New Zealand*. Botany Division, Department of Scientific and Industrial Research, Wellington, New Zealand.

Fisher, R. A., 1992 Statistical methods for research workers, pp. 66–70 in *Breakthroughs in statistics*, edited by S. Kotz and N.L. Johnson. Springer, New York.

Fitch, W. M., 1970 Distinguishing Homologous from Analogous Proteins. Systematic Zoology 19: 99–113. https://doi.org/10.2307/2412448

Foster, J. T., G. J. Allan, A. P. Chan, P. D. Rabinowicz, J. Ravel, *et al.*, 2010 Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis*). BMC Plant Biology 10: 13. https://doi.org/10.1186/1471-2229-10-13

Fox, R. J., J. M. Donelson, C. Schunter, T. Ravasi, and J. D. Gaitán-Espitia, 2019 Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change. Philosophical Transactions of the Royal Society B 374: 20180174. https://doi.org/10.1098/rstb.2018.0174

Fracasso, A., L. M. Trindade, and S. Amaducci, 2016 Drought stress tolerance strategies revealed by RNA-Seq in two sorghum genotypes with contrasting WUE. BMC Plant Biology 16: 115. https://doi.org/10.1186/s12870-016-0800-x

Franklin, J., F. W. Davis, M. Ikegami, A. D. Syphard, L. E. Flint, *et al.*, 2013 Modeling plant species distributions under future climates: How fine scale do climate projections need to be? Global Change Biology 19: 473–483. https://doi.org/10.1111/gcb.12051

Freedman, A., 2016 *Best Practices for De Novo Transcriptome Assembly with Trinity.* Available at: https://informatics.fas.harvard.edu/best-practices-for-de-novo-transcriptome-assembly-with-trinity.html (Accessed: 23 January 2021)

Gabaldón, T., and E. V Koonin, 2013 Functional and evolutionary implications of gene orthology. Nature Reviews Genetics 14: 360–366. https://doi.org/10.1038/nrg3456

Ganley, A. R. D., and T. Kobayashi, 2007 Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. Genome Research 17: 184–191. https://doi.org/10.1101/gr.5457707

Gao, S. B., L. D. Mo, L. H. Zhang, J. L. Zhang, J. B. Wu, *et al.*, 2018 Phenotypic plasticity vs. local adaptation in quantitative traits differences of *Stipa grandis* in semi-arid steppe, China. Scientific Reports 8: 3148. https://doi.org/10.1038/s41598-018-21557-w

Gene Ontology Consortium, 2004 The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research 32: D258–D261. https://doi.org/10.1093/nar/gkh036

Gene Ontology Consortium, 2019 The gene ontology resource: 20 years and still GOing strong. Nucleic Acids Research 47: D330–D338. https://doi.org/10.1093/nar/gky1055.

Glover, N. M., H. Redestig, and C. Dessimoz, 2016 Homoeologs: What Are They and How Do We Infer Them? Trends in Plant Science 21: 609–621. https://doi.org/10.1016/j.tplants.2016.02.005

Gonçalves, L. P., R. L. B. Camargo, M. A. Takita, M. A. Machado, W. S. dos Soares Filho, *et al.*, 2019 Rootstock-induced molecular responses associated with drought tolerance in sweet orange as revealed by RNA-Seq. BMC Genomics 20: 110. https://doi.org/10.1186/s12864-019-5481-z

Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes, *et al.*, 2012 Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research 40: D1178–D1186. https://doi.org/10.1093/nar/gkr944

Goodswen, S. J., P. J. Kennedy, and J. T. Ellis, 2018 A Gene-Based Positive Selection Detection Approach to Identify Vaccine Candidates Using *Toxoplasma gondii* as a Test Case Protozoan Pathogen. Frontiers in Genetics 9: 332. https://doi.org/10.3389/fgene.2018.00332

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29: 644–652. https://doi.org/10.1038/nbt.1883

Greenwood, R. M., and I. A. E. Atkinson, 1977 Evolution of Divaricating Plants in New Zealand in Relation to moa browsing. Proceedings of the New Zealand Ecological Society 24: 21–33.

GuhaThakurta, D., T. Xie, M. Anand, S. W. Edwards, G. Li, *et al.*, 2006 Cis-regulatory variations: A study of SNPs around genes showing cis-linkage in segregating mouse populations. BMC Genomics 7: 235. https://doi.org/10.1186/1471-2164-7-235

Haag, C. R., and D. Ebert, 2004 A new hypothesis to explain geographic parthenogenesis. Annales Zoologici Fennici 41: 539–544.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, *et al.*, 2013 *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8: 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hahn, M. W., 2009 Distinguishing Among Evolutionary Models for the Maintenance of Gene Duplicates. Journal of Heredity 100: 605–617. https://doi.org/10.1093/jhered/esp047

Halloy, S., and A. Mark, 2003 Climate-Change Effects on Alpine Plant Biodiversity: A New Zealand Perspective on Quantifying the Threat. Arctic, Antarctic, and Alpine Research 35: 248. https://doi.org/10.1657/1523-0430(2003)035[0248:CEOAPB]2.0.CO;2

Hamm, A., J. Veeck, N. Bektas, P. J. Wild, A. Hartmann, *et al.*, 2008 Frequent expression loss of Inter-alpha-trypsin inhibitor heavy chain (ITIH) genes in multiple human solid tumors: a systematic expression analysis. BMC Cancer 8: 25. https://doi.org/10.1186/1471-2407-8-25

Han, F., Y. Peng, L. Xu, and P. Xiao, 2014 Identification, characterization, and utilization of single copy genes in 29 angiosperm genomes. BMC Genomics 15: 504. https://doi.org/10.1186/1471-2164-15-504

Hartung, J., J. Wagener, R. Ruser, and H.-P. Piepho, 2019 Blocking and re-arrangement of pots in greenhouse experiments: which approach is more effective? Plant Methods 15: 143. https://doi.org/10.1186/s13007-019-0527-4

Hasegawa, M., H. Kishino, and T. Yano, 1985 Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22: 160–174. https://doi.org/10.1007/BF02101694

Hata, H., Y. Uemura, K. Ouchi, and H. Matsuba, 2019 Hybridization between an endangered freshwater fish and an introduced congeneric species and consequent genetic introgression. PLoS ONE 14: e0212452. https://doi.org/10.1371/journal.pone.0212452

Heenan, P. B., P. J. Lockhart, N. Kirkham, K. McBreen, and D. Havell, 2006 Relationships in the alpine *Ranunculus haastii* (Ranunculaceae) complex and recognition of *R. piliferus* and *R. acraeus* from southern New Zealand. New Zealand Journal of Botany 44: 425–441. https://doi.org/10.1080/0028825x.2006.9513034

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 37: 105–113.

Heiser, C. B., 1949 Natural hybridization with particular reference to introgression. The Botanical Review 15: 645–687. https://doi.org/10.1663/0006-8101(2003)069

Heiser, C. B., 1973 Introgression re-examined. The Botanical Review 39: 347–366. https://doi.org/10.1007/BF02859160

Hershkovitz, M. A., and E. A. Zimmer, 1996 Conservation Patterns in Angiosperm rDNA ITS2 Sequences. Nucleic Acids Research 24: 2857–2867. https://doi.org/10.1093/nar/24.15.2857

Hewezi, T., T. Lane, S. Piya, A. Rambani, J. H. Rice, *et al.*, 2017 Cyst Nematode Parasitism Induces Dynamic Changes in the Root Epigenome. Plant Physiology 174: 405–420. https://doi.org/10.1104/pp.16.01948

Hiatt, D., and S. L. Flory, 2020 Populations of a widespread invader and co-occurring native species vary in phenotypic plasticity. New Phytologist 225: 584–594. https://doi.org/10.1111/nph.16225

Hibbins, M. S., and M. W. Hahn, 2019 The Timing and Direction of Introgression Under the Multispecies Network Coalescent. Genetics 211: 1059–1073. https://doi.org/10.1534/genetics.118.301831

Hoang, N. V, A. Furtado, P. J. Mason, A. Marquardt, L. Kasirajan, *et al.*, 2017 A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. BMC Genomics 18: 395. https://doi.org/10.1186/s12864-017-3757-8

Hodges, S. A., and N. J. Derieg, 2009 Adaptive radiations: From field to genomic studies. Proceedings of the National Academy of Sciences 106: 9947–9954. https://doi.org/10.1073/pnas.0901594106

Hoffmann, A. A., and C. M. Sgró, 2011 Climate change and evolutionary adaptation. Nature 470: 479–485. https://doi.org/10.1038/nature09670

Hoffmann, A. A., P. Griffin, S. Dillon, R. Catullo, R. Rane, *et al.*, 2015 A framework for incorporating evolutionary genomics into biodiversity conservation and management. Climate Change Responses 2: 1. https://doi.org/10.1186/s40665-014-0009-x

Hokken, M. W. J., J. Zoll, J. P. M. Coolen, B. J. Zwaan, P. E. Verweij, *et al.*, 2019 Phenotypic plasticity and the evolution of azole resistance in *Aspergillus fumigatus*; an expression profile of clinical isolates upon exposure to itraconazole. BMC Genomics 20: 28. https://doi.org/10.1186/s12864-018-5255-z

Hölzer, M., and M. Marz, 2019 *De novo* transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. GigaScience 8: giz039. https://doi.org/10.1093/gigascience/giz039

Horandl, E., O. Paun, J. T. Johansson, C. Lehnebach, T. Armstrong, *et al.*, 2005 Phylogenetic relationships and evolutionary traits in *Ranunculus* s.l. (Ranunculaceae) inferred from ITS sequence analysis. Molecular Phylogenetics and Evolution 36: 305–327. https://doi.org/10.1016/j.ympev.2005.02.009

Hsieh, P.-H., Y.-J. Oyang, and C.-Y. Chen, 2019 Effect of *de novo* transcriptome assembly on transcript quantification. Scientific Reports 9: 8304. https://doi.org/10.1038/s41598-019-44499-3

Hu, Z., Y. Zhang, Y. He, Q. Cao, T. Zhang, *et al.*, 2020 Full-Length Transcriptome Assembly of Italian Ryegrass Root Integrated with RNA-seq to Identify Genes in Response to Plant Cadmium Stress. International Journal of Molecular Sciences 21: 1067. https://doi.org/10.3390/ijms21031067

Hubner, N., C. A. Wallace, H. Zimdahl, E. Petretto, H. Schulz, *et al.*, 2005 Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. Nature Genetics 37: 243–253. https://doi.org/10.1038/ng1522

Hurst, J. M., and R. Allen, 2007 *The Recce method for describing New Zealand vegetation: field protocols*. Manaaki Whenua Press, Lincoln, New Zealand.

Huson, D. H., and D. Bryant, 2006 Application of Phylogenetic Networks in Evolutionary Studies. Molecular Biology and Evolution 23: 254–267. https://doi.org/10.1093/molbev/msj030

Ilut, D. C., J. E. Coate, A. K. Luciano, T. G. Owens, G. D. May, *et al.*, 2012 A comparative transcriptomic study of an allotetraploid and its diploid progenitors illustrates the unique advantages and challenges of RNA-seq in plant species. American Journal of Botany 99: 383–396. https://doi.org/10.3732/ajb.1100312

Iovieno, P., P. Punzo, G. Guida, C. Mistretta, M. J. Van Oosten, *et al.*, 2016 Transcriptomic Changes Drive Physiological Responses to Progressive Drought Stress and Rehydration in Tomato. Frontiers in Plant Science 7: 371. https://doi.org/10.3389/fpls.2016.00371

Jay, P., A. Whibley, L. Frézal, M. Á. Rodríguez de Cara, R. W. Nowell, *et al.*, 2018 Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. Current Biology 28: 1839-1845.e3. https://doi.org/10.1016/j.cub.2018.04.072

Jezkova, T., and J. J. Wiens, 2016 Rates of change in climatic niches in plant and animal populations are much slower than projected climate change. Proceedings of the Royal Society B 283: 20162104. https://doi.org/10.1098/rspb.2016.2104

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. https://doi.org/10.1038/nature09916

Jiao, W.-B., and K. Schneeberger, 2017 The impact of third generation genomic technologies on plant genome assembly. Current Opinion in Plant Biology 36: 64–70. https://doi.org/10.1016/j.pbi.2017.02.002

Johnson, D. J., R. Condit, S. P. Hubbell, and L. S. Comita, 2017 Abiotic niche partitioning and negative density dependence drive tree seedling survival in a tropical forest. Proceedings of the Royal Society B 284: 20172210. https://doi.org/10.1098/rspb.2017.2210

Johnson, M. G., L. Pokorny, S. Dodsworth, L. R. Botigue, R. S. Cowan, *et al.*, 2019 A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using k-Medoids Clustering. Systematic Biology 68: 594–606. https://doi.org/10.1093/sysbio/syy086

Johnstun, J. A., V. Shankar, S. S. Mokashi, L. T. Sunkara, U. E. Ihearahu, *et al.*, 2021 Functional Diversification, Redundancy, and Epistasis among Paralogs of the *Drosophila melanogaster Obp50a–d* Gene Cluster. Molecular Biology and Evolution 38: 2030–2044. https://doi.org/10.1093/molbev/msab004

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. https://doi.org/10.1016/j.ympev.2009.02.015

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. https://doi.org/10.1086/600082

Joly, S., 2012 JML: Testing hybridization from species trees. Molecular Ecology Resources 12: 179–184. https://doi.org/10.1111/j.1755-0998.2011.03065.x

Joly, S., P. B. Heenan, and P. J. Lockhart, 2014 Species radiation by niche shifts in New Zealand's rockcresses (*Pachycladon*, Brassicaceae). Systematic Biology 63: 192–202. https://doi.org/10.1093/sysbio/syt104

Jump, A. S., and J. Penuelas, 2005 Running to stand still: adaptation and the response of plants to rapid climate change. Ecology Letters 8: 1010–1020. https://doi.org/10.1111/j.1461-0248.2005.00796.x

Kachroo, A. H., J. M. Laurent, C. M. Yellman, A. G. Meyer, C. O. Wilke, *et al.*, 2015 Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science 348: 921–925. https://doi.org/10.1126/science.aaa0769

Kadam, N. N., X. Yin, P. S. Bindraban, P. C. Struik, and K. S. V Jagadish, 2015 Does morphological and anatomical plasticity during the vegetative stage make wheat more tolerant of water deficit stress than rice? Plant Physiology 167: 1389–1401. https://doi.org/10.1104/pp.114.253328

Kalinowska, E., M. Chodorska, E. Paduch-Cichal, and K. Mroczkowska, 2012 An improved method for RNA isolation from plants using commercial extraction kits. Acta Biochimica Polonica 59.

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. https://doi.org/10.1093/bioinformatics/bts199

Kennedy, M., B. R. Holland, R. D. Gray, and H. G. Spencer, 2005 Untangling Long Branches: Identifying Conflicting Phylogenetic Signals Using Spectral Analysis, Neighbor-Net, and Consensus Networks. Systematic Biology 54: 620–633. https://doi.org/10.1080/106351591007462

Ketterings, Q., S. Reid, and R. Rao, 2007 *Cation exchange capacity (CEC). Fact sheet 22*. Available at http://nmsp.cals.cornell.edu/publications/factsheets/factsheet22.pdf (Accessed: 18 May 2021)

Khadka, V. S., K. Vaughn, J. Xie, P. Swaminathan, Q. Ma, *et al.*, 2019 Transcriptomic response is more sensitive to water deficit in shoots than roots of *Vitis riparia* (Michx.). BMC Plant Biology 19: 72. https://doi.org/10.1186/s12870-019-1664-7

Khoo, H. E., A. Azlan, S. T. Tang, and S. M. Lim, 2017 Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. Food & Nutrition Research 61: 1361779. https://doi.org/10.1080/16546628.2017.1361779

Kim, J., Y. Joo, J. Kyung, M. Jeon, J. Y. Park, *et al.*, 2018 A molecular basis behind heterophylly in an amphibious plant, *Ranunculus trichophyllus*. PLoS Genetics 14: e1007208. https://doi.org/10.1371/journal.pgen.1007208

Kimura, M., 1968 Evolutionary Rate at the Molecular Level. Nature 217: 624–626. https://doi.org/10.1038/217624a0

Kipling, R. P., and J. Warren, 2014 How generalists coexist: the role of floral phenotype and spatial factors in the pollination systems of two *Ranunculus* species. Journal of Plant Ecology 7: 480–489. https://doi.org/10.1093/jpe/rtt040

Kirchner, M., T. Faus-Kessler, G. Jakobi, M. Leuchner, L. Ries, *et al.*, 2013 Altitudinal temperature lapse rates in an Alpine valley: trends and the influence of season and weather patterns. International Journal of Climatology 33: 539–555. https://doi.org/10.1002/joc.3444

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. https://doi.org/10.3390/ijms12063966

Kirkman, J. H., A. Basker, A. Surapaneni, and A. N. MacGregor, 1994 Potassium in the soils of New Zealand - A review. New Zealand Journal of Agricultural Research 37: 207–227. https://doi.org/10.1080/00288233.1994.9513059

Kitajima, M., and W. L. Butler, 1975 Quenching of chlorophyll fluorescence and primary photochemistry in chloroplasts by dibromothymoquinone. Biochimica et Biophysica Acta (BBA)-Bioenergetics 376: 105–115. https://doi.org/10.1016/0005-2728(75)90209-1

Klein, A., I. Steffan-Dewenter, and T. Tscharntke, 2003 Bee pollination and fruit set of *Coffea arabica* and *C. canephora* (Rubiaceae). American Journal of Botany 90: 153–157. https://doi.org/10.3732/ajb.90.1.153

Kluyver, T., B. Ragan-Kelley, F. Pérez, B. E. Granger, M. Bussonnier, *et al.*, 2016 Jupyter Notebooks-a publishing format for reproducible computational workflows, pp. 87–90 in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, edited by Loizides F., Scmidt B. IOS Press, Amsterdam.

Koenig, D., J. M. Jiménez-Gómez, S. Kimura, D. Fulop, D. H. Chitwood, *et al.*, 2013 Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proceedings of the National Academy of Sciences 110: E2655–E2662. https://doi.org/10.1073/pnas.1309606110

Kozlov, A. M., D. Darriba, T. Flouri, B. Morel, and A. Stamatakis, 2019 RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35: 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Kriventseva, E. V, D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, *et al.*, 2018 OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Research 47: 807–811. https://doi.org/10.1093/nar/gky1053

Kryvokhyzha, D., P. Milesi, T. Duan, M. Orsucci, S. I. Wright, *et al.*, 2019 Towards the new normal: Transcriptomic convergence and genomic legacy of the two subgenomes of an allopolyploid weed (*Capsella bursa-pastoris*). PLoS Genetics 15: e1008131. https://doi.org/10.1371/journal.pgen.1008131

Kuhner M. K., and J. Felsenstein, 1994 A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Molecular Biology and Evolution 11: 459–468. https://doi.org/10.1093/oxfordjournals.molbev.a040126

Kumar S., and T. Mohapatra, 2021 Dynamics of DNA Methylation and Its Functions in Plant Growth and Development. Frontiers in Plant Science 12: 858. https://doi.org/10.3389/fpls.2021.596236

Kyriakidou, M., H. H. Tai, N. L. Anglin, D. Ellis, and M. V Strömvik, 2018 Current Strategies of Polyploid Plant Genome Sequence Assembly. Frontiers in Plant Science 9: 1660. https://doi.org/10.3389/fpls.2018.01660

Kyriakidou, M., N. L. Anglin, D. Ellis, H. H. Tai, and M. V Strömvik, 2020 Genome assembly of six polyploid potato genomes. Scientific Data 7: 88. https://doi.org/10.1038/s41597-020-0428-4

Lafond, M., M. Meghdari Miardan, and D. Sankoff, 2018 Accurate prediction of orthologs in the presence of divergence after duplication. Bioinformatics 34: i366–i375. https://doi.org/10.1093/bioinformatics/bty242

Lai, Y. T., C. K. L. Yeung, K. E. Omland, E. L. Pang, Y. Hao, *et al.*, 2019 Standing genetic variation as the predominant source for adaptation of a songbird. Proceedings of the National Academy of Sciences of the United States of America 116: 2152–2157. https://doi.org/10.1073/pnas.1813597116

Lamarre, S., P. Frasse, M. Zouine, D. Labourdette, E. Sainderichin, *et al.*, 2018 Optimization of an RNA-Seq Differential Gene Expression Analysis Depending on Biological Replicate Number and Library Size. Frontiers in Plant Science 9: 108. https://doi.org/10.3389/fpls.2018.00108

Landis, M. J., J. G. Schraiber, and M. Liang, 2013 Phylogenetic Analysis Using Lévy Processes: Finding Jumps in the Evolution of Continuous Traits. Systematic Biology 62: 193–204. https://doi.org/10.1093/sysbio/sys086

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nature methods 9: 357–359. https://doi.org/10.1038/nmeth.1923

Lawlor, D. W., 2012 Genetic engineering to improve plant performance under drought: physiological evaluation of achievements, limitations, and possibilities. Journal of Experimental Botany 64: 83–108. https://doi.org/10.1093/jxb/ers326

Lee-Yaw, J. A., C. J. Grassa, S. Joly, R. L. Andrew, and L. H. Rieseberg, 2019 An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). New Phytologist 221: 515–526. https://doi.org/10.1111/nph.15386

Lehnebach, C. A., 2008 *Phylogenetic affinities, species delimitation and adaptive radius of New Zealand* Ranunculus. Massey University, Palmerston North, New Zealand.

Leinweber, A., M. Weigert, and R. Kümmerli, 2018 The bacterium *Pseudomonas aeruginosa* senses and gradually responds to interspecific competition for iron. Evolution 72: 1515–1528. https://doi.org/10.1111/evo.13491

Leitch, I. J., and M. D. Bennett, 2004 Genome downsizing in polyploid plants. Biological Journal of the Linnean Society 82: 651–663. https://doi.org/10.1111/j.1095-8312.2004.00349.x

Leitch, I. J., L. Hanson, K. Y. Lim, A. Kovarik, M. W. Chase, *et al.*, 2008 The Ups and Downs of Genome Size Evolution in Polyploid Species of *Nicotiana* (Solanaceae). Annals of Botany 101: 805–814.

Lewontin, R. C., and L. C. Birch, 1966 Hybridization as a Source of Variation for Adaptation to New Environments. Evolution 20: 315–336. https://doi.org/10.1111/j.1558-5646.1966.tb03369.x

Li, W., and A. Godzik, 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, B., and C. N. Dewey, 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323. https://doi.org/10.1186/1471-2105-12-323

Li, D.-Z., L.-M. Gao, H.-T. Li, H. Wang, X.-J. Ge, *et al.*, 2011 Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proceedings of the National Academy of Sciences 108: 19641–19646. https://doi.org/10.1073/pnas.1104551108

Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997 (Preprint posted March 16, 2013).

Li, B., N. Fillmore, Y. Bai, M. Collins, J. A. Thomson, *et al.*, 2014 Evaluation of *de novo* transcriptome assemblies from RNA-Seq data. Genome Biology 15: 553. https://doi.org/10.1186/s13059-014-0553-5

Li, B., L. Wang, K. F. Kaseke, L. Li, and M. K. Seely, 2016 The impact of rainfall on soil moisture dynamics in a foggy desert. PLoS ONE 11: e0164982. https://doi.org/10.1371/journal.pone.0164982

Li, F.-D., W. Tong, E.-H. Xia, and C.-L. Wei, 2019a Optimized sequencing depth and de novo assembler for deeply reconstructing the transcriptome of the tea plant, an economically important plant species. BMC Bioinformatics 20: 553. https://doi.org/10.1186/s12859-019-3166-x

Li, F., C. Wu, M. Gao, M. Jiao, C. Qu, *et al.*, 2019b Transcriptome sequencing, molecular markers, and transcription factor discovery of *Platanus acerifolia* in the presence of *Corythucha ciliata*. Scientific Data 6: 128. https://doi.org/10.1038/s41597-019-0111-9

Li, H., C. Testerink, and Y. Zhang, 2021 How roots and shoots communicate through stressful times. Trends in Plant Science. https://doi.org/10.1016/j.tplants.2021.03.005

Libohova, Z., C. Seybold, D. Wysocki, S. Wills, P. Schoeneberger, *et al.*, 2018 Reevaluating the effects of soil organic matter and other properties on available water-holding capacity using the National Cooperative Soil Survey Characterization Database. Journal of Soil and Water Conservation 73: 411–421. https://doi.org/10.2489/jswc.73.4.411

Liu, Z., B. Dugan, C. A. Masiello, and H. M. Gonnermann, 2017 Biochar particle size, shape, and porosity act together to influence soil water properties. PLoS ONE 12: e0179079. https://doi.org/10.1371/journal.pone.0179079

Liu, X., R. Zhang, H. Ou, Y. Gui, J. Wei, *et al.*, 2018 Comprehensive transcriptome analysis reveals genes in response to water deficit in the leaves of *Saccharum narenga* (Nees ex Steud.) hack. BMC Plant Biology 18: 250. https://doi.org/10.1186/s12870-018-1428-9

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. https://doi.org/10.1016/j.ympev.2020.106784

Llopart, A., D. Herrig, E. Brud, and Z. Stecklein, 2014 Sequential adaptive introgression of the mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*. Molecular Ecology 23: 1124–1136. https://doi.org/10.1111/mec.12678

Lockhart, P. J., A. W. D. Larkum, M. Steel, P. J. Waddell, and D. Penny, 1996 Evolution of chlorophyll and bacteriochlorophyll: The problem of invariant sites in sequence analysis. Proceedings of the National Academy of Sciences 93: 1930 LP – 1934. https://doi.org/10.1073/pnas.93.5.1930

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 88: 458–477. https://doi.org/10.2307/3298586

Lockhart, P. J., A. W. D. Larkum, M. Becker, and D. Penny, 2014 We are Still Learning About the Nature of Species and Their Evolutionary Relationships. Annals of the Missouri Botanical Garden 100: 6–13. https://doi.org/10.3417/2012084

Lord, M. S., J. Melrose, A. J. Day, and J. M. Whitelock, 2020 The Inter-α-Trypsin Inhibitor Family: Versatile Molecules in Biology and Pathology. Journal of Histochemistry & Cytochemistry 68: 907–927. https://doi.org/10.1369/0022155420940067

Love, M. I., W. Huber, and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15: 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, T., H. Yu, Q. Li, L. Chai, and W. Jiang, 2019 Improving Plant Growth and Alleviating Photosynthetic Inhibition and Oxidative Stress From Low-Light Stress With Exogenous GR24 in Tomato (*Solanum lycopersicum* L.) Seedlings. Frontiers in Plant Science 10: 490.

Luo, D., Q. Zhou, Y. Wu, X. Chai, W. Liu, *et al.*, 2019a Full-length transcript sequencing and comparative transcriptomic analysis to evaluate the contribution of osmotic and ionic stress components towards salinity tolerance in the roots of cultivated alfalfa (*Medicago sativa* L.). BMC Plant Biology 19: 32. https://doi.org/10.1186/s12870-019-1630-4

Luo, Y., B. Fang, W. Wang, Y. Yang, L. Rao, *et al.*, 2019b Genome-wide analysis of the rice J-protein family: identification, genomic organization, and expression profiles under multiple stresses. 3 Biotech 9: 358. https://doi.org/10.1007/s13205-019-1880-8

Lusk, C. H., S. K. Wiser, and D. C. Laughlin, 2020 Macroclimate and Topography Interact to Influence the Abundance of Divaricate Plants in New Zealand. Frontiers in Plant Science 11: 507. https://doi.org/10.3389/fpls.2020.00507

Lute, A. C., and J. T. Abatzoglou, 2021 Best practices for estimating near-surface air temperature lapse rates. International Journal of Climatology 41: E110–E125. https://doi.org/10.1002/joc.6668

Ma, J., R. Q. Li, H. G. Wang, D. X. Li, X. Y. Wang, *et al.*, 2017 Transcriptomics Analyses Reveal Wheat Responses to Drought Stress during Reproductive Stages under Field Conditions. Frontiers in Plant Science 8: 13. https://doi.org/10.3389/fpls.2017.00592

Mable, B. K., 2019 Conservation of adaptive potential and functional diversity: integrating old and new approaches. Conservation Genetics 20: 89–100. https://doi.org/10.1007/s10592-018-1129-9

MacManes, M. D., 2014 On the optimal trimming of high-throughput mRNA sequence data. Frontiers in Denetics 5: 13. https://doi.org/10.3389/fgene.2014.00013

Madden, T. L., B. Busby, and J. Ye, 2019 Reply to the paper: Misunderstood parameters of NCBI BLAST impacts the correctness of bioinformatics workflows. Bioinformatics 35: 2699–2700. https://doi.org/10.1093/bioinformatics/bty1026

Maddison, W. P., 1997 Gene Trees in Species Trees. Systematic Biology 46: 523–536. https://doi.org/10.1093/sysbio/46.3.523

Madritsch, S., A. Burg, and E. M. Sehr, 2021 Comparing de novo transcriptome assembly tools in di-and autotetraploid non-model plant species. BMC Bioinformatics 22: 146. https://doi.org/10.1186/s12859-021-04078-8

Mahmood, K., J. Orabi, P. S. Kristensen, P. Sarup, L. N. Jørgensen, *et al.*, 2020 De novo transcriptome assembly, functional annotation, and expression profiling of rye (*Secale cereale* L.) hybrids inoculated with ergot (*Claviceps purpurea*). Scientific Reports 10: 13475. https://doi.org/10.1038/s41598-020-70406-2

Malik, L., F. Almodaresi, and R. Patro, 2018 Grouper: graph-based clustering and annotation for improved de novo transcriptome analysis. Bioinformatics 34: 3265–3272. https://doi.org/10.1093/bioinformatics/bty378.

Mallet, J., N. Besansky, and M. W. Hahn, 2016 How reticulated are species? BioEssays 38: 140–149. https://doi.org/10.1002/bies.201500149

Manaaki Whenua – Landcare Research, 2020 *The New Zealand SoilsMapViewer*. Available at: https://soils-maps.landcareresearch.co.nz/ (Accessed: 20 May 2021)

Maxwell, K., and G. N. Johnson, 2000 Chlorophyll fluorescence—a practical guide. Journal of Experimental Botany 51: 659–668. https://doi.org/10.1093/jexbot/51.345.659

McFarlane, S. E., and J. M. Pemberton, 2019 Detecting the True Extent of Introgression during Anthropogenic Hybridization. Trends in Ecology & Evolution 34: 315–326. https://doi.org/10.1016/j.tree.2018.12.013

McGlone, M. S., and C. J. Webb, 1981 Selective forces influencing the evolution of divaricating plants. New Zealand Journal of Ecology 4: 20–28.

McGlone, M., and S. Walker, 2011 *Potential effects of climate change on New Zealand's terrestrial biodiversity and policy recommendations for mitigation, adaptation and research.* Department of Conservation, Wellington, New Zealand.

McLaren, R. G., and K. C. Cameron, 1996 *Soil Science: Sustainable production and environmental protection*. Oxford University Press, Auckland, New Zealand.

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: 336. https://doi.org/10.3389/fpls.2021.637214

Minasny, B., and A. B. McBratney, 2018 Limited effect of organic matter on soil available water capacity. European Journal of Soil Science 69: 39–47. https://doi.org/10.1111/ejss.12475

Minder, J. R., P. W. Mote, and J. D. Lundquist, 2010 Surface temperature lapse rates over complex terrain: Lessons from the Cascade Mountains. Journal of Geophysical Research: Atmospheres 115. https://doi.org/10.1029/2009JD013493

Ministry for the Environment, 2018 *Climate Change Projections for New Zealand: Atmosphere Projections Based on Simulations from the IPCC Fifth Assessment, 2nd edition*. Ministry for the Environment, Wellington, New Zealand.

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. https://doi.org/10.1038/s41598-019-43119-4

Moreno-Hagelsieb, G., and K. Latimer, 2008 Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatics 24: 319–324. https://doi.org/10.1093/bioinformatics/btm585

Moreno-Santillán, D. D., C. Machain-Williams, G. Hernández-Montes, and J. Ortega, 2019 De Novo Transcriptome Assembly and Functional Annotation in Five Species of Bats. Scientific Reports 9: 6222. https://doi.org/10.1038/s41598-019-42560-9

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. https://doi.org/10.1080/03014220909510561

Moura, R. F., D. Queiroga, E. Vilela, and A. P. Moraes, 2021 Polyploidy and high environmental tolerance increase the invasive success of plants. Journal of Plant Research 134: 105–114. https://doi.org/10.1007/s10265-020-01236-6

Muktar, M. S., A. Teshome, J. Hanson, A. T. Negawo, E. Habte, *et al.*, 2019 Genotyping by sequencing provides new insights into the diversity of Napier grass (*Cenchrus purpureus*) and reveals variation in genome-wide LD patterns between collections. Scientific Reports 9: 6936. https://doi.org/10.1038/s41598-019-43406-0

Mummenhoff, K., P. Linder, N. Friesen, J. L. Bowman, J. Lee, *et al.*, 2004 Molecular evidence for bicontinental hybridogenous genomic constitution in *Lepidium* sensu stricto (Brassicaceae) species from Australia and New Zealand. American Journal of Botany 91: 254–261. https://doi.org/10.3732/ajb.91.2.254

Murchie, E. H., and T. Lawson, 2013 Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. Journal of Experimental Botany 64: 3983–3998. https://doi.org/10.1093/jxb/ert208

Naidoo, K., E. T. Steenkamp, M. P. A. Coetzee, M. J. Wingfield, and B. D. Wingfield, 2013 Concerted Evolution in the Ribosomal RNA Cistron. PLoS ONE 8: e59355. https://doi.org/10.1371/journal.pone.0059355

Nei, M., and A. P. Rooney, 2005 Concerted and Birth-and-Death Evolution of Multigene Families. Annual Review of Genetics 39: 121–152. https://doi.org/10.1146/annurev.genet.39.073003.112240

Neji, M., A. Gorel, D. I. Ojeda, J. Duminil, C. Kastally, *et al.*, 2019 Comparative analysis of two sister *Erythrophleum* species (Leguminosae) reveal contrasting transcriptome-wide responses to early drought stress. Gene 694: 50–62. https://doi.org/10.1016/j.gene.2019.01.027

Nevado, B., E. L. Y. Wong, O. G. Osborne, and D. A. Filatov, 2019 Adaptive Evolution Is Common in Rapid Evolutionary Radiations. Current Biology 29: 3081-3086.e5. https://doi.org/10.1016/j.cub.2019.07.059

Newnham, R. M., D. J. Lowe, and P. W. Williams, 1999 Quaternary environmental change in New Zealand: a review. Progress in Physical Geography 23: 567–610. https://doi.org/10.1177/030913339902300406

Nichio, B. T. L., J. N. Marchaukoski, and R. T. Raittz, 2017 New Tools in Orthology Analysis: A Brief Review of Promising Perspectives. Frontiers in Genetics 8: 165. https://doi.org/10.3389/fgene.2017.00165

Nielsen, S. K. D., T. L. Koch, F. Hauser, A. Garm, and C. J. P. Grimmelikhuijzen, 2019 De novo transcriptome assembly of the cubomedusa *Tripedalia cystophora*, including the analysis of a set of genes involved in peptidergic neurotransmission. BMC Genomics 20: 175. https://doi.org/10.1186/s12864-019-5514-7

Niu, Y., G. Chen, D. Peng, B. Song, Y. Yang, *et al.*, 2014 Grey leaves in an alpine plant: a cryptic colouration to avoid attack? New Phytologist 203: 953–963. https://doi.org/10.1111/nph.12834

Niu, Y., Z. Chen, M. Stevens, and H. Sun, 2017 Divergence in cryptic leaf colour provides local camouflage in an alpine plant. Proceedings of the Royal Society B 284: 20171654. https://doi.org/10.1098/rspb.2017.1654

NIWA, 2010 *Climate Summaries.* Available at https://niwa.co.nz/education-and-training/schools/resources/climate/summary (Accessed: 30 March 2019)

NIWA, 2021a *Monthly climate summaries from December 2001 to the present.* Available at https://niwa.co.nz/climate/monthly (Accessed: 13 May 2021)

NIWA, 2021b *Seasonal climate summaries from summer 2001 to the present.* Available at https://niwa.co.nz/climate/summaries/seasonal (Accessed: 13 May 2021)

Njuguna, W., A. Liston, R. Cronn, T.-L. Ashman, and N. Bassil, 2013 Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. Molecular Phylogenetics and Evolution 66: 17–29. https://doi.org/10.1016/j.ympev.2012.08.026

Norton, D. A., 1985 A Multivariate Technique for Estimating New Zealand Temperature Normals. Weather and Climate 5: 64–74. https://doi.org/10.2307/44279988

O'Brien, E. K., M. Higgie, A. Reynolds, A. A. Hoffmann, and J. R. Bridle, 2017 Testing for local adaptation and evolutionary potential along altitudinal gradients in rainforest *Drosophila*: beyond laboratory estimates. Global Change Biology 23: 1847–1860. https://doi.org/10.1111/gcb.13553

O'Kelly, B. C., 2005 Oven-Drying Characteristics of Soils of Different Origins. Drying Technology 23: 1141–1149. https://doi.org/10.1081/DRT-200059149

O'Neil, S. T., and S. J. Emrich, 2013 Assessing *De Novo* transcriptome assembly metrics for consistency and utility. BMC Genomics 14: 465. https://doi.org/10.1186/1471-2164-14-465

Ouyang, K., J. Li, H. Huang, Q. Que, P. Li, *et al.*, 2014 A simple method for RNA isolation from various tissues of the tree *Neolamarckia cadamba*. Biotechnology & Biotechnological Equipment 28: 1008–1013. https://doi.org/10.1080/13102818.2014.981086

Patro, R., G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford, 2017 Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods 14: 417–419. https://doi.org/10.1038/nmeth.4197

Payá-Milans, M., J. W. Olmstead, G. Nunez, T. A. Rinehart, and M. Staton, 2018 Comprehensive evaluation of RNA-seq analysis pipelines in diploid and polyploid species. GigaScience 7: giy132. https://doi.org/10.1093/gigascience/giy132

Pease J. B., and M. W. Hahn, 2015 Detection and Polarization of Introgression in a Five-Taxon Phylogeny. Systematic Biology 64: 651–662. https://doi.org/10.1093/sysbio/syv023

Peng, Z., S. He, W. Gong, F. Xu, Z. Pan, *et al.*, 2018 Integration of proteomic and transcriptomic profiles reveals multiple levels of genetic regulation of salt tolerance in cotton. BMC Plant Biology 18: 128. https://doi.org/10.1186/s12870-018-1350-1

Penny, D., L. R. Foulds, and M. D. Hendy, 1982 Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. Nature 297: 197–200. https://doi.org/10.1038/297197a0

Penny, D., and M. D. Hendy, 1985 The Use of Tree Comparison Metrics. Systematic Zoology 34: 75–82. https://doi.org/10.2307/2413347

Pickering, C. M., 1997 Breeding systems of Australian Ranunculus in the alpine region. Nordic Journal of Botany 17: 613–620. https://doi.org/10.1111/j.1756-1051.1997.tb00357.x

Pineiro, M., M. Andres, M. Iturralde, S. Carmona, J. Hirvonen, *et al.*, 2004 ITIH4 (inter-alpha-trypsin inhibitor heavy chain 4) is a new acute-phase protein isolated from cattle during experimental infection. Infection and Immunity 72: 3777–3782. https://doi.org/10.1128/IAI.72.7.3777-3782.2004

Pirie, M. D., K. M. Lloyd, W. G. Lee, and H. P. Linder, 2010 Diversification of *Chionochloa* (Poaceae) and biogeography of the New Zealand Southern Alps. Journal of Biogeography 37: 379–392. https://doi.org/10.1111/j.1365-2699.2009.02205.x

Polz, M. F., and C. M. Cavanaugh, 1998 Bias in Template-to-Product Ratios in Multitemplate PCR. Applied and Environmental Microbiology 64: 3724–3730. https://doi.org/10.1128/AEM.64.10.3724-3730.1998.

Posada, D., and T. R. Buckley, 2004 Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. Systematic biology 53: 793–808. https://doi.org/10.1080/10635150490522304

Pruitt, K. D., T. Tatusova, and D. R. Maglott, 2007 NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research 35: D61–D65. https://doi.org/10.1093/nar/gkl842

Puértolas, J., E. K. Larsen, W. J. Davies, and I. C. Dodd, 2017 Applying "drought" to potted plants by maintaining suboptimal soil moisture improves plant water relations. Journal of Experimental Botany 68: 2413–2424. https://doi.org/10.1093/jxb/erx116

Puglia, G. D., A. D. Prjibelski, D. Vitale, E. Bushmanova, K. J. Schmid, *et al.*, 2020 Hybrid transcriptome sequencing approach improved assembly and gene annotation in *Cynara cardunculus* (L.). BMC Genomics 21: 317. https://doi.org/10.1186/s12864-020-6670-5

Pulido, P., and D. Leister, 2018 Novel DNAJ-related proteins in *Arabidopsis thaliana*. New Phytologist 217: 480–490. https://doi.org/10.1111/nph.14827

Qiao, D., C. Yang, J. Chen, Y. Guo, Y. Li, *et al.*, 2019 Comprehensive identification of the full-length transcripts and alternative splicing related to the secondary metabolism pathways in the tea plant (*Camellia sinensis*). Scientific Reports 9: 2709. https://doi.org/10.1038/s41598-019-39286-z

Quansah, C., 1981 The Effect of Soil Type, Slope, Rain Intensity and Their Interactions on Splash Detachment and Transport. Journal of Soil Science 32: 215–224. https://doi.org/10.1111/j.1365-2389.1981.tb01701.x

Quek, Z. B. R., and D. Huang, 2019 Effects of missing data and data type on phylotranscriptomic analysis of stony corals (Cnidaria: Anthozoa: Scleractinia). Molecular Phylogenetics and Evolution 134: 12–23. https://doi.org/10.1016/j.ympev.2019.01.012

R Core Team, 2020 *R: A language and environment for statistical computing*

Raeymaekers, J. A. M., A. Chaturvedi, P. I. Hablützel, I. Verdonck, B. Hellemans, *et al.*, 2017 Adaptive and non-adaptive divergence in a common landscape. Nature Communications 8: 267. https://doi.org/10.1038/s41467-017-00256-6

Ramsey, J., 2011 Polyploidy and ecological adaptation in wild yarrow. Proceedings of the National Academy of Sciences 108: 7096–7101. https://doi.org/10.1073/pnas.1016631108

Rana, K., C. Atri, J. Akhatar, R. Kaur, A. Goyal, *et al.*, 2019 Detection of first marker trait associations for resistance against *Sclerotinia sclerotiorum* in *Brassica juncea–Erucastrum cardaminoides* introgression lines. Frontiers in Plant Science 10: 1015. https://doi.org/10.3389/fpls.2019.01015

Raven, P. H., 1973 Evolution of subalpine and alpine plant groups in New Zealand. New Zealand Journal of Botany 11: 177–200. https://doi.org/10.1080/0028825X.1973.10430272

Renny-Byfield, S., and J. F. Wendel, 2014 Doubling down on genomes: Polyploidy and crop plants. American Journal of Botany 101: 1711–1725. https://doi.org/10.3732/ajb.1400119

Renwick, J., B. Mullan, L. Wilcocks, C. Zammit, J. Sturman, *et al.*, 2013 *Four Degrees of Global Warming : Effects on the New Zealand Primary Sector.* National Institute of Water and Atmospheric Research, Wellington, New Zealand.

Rhymer, J. M., and D. Simberloff, 1996 Extinction by Hybridization and Introgression. Annual Review of Ecology and Systematics 27: 83–109. https://doi.org/10.1146/annurev.ecolsys.27.1.83

Robinson, D. F., and L. R. Foulds, 1981 Comparison of phylogenetic trees. Mathematical Biosciences 53: 131–147. https://doi.org/10.1016/0025-5564(81)90043-2

Robinson, M. D., D. J. McCarthy, and G. K. Smyth, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rodrigues, D. M., C. Turchetto, J. S. Lima, and L. B. Freitas, 2019 Diverse yet endangered: Pollen dispersal and mating system reveal inbreeding in a narrow endemic plant. Plant Ecology & Diversity 12: 169–180. https://doi.org/10.1080/17550874.2019.1610914

Rohlfs, R. V, P. Harrigan, and R. Nielsen, 2013 Modeling Gene Expression Evolution with an Extended Ornstein – Uhlenbeck Process Accounting for Within-Species Variation. Molecular Biology and Evolution 31: 201–211. https://doi.org/10.1093/molbev/mst190

Rohlfs, R. V., and R. Nielsen, 2015 Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution. Systematic Biology 64: 695–708. https://doi.org/10.1093/sysbio/syv042

Román-Palacios, C., and J. J. Wiens, 2020 Recent responses to climate change reveal the drivers of species extinction and survival. Proceedings of the National Academy of Sciences 117: 4211–4217. https://doi.org/10.1073/pnas.1913007117

Rooney, A. P., and T. J. Ward, 2005 Evolution of a large ribosomal RNA multigene family in filamentous fungi: Birth and death of a concerted evolution paradigm. Proceedings of the National Academy of Sciences 102: 5084–5089. https://doi.org/10.1073/pnas.0409689102

Roy, A., A. Dutta, D. Roy, P. Ganguly, R. Ghosh, *et al.*, 2016 Deciphering the role of the AT-rich interaction domain and the HMG-box domain of ARID-HMG proteins of *Arabidopsis thaliana*. Plant Molecular Biology 92: 371–388. https://doi.org/10.1007/s11103-016-0519-y

RStudio Team, 2020 *RStudio: Integrated Development Environment for R*, PBC, Boston, MA URL http://www.rstudio.com/

Salas-Muñoz, S., A. A. Rodríguez-Hernández, M. A. Ortega-Amaro, F. B. Salazar-Badillo, and J. F. Jiménez-Bremont, 2016 *Arabidopsis AtDjA3* Null Mutant Shows Increased Sensitivity to Abscisic Acid, Salt, and Osmotic Stress in Germination and Post-germination Stages. Frontiers in Plant Science 7: 220. https://doi.org/10.3389/fpls.2016.00220

Salzman, R. A., T. Fujita, K. Zhu-Salzman, P. M. Hasegawa, and R. A. Bressan, 1999 An Improved RNA Isolation Method for Plant Tissues Containing High Levels of Phenolic Compounds or Carbohydrates. Plant Molecular Biology Reporter 17: 11–17. https://doi.org/10.1023/A:1007520314478

Sánchez, C., J. Villacreses, N. Blanc, L. Espinoza, C. Martinez, *et al.*, 2016 High quality RNA extraction from Maqui berry for its application in next-generation sequencing. SpringerPlus 5: 1243. https://doi.org/10.1186/s40064-016-2906-x

Sanders, S. M., and P. Cartwright, 2015 Interspecific Differential Expression Analysis of RNA-Seq Data Yields Insight into Life Cycle Variation in Hydractiniid Hydrozoans. Genome Biology and Evolution 7: 2417–2431. https://doi.org/10.1093/gbe/evv153

Sanger, F., S. Nicklen, and A. R. Coulson, 1977 DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74: 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Saxton, K. E., and W. J. Rawls, 2006 Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. Soil Science Society of America Journal 70: 1569–1578. https://doi.org/10.2136/sssaj2005.0117

Sayers, E. W., R. Agarwala, E. E. Bolton, J. R. Brister, K. Canese, *et al.*, 2019a Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 47: D23–D28. https://doi.org/10.1093/nar/gky1069

Sayers, E. W., M. Cavanaugh, K. Clark, J. Ostell, K. D. Pruitt, *et al.*, 2019b GenBank. Nucleic Acids Research 47: D94–D99. https://doi.org/10.1093/nar/gky989

Scalabrin, S., L. Toniutti, G. Di Gaspero, D. Scaglione, G. Magris, *et al.*, 2020 A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. Scientific Reports 10: 4642. https://doi.org/10.1038/s41598-020-61216-7

Scheepens, J. F., Y. Deng, and O. Bossdorf, 2018 Phenotypic plasticity in response to temperature fluctuations is genetically variable, and relates to climatic variability of origin, in *Arabidopsis thaliana*. AoB Plants 10: ply043. https://doi.org/10.1093/aobpla/ply043

Scheiner, S. M., M. Barfield, and R. D. Holt, 2017 The genetics of phenotypic plasticity. XV. Genetic assimilation, the Baldwin effect, and evolutionary rescue. Ecology and Evolution 7: 8788–8803. https://doi.org/10.1002/ece3.3429

Scherff, E. J., C. Galen, and M. L. Stanton, 1994 Seed Dispersal, Seedling Survival and Habitat Affinity in a Snowbed Plant: Limits to the Distribution of the Snow Buttercup, Ranunculus adoneus. Oikos 69: 405–413. https://doi.org/10.2307/3545853

Schliep, K. P., 2011 phangorn: phylogenetic analysis in R. Bioinformatics 27: 592–593. https://doi.org/10.1093/bioinformatics/btq706

Schluter, D., 2000 *The Ecology of Adaptive Radiation*. Oxford University Press, New York.

Schmickl, R., S. Marburger, S. Bray, and L. Yant, 2017 Hybrids and horizontal transfer: introgression allows adaptive allele discovery. Journal of Experimental Botany 68: 5453–5470. https://doi.org/10.1093/jxb/erx297

Schmidt, T. L., M. Jasper, A. R. Weeks, and A. A. Hoffmann, 2021 Unbiased population heterozygosity estimates from genome-wide sequence data. Methods in Ecology and Evolution. https://doi.org/10.1111/2041-210X.13659

Schnable, J. C., N. M. Springer, and M. Freeling, 2011 Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences 108: 4069–4074. https://doi.org/10.1073/pnas.1101368108

Schrieber, K., S. Wolf, C. Wypior, D. Höhlig, I. Hensen, *et al.*, 2017 Adaptive and non-adaptive evolution of trait means and genetic trait correlations for herbivory resistance and performance in an invasive plant. Oikos 126: 572–582. https://doi.org/10.1111/oik.03781

Schrieber, K., S. C. Paul, L. V. Höche, A. C. Salas, R. Didszun, *et al.*, 2021 Inbreeding in a dioecious plant has sex-and population origin-specific effects on its interactions with pollinators. eLife 10: e65610. https://doi.org/10.7554/eLife.65610

Schurch, N. J., P. Schofield, M. Gierliński, C. Cole, A. Sherstnev, *et al.*, 2016 How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA 22: 839–851. https://doi.org/10.1261/rna.053959.115

Sexton, J. P., P. J. McIntyre, A. L. Angert, and K. J. Rice, 2009 Evolution and Ecology of Species Range Limits. Annual Review of Ecology, Evolution, and Systematics 40: 415–436. https://doi.org/10.1146/annurev.ecolsys.110308.120317

Shah, N., M. G. Nute, T. Warnow, and M. Pop, 2019 Misunderstood parameter of NCBI BLAST impacts the correctness of bioinformatics workflows. Bioinformatics 35: 1613–1614. https://doi.org/10.1093/bioinformatics/bty833

Shi, C., N. Hu, H. Huang, J. Gao, Y.-J. Zhao, *et al.*, 2012 An Improved Chloroplast DNA Extraction Procedure for Whole Plastid Genome Sequencing. PLoS ONE 7: e31468. https://doi.org/10.1371/journal.pone.0031468

Shi, C., S. Wang, E.-H. Xia, J.-J. Jiang, F.-C. Zeng, *et al.*, 2016 Full transcription of the chloroplast genome in photosynthetic eukaryotes. Scientific Reports 6: 30135. https://doi.org/10.1038/srep30135

Silvertown, J., Y. Araya, and D. Gowing, 2015 Hydrological niches in terrestrial plant communities: a review. Journal of Ecology 103: 93–108. https://doi.org/10.1111/1365-2745.12332

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V Kriventseva, and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simmons, M. P., H. Ochoterena, and J. V Freudenstein, 2002 Amino acid vs. nucleotide characters: challenging preconceived notions. Molecular Phylogenetics and Evolution 24: 78–90. https://doi.org/10.1016/S1055-7903(02)00202-6

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. https://doi.org/10.1080/00288250709509723

Smissen, R. D., S. J. Richardson, C. W. Morse, and P. B. Heenan, 2014 Relationships, gene flow and species boundaries among New Zealand *Fuscospora* (Nothofagaceae: southern beech). New Zealand Journal of Botany 52: 389–406. https://doi.org/10.1080/0028825X.2014.960946

Snell-Rood, E. C., M. E. Kobiela, K. L. Sikkink, and A. M. Shephard, 2018 Mechanisms of Plastic Rescue in Novel Environments. Annual Review of Ecology, Evolution, and Systematics 49: 331–354. https://doi.org/10.1146/annurev-ecolsys-110617-062622

Soltis, D. E., and P. S. Soltis, 1999 Polyploidy: recurrent formation and genome evolution. Trends in Ecology & Evolution 14: 348–352. https://doi.org/10.1016/S0169-5347(99)01638-9

Soltis, P. S., and D. E. Soltis, 2000 The role of genetic and genomic attributes in the success of polyploids. Proceedings of the National Academy of Sciences 97: 7051–7057. https://doi.org/10.1073/pnas.97.13.7051

Soneson, C., M. I. Love, and M. D. Robinson, 2015 Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 1; peer review: 2 approved]. F1000Research 4: 1521. https://doi.org/10.12688/f1000research.7563.1

Song, L., and L. Florea, 2015 Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. GigaScience 4: 48. https://doi.org/10.1186/s13742-015-0089-y

Soutourina, J., S. Wydau, Y. Ambroise, C. Boschiero, and M. Werner, 2011 Direct Interaction of RNA Polymerase II and Mediator Required for Transcription in Vivo. Science 331: 1451–1454. https://doi.org/10.1126/science.1200188

Sprenger, H., A. Erban, S. Seddig, K. Rudack, A. Thalhammer, *et al.*, 2018 Metabolite and transcript markers for the prediction of potato drought tolerance. Plant Biotechnology Journal 16: 939–950. https://doi.org/10.1111/pbi.12840

Steel, M., and D. Penny, 2000 Parsimony, Likelihood, and the Role of Models in Molecular Phylogenetics. Molecular Biology and Evolution 17: 839–850. https://doi.org/10.1093/oxfordjournals.molbev.a026364

Steinbach, K., and G. Gottsberger, 1994 Phenology and pollination biology of five *Ranunculus* species in Giessen, Central Germany. Phyton 34: 203–218.

Strauss, S. Y., N. I. Cacho, M. W. Schwartz, A. C. Schwartz, and K. C. Burns, 2015 Apparency revisited. Entomologia Experimentalis et Applicata 157: 74–85. https://doi.org/10.1111/eea.12347

Suarez-Gonzalez, A., C. Lexer, and Q. C. B. Cronk, 2018 Adaptive introgression: A plant perspective. Biology Letters 14: 20170688. https://doi.org/10.1098/rsbl.2017.0688

Suggate, R. P., 1990 Late pliocene and quaternary glaciations of New Zealand. Quaternary Science Reviews 9: 175–197. https://doi.org/10.1016/0277-3791(90)90017-5

Sugita, M., and M. Sugiura, 1996 Regulation of gene expression in chloroplasts of higher plants. Plant Molecular Biology 32: 315–326. https://doi.org/10.1007/BF00039388

Suzuki, S., A. D. Noble, S. Ruaysoongnern, and N. Chinabut, 2007 Improvement in Water-Holding Capacity and Structural Stability of a Sandy Soil in Northeast Thailand. Arid Land Research and Management 21: 37–49. https://doi.org/10.1080/15324980601087430

Swofford, D. L., P. J. Waddell, J. P. Huelsenbeck, P. G. Foster, P. O. Lewis, *et al.*, 2001 Bias in Phylogenetic Estimation and Its Relevance to the Choice between Parsimony and Likelihood Methods. Systematic Biology 50: 525–539. https://doi.org/10.1080/10635150117959

Tait, A., J. Sturman, and M. Clark, 2012 An assessment of the accuracy of interpolated daily rainfall for New Zealand. Journal of Hydrology (New Zealand) 51: 25–44.

Talbi, S., M. C. Romero-Puertas, A. Hernandez, L. Terron, A. Ferchichi, *et al.*, 2015 Drought tolerance in a Saharian plant *Oudneya africana*: Role of antioxidant defences. Environmental and Experimental Botany 111: 114–126. https://doi.org/10.1016/j.envexpbot.2014.11.004

Tang, J., and D. C. Bassham, 2018 Autophagy in crop plants: what's new beyond *Arabidopsis*? Open Biology 8: 180162. https://doi.org/10.1098/rsob.180162

Taseski, G. M., D. A. Keith, R. L. Dalrymple, and W. K. Cornwell, 2021 Shifts in fine root traits within and among species along a fine-scale hydrological gradient. Annals of Botany 127: 473–481. https://doi.org/10.1093/aob/mcaa175

Tate, J. A., P. Joshi, K. A. Soltis, P. S. Soltis, and D. E. Soltis, 2009 On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). BMC Plant Biology 9: 80. https://doi.org/10.1186/1471-2229-9-80

Telfer, E., N. Graham, L. Macdonald, S. Sturrock, P. Wilcox, *et al.*, 2018 Approaches to variant discovery for conifer transcriptome sequencing. PLoS ONE 13: e0205835. https://doi.org/10.1371/journal.pone.0205835

Thakur, S., S. Choudhary, and P. Bhardwaj, 2019 Comparative Transcriptome Profiling Under Cadmium Stress Reveals the Uptake and Tolerance Mechanism in *Brassica juncea*. Journal of Plant Growth Regulation 38: 1141–1152. https://doi.org/10.1007/s00344-019-09919-8

Thomas, B. C., B. Pedersen, and M. Freeling, 2006 Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Research 16: 934–946. https://doi.org/10.1101/gr.4708406

Tian, T., Y. Liu, H. Yan, Q. You, X. Yi, *et al.*, 2017 agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Research 45: W122–W129. https://doi.org/10.1093/nar/gkx382

Tigano, A., and V. L. Friesen, 2016 Genomics of local adaptation with gene flow. Molecular Ecology 25: 2144–2164. https://doi.org/10.1111/mec.13606

Tiwari, J. K., T. Buckseth, R. Zinta, A. Saraswati, R. K. Singh, *et al.*, 2020 Transcriptome analysis of potato shoots, roots and stolons under nitrogen stress. Scientific Reports 10: 1152. https://doi.org/10.1038/s41598-020-58167-4

Todd, E. V, M. A. Black, and N. J. Gemmell, 2016 The power and promise of RNA-seq in ecology and evolution. Molecular Ecology 25: 1224–1241. https://doi.org/10.1111/mec.13526

Turner, M. E., J. C. Stephens, and W. W. Anderson, 1982 Homozygosity and patch structure in plant populations as a result of nearest-neighbor pollination. Proceedings of the National Academy of Sciences 79: 203–207. https://doi.org/10.1073/pnas.79.1.203

Udall, J. A., and J. F. Wendel, 2006 Polyploidy and Crop Improvement. Crop Science 46: S3–S14. https://doi.org/10.2135/cropsci2006.07.0489tpg

Undin, M., P. J. Lockhart, S. F. K. Hills, and I. Castro, 2021 Genetic rescue and the plight of Ponui hybrids. Frontiers in Conservation Science 1: 8. https://doi.org/10.3389/fcosc.2020.622191

UniProt Consortium, 2018 UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research 47: D506–D515. https://doi.org/10.1093/nar/gky1049

Vennapusa, A. R., I. M. Somayanda, C. J. Doherty, and S. V. K. Jagadish, 2020 A universal method for high-quality RNA extraction from plant tissues rich in starch, proteins and fiber. Scientific Reports 10: 16887. https://doi.org/10.1038/s41598-020-73958-5

Verheijen, F. G. A., A. Zhuravel, F. C. Silva, A. Amaro, M. Ben-Hur, *et al.*, 2019 The influence of biochar particle size and concentration on bulk density and maximum water holding capacity of sandy vs sandy loam soil in a column experiment. Geoderma 347: 194–202. https://doi.org/10.1016/j.geoderma.2019.03.044

Voelckel, C., P. B. Heenan, B. Janssen, M. Reichelt, K. Ford, *et al.*, 2008 Transcriptional and biochemical signatures of divergence in natural populations of two species of New Zealand alpine *Pachycladon*. Molecular Ecology 17: 4740–4753. https://doi.org/10.1111/j.1365-294X.2008.03933.x

Voelckel, C., M. Mirzaei, M. Reichelt, Z. Luo, D. Pascovici, *et al.*, 2010 Transcript and protein profiling identify candidate gene sets of potential adaptive significance in New Zealand *Pachycladon*. BMC Evolutionary Biology 10: 151. https://doi.org/10.1186/1471-2148-10-151

Voelckel, C., N. Gruenheit, P. Biggs, O. Deusch, and P. J. Lockhart, 2012 Chips and tags suggest plant-environment interactions differ for two alpine *Pachycladon* species. BMC Genomics 13: 322. https://doi.org/10.1186/1471-2164-13-322

Voelckel, C., N. Gruenheit, and P. J. Lockhart, 2017 Evolutionary Transcriptomics and Proteomics: Insight into Plant Adaptation. Trends in Plant Science 22: 462–471. https://doi.org/10.1016/j.tplants.2017.03.001

Vollmer, S. V, and S. R. Palumbi, 2004 Testing the utility of internally transcribed spacer sequences in coral phylogenetics. Molecular Ecology 13: 2763–2772. https://doi.org/10.1111/j.1365-294X.2004.02265.x

Wagner, G. P., K. Kin, and V. J. Lynch, 2012 Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory in Biosciences 131: 281–285. https://doi.org/10.1007/s12064-012-0162-3

Wagstaff, S. J., M. J. Bayly, P. J. Garnock-Jones, and D. C. Albach, 2002 Classification, origin, and diversification of the New Zealand hebes (Scrophulariaceae). Annals of the Missouri Botanical Garden 89: 38–63. https://doi.org/10.2307/3298656

Wang, G., G. Cai, F. Kong, Y. Deng, N. Ma, *et al.*, 2014 Overexpression of tomato chloroplast-targeted DnaJ protein enhances tolerance to drought stress and resistance to *Pseudomonas solanacearum* in transgenic tobacco. Plant Physiology and Biochemistry 82: 95–104. https://doi.org/10.1016/j.plaphy.2014.05.011

Wang, K., W. Hong, H. Jiao, and H. Zhao, 2017 Transcriptome sequencing and phylogenetic analysis of four species of luminescent beetles. Scientific Reports 7: 1814. https://doi.org/10.1038/s41598-017-01835-9

Wang, G., G. Cai, N. Xu, L. Zhang, X. Sun, *et al.*, 2019a Novel DnaJ Protein Facilitates Thermotolerance of Transgenic Tomatoes. International Journal of Molecular Sciences 20: 367. https://doi.org/10.3390/ijms20020367

Wang, Y., M. Q. Shahid, F. Ghouri, S. Ercişli, F. S. Baloch, *et al.*, 2019b Transcriptome analysis and annotation: SNPs identified from single copy annotated unigenes of three polyploid blueberry crops. PLoS ONE 14: e0216299. https://doi.org/10.1371/journal.pone.0216299

Wang, Y., B. Liu, Z. Nie, H. Chen, F. Chen, *et al.*, 2020 Major clades and a revised classification of *Magnolia* and Magnoliaceae based on whole plastid genome sequences via genome skimming. Journal of Systematics and Evolution 58: 673–695. https://doi.org/10.1111/jse.12588

Ward, Jr J. H., 1963 Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58: 236–244.

Warnes, M. G. R., B. Bolker, L. Bonebakker, R. Gentleman, and W. Huber, 2016 *Package 'gplots.' Various R programming tools for plotting data*.

Watermann, L. Y., M. Hock, C. Blake, and A. Erfmeier, 2020 Plant invasion into high elevations implies adaptation to high UV-B environments: a multi-species experiment. Biological Invasions 22: 1203–1218. https://doi.org/10.1007/s10530-019-02173-9

Watson, M., and A. Warr, 2019 Errors in long-read assemblies can critically affect protein prediction. Nature Biotechnology 37: 124–126. https://doi.org/10.1038/s41587-018-0004-z

Webb, C. J., W. R. Sykes, P. J. Garnock-Jones, and D. R. Given, 1988 *Flora of New Zealand. Volume IV, Naturalised pteridophytes, gymnosperms, dicotyledons*. Botany Division, Department of Scientific and Industrial Research, Christchurch, New Zealand.

Webb, T. H., 1995 *A manual of land characteristics for evaluation of rural land*. Manaaki Whenua Press, Lincoln, Canterbury, New Zealand.

Wettewa, E., and L. E. Wallace, 2021 Molecular phylogeny and ancestral biogeographic reconstruction of *Platanthera* subgenus *Limnorchis* (Orchidaceae) using target capture methods. Molecular Phylogenetics and Evolution 157: 107070. https://doi.org/10.1016/j.ympev.2021.107070

White, T. J., T. Bruns, S. Lee, and J. Taylor, 1990 *Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics*. Academic Press, San Diego.

Whitney, K. D., K. W. Broman, N. C. Kane, S. M. Hovick, R. A. Randell, *et al.*, 2015 Quantitative trait locus mapping identifies candidate alleles involved in adaptive introgression and range expansion in a wild sunflower. Molecular Ecology 24: 2194–2211. https://doi.org/10.1111/mec.13044

Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, *et al.*, 2014 Phylotranscriptomic analysis of the origin and early diversification of land plants. Proceedings of the National Academy of Sciences 111: E4859–E4868. https://doi.org/10.1073/pnas.1323926111

Wickham, H., 2011 ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics 3: 180–185. https://doi.org/10.1002/wics.147

Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. McGowan, *et al.*, 2019 Welcome to the Tidyverse. Journal of Open Source Software 4: 1686. https://doi.org/10.21105/joss.01686

Winkworth, R. C., A. W. Robertson, F. Ehrendorfer, and P. J. Lockhart, 1999 The importance of dispersal and recent speciation in the flora of New Zealand. Journal of Biogeography 26: 1323–1325. https://doi.org/10.1046/j.1365-2699.1999.00392.x

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 and Evolution 5: 237–247. https://doi.org/10.1016/j.ode.2004.12.001

Wolfe, K. H., 2001 Yesterday's polyploids and the mystery of diploidization. Nature Reviews Genetics 2: 333–341. https://doi.org/10.1038/35072009

Xiao, L.-Q., M. Möller, and H. Zhu, 2010 High nrDNA ITS polymorphism in the ancient extant seed plant *Cycas*: Incomplete concerted evolution and the origin of pseudogenes. Molecular Phylogenetics and Evolution 55: 168–177. https://doi.org/10.1016/j.ympev.2009.11.020

Xu, Y., W. Zong, X. Hou, J. Yao, H. Liu, *et al.*, 2015 OsARID 3, an AT-rich Interaction Domain-containing protein, is required for shoot meristem development in rice. The Plant Journal 83: 806–817. https://doi.org/10.1111/tpj.12927

Xu, B., X.-M. Zeng, X.-F. Gao, D.-P. Jin, and L.-B. Zhang, 2017 ITS non-concerted evolution and rampant hybridization in the legume genus *Lespedeza* (Fabaceae). Scientific Reports 7: 1–15. https://doi.org/10.1038/srep40057

Xu, Y., and B. Huang, 2018 Comparative transcriptomic analysis reveals common molecular factors responsive to heat and drought stress in *Agrostis stolonifera*. Scientific Reports 8: 15181. https://doi.org/10.1038/s41598-018-33597-3

Yang, Y., and S. A. Smith, 2013 Optimizing *de novo* assembly of short-read RNA-seq data for phylogenomics. BMC Genomics 14: 328. https://doi.org/10.1186/1471-2164-14-328

Yang, T., J. Cao, Y. Wang, and Y. Liu, 2017a Soil moisture influences vegetation distribution patterns in sand dunes of the Horqin Sandy Land, Northeast China. Ecological Engineering 105: 95–101. https://doi.org/10.1016/j.ecoleng.2017.04.035

Yang, X., J. Song, Q. You, D. R. Paudel, J. Zhang, *et al.*, 2017b Mining sequence variations in representative polyploid sugarcane germplasm accessions. BMC Genomics 18: 594. https://doi.org/10.1186/s12864-017-3980-3

Yang, L., Y. Jin, W. Huang, Q. Sun, F. Liu, *et al.*, 2018a Full-length transcriptome sequences of ephemeral plant *Arabidopsis pumila* provides insight into gene expression dynamics during continuous salt stress. BMC Genomics 19: 717.

Yang, Y., T. Ma, Z. Wang, Z. Lu, Y. Li, *et al.*, 2018b Genomic effects of population collapse in a critically endangered ironwood tree *Ostrya rehderiana*. Nature Communications 9: 5449. https://doi.org/10.1038/s41467-018-07913-4

Yao, Z., F. M. You, A. N'Diaye, R. E. Knox, C. McCartney, *et al.*, 2020 Evaluation of variant calling tools for large plant genome re-sequencing. BMC Bioinformatics 21: 360.

https://doi.org/10.1186/s12859-020-03704-1

Ye, G., Y. Ma, Z. Feng, and X. Zhang, 2018 Transcriptomic analysis of drought stress responses of sea buckthorn (*Hippophae rhamnoides* subsp. *sinensis*) by RNA-Seq. PLoS ONE 13: e0202213. https://doi.org/10.1371/journal.pone.0202213

Zhang, Y., M. Fischer, V. Colot, and O. Bossdorf, 2013 Epigenetic variation creates potential for evolution of plant phenotypic plasticity. New Phytologist 197: 314–322. https://doi.org/10.1111/nph.12010

Zhang, Z. F., Y. Y. Li, and B. Z. Xiao, 2016 Comparative transcriptome analysis highlights the crucial roles of photosynthetic system in drought stress adaptation in upland rice. Scientific Reports 6: 19349. https://doi.org/10.1038/srep19349

Zhang, Q.-P., J. Wang, H.-L. Gu, Z.-G. Zhang, and Q. Wang, 2018 Effects of Continuous Slope Gradient on the Dominance Characteristics of Plant Functional Groups and Plant Diversity in Alpine Meadows. Sustainability 10: 4805. https://doi.org/10.3390/su10124805

Zhang, H., C. Jain, and S. Aluru, 2020a A comprehensive evaluation of long read error correction methods. BMC Genomics 21: 889. https://doi.org/10.1186/s12864-020-07227-0

Zhang, X., Y. Yao, X. Li, L. Zhang, and S. Fan, 2020b Transcriptomic analysis identifies novel genes and pathways for salt stress responses in *Suaeda salsa* leaves. Scientific Reports 10: 4236. https://doi.org/10.1038/s41598-020-61204-x

Zhao, Q.-Y., Y. Wang, Y.-M. Kong, D. Luo, X. Li, *et al.*, 2011 Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics 12: S2. https://doi.org/10.1186/1471-2105-12-S14-S2

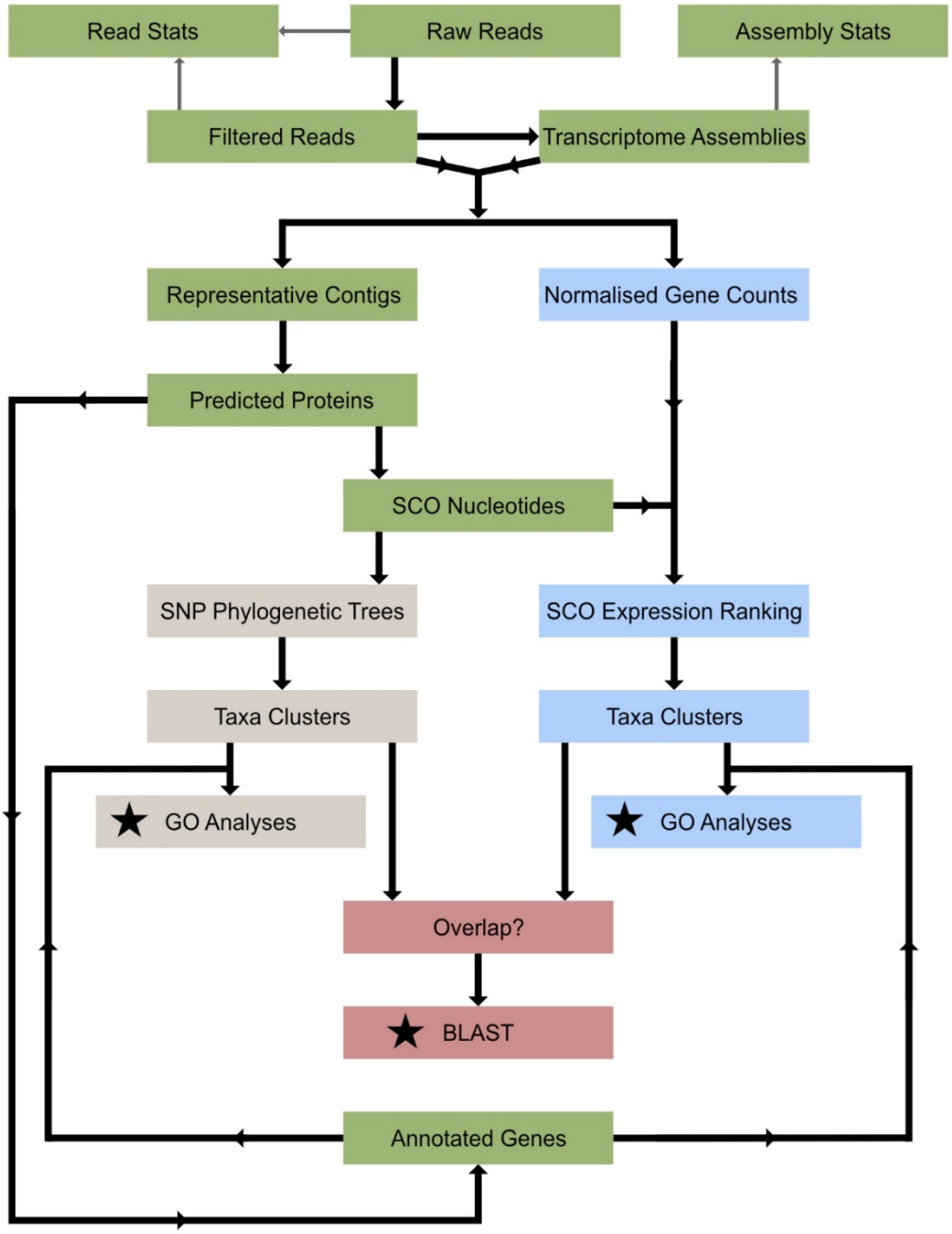
Zhao, S.-Y., L.-Y. Chen, Y.-L. Wei, Q.-F. Wang, and M. L. Moody, 2016 RNA-seq of *Ranunculus sceleratus* and Identification of Orthologous Genes among Four *Ranunculus* Species. Frontiers in Plant Science 7: 732. https://doi.org/10.3389/fpls.2016.00732

Zhao, Y., K. Wang, W. Wang, T. Yin, W. Dong, *et al.*, 2019 A high-throughput SNP discovery strategy for RNA-seq data. BMC Genomics 20: 160. https://doi.org/10.1186/s12864-019-5533-4

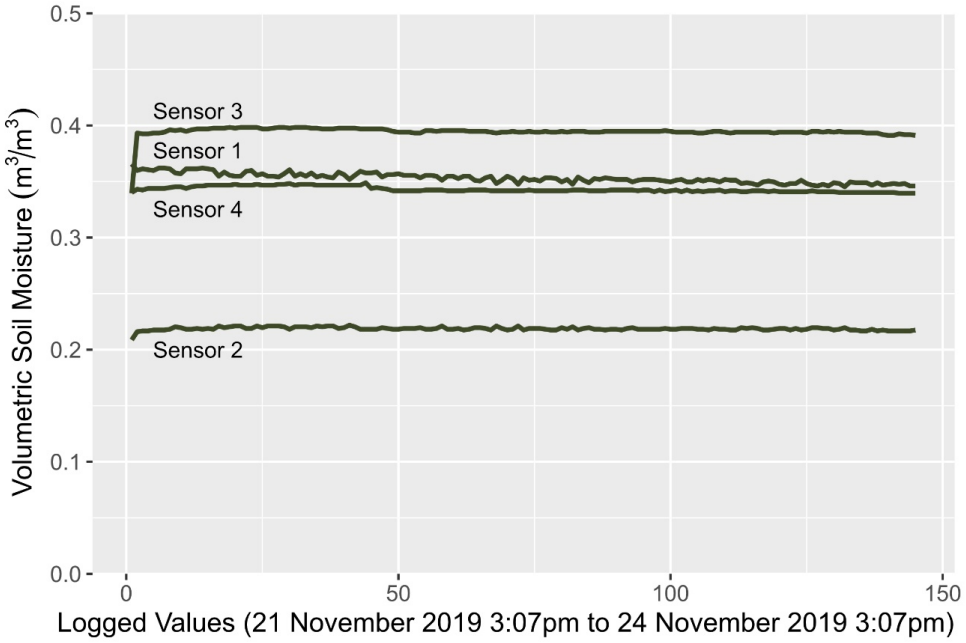
Zhao, S., Z. Ye, and R. Stanton, 2020 Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. RNA 26: 903–909. https://doi.org/10.1261/rna.074922.120

Zhou, K., B. B. Liu, Y. L. Wang, X. Q. Zhang, and G. L. Sun, 2019 Evolutionary mechanism of genome duplication enhancing natural autotetraploid sea barley adaptability to drought stress. Environmental and Experimental Botany 159: 44–54. https://doi.org/10.1016/j.envexpbot.2018.12.005

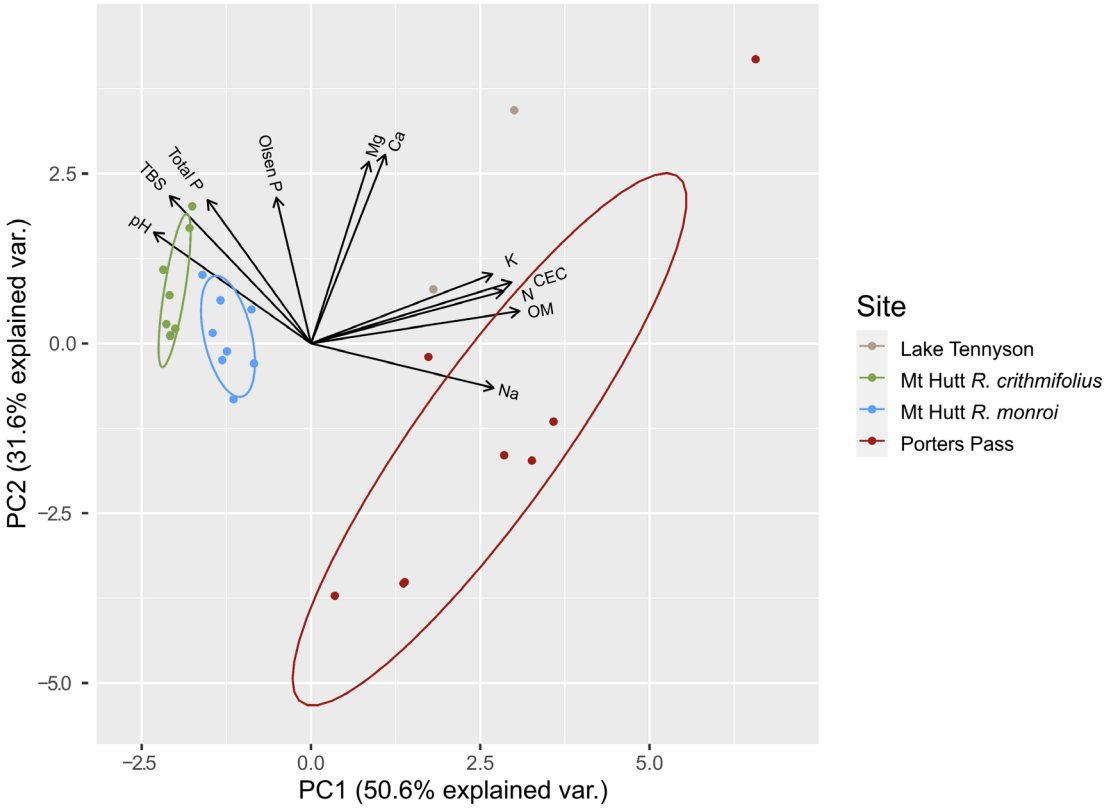
Appendix I: Figures



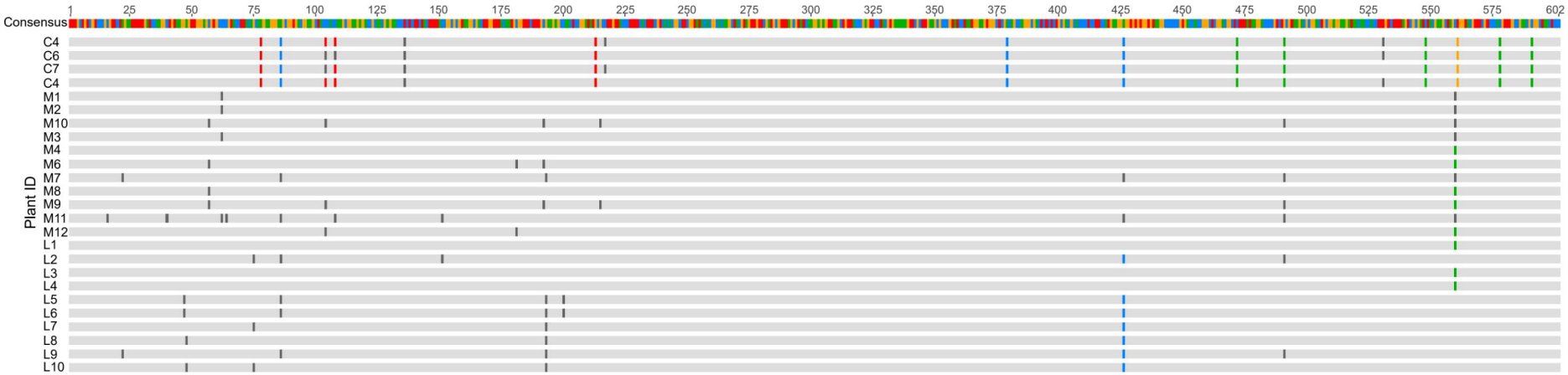
Supplementary Figure 1 Bioinformatic pipeline for aggregated assembly SCO analyses. This figure illustrates the steps taken, from raw sequence data to final outputs, in this study to analyse single-copy orthologues (SCOs). Green boxes indicate steps common to all analyses. Brown boxes indicate steps taken for nucleotide variation analyses. Blue boxes indicate steps taken for gene expression analyses. Red boxes indicate steps taken for finding and identifying genes that occur in the taxonomic split for both gene expression and nucleotide variation analyses. Stars indicate final outputs; either gene ontology (GO) term enrichment analyses or basic local alignment search tool (BLAST) outputs from the National Center for Biotechnology Information (NCBI) website.



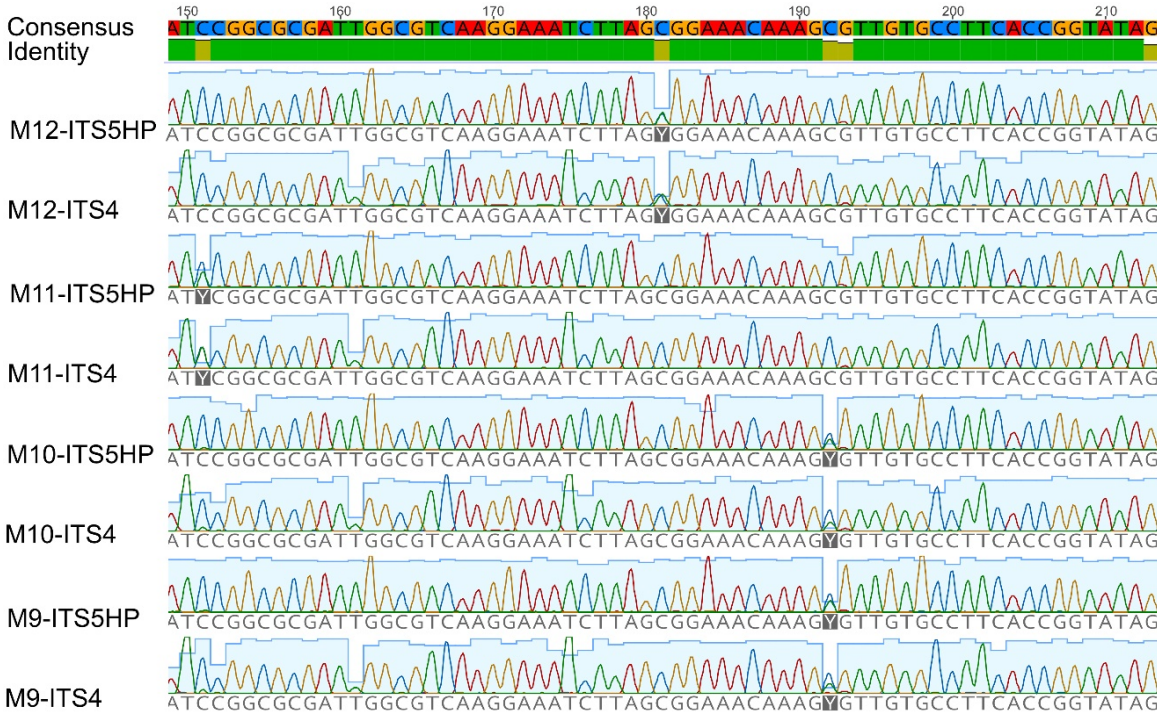
Supplementary Figure 2 Soil moisture measurements at Porters Pass prior to data logger failure. This plot shows three days of volumetric soil moisture readings for the four moisture sensors. Data logger failure occurred after exactly three days of measurement.



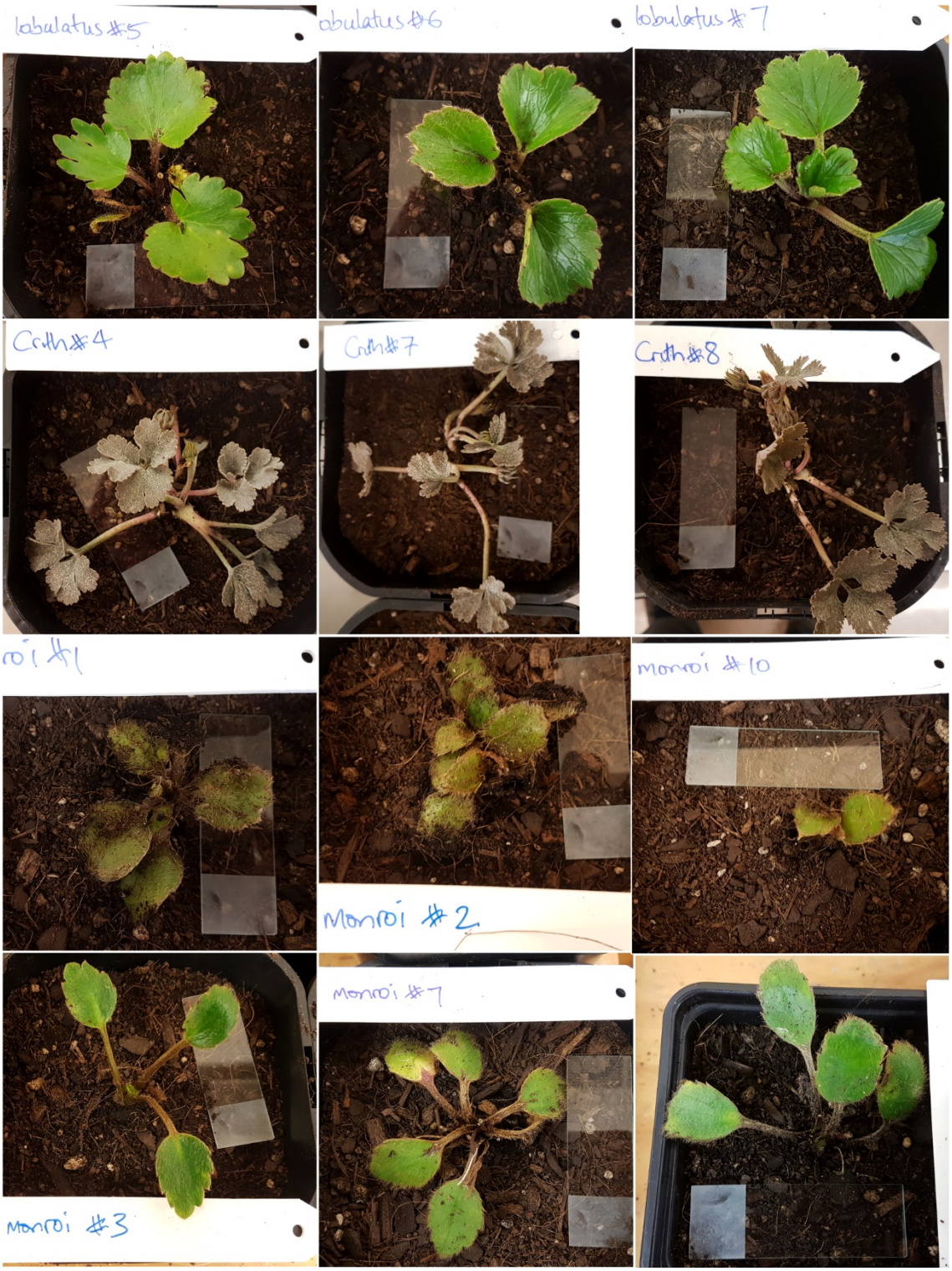
Supplementary Figure 3 Principal coordinate analysis of sample site soil chemistry with all samples retained. Dots represent soil samples. Vector directions and lengths indicate variable loadings for each axis. Each ellipse represents the 68% confidence interval of the group core.



Supplementary Figure 4 Alignment of ITS2 sequences for plants used in this study. Light grey bars indicate agreement with the consensus sequence of the multiple sequence alignment (MSA) of the internal transcribed spacer 2 (ITS2) region. Coloured blocks represent polymorphisms: red = A, blue = C, orange = G, green = T. Dark grey blocks represent ambiguous polymorphisms that have been assigned one of the International Union of Pure and Applied Chemistry (IUPAC) nucleotide codes.

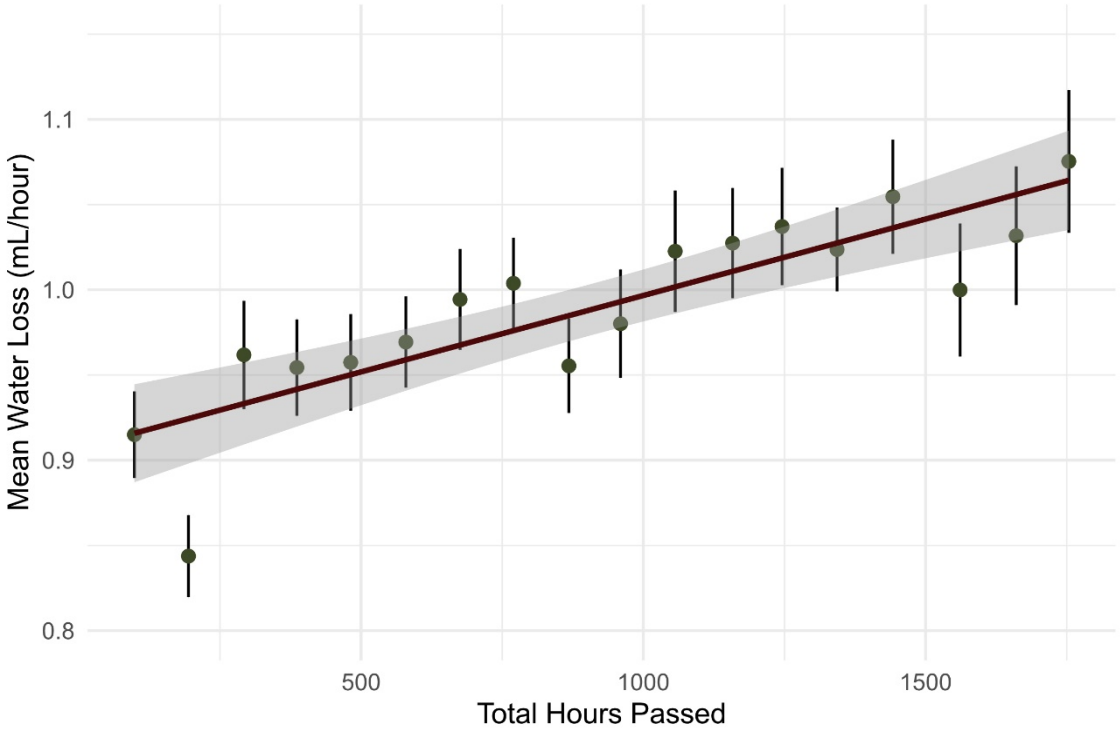


Supplementary Figure 5 Close-up section of the ITS2 multiple sequence alignment. This plot includes chromatograms output from the Applied Biosystems™ 3730 Genetic Analyzer of a partial internal transcribed spacer 2 (ITS2) multiple sequence alignment (MSA). Sequence names refer to the plant ID as well as primers used for PCR and Sanger sequencing. At sites deemed ambiguous (Y), the sites are ambiguous in both sequencing directions. Additionally, chromatograms indicate the presence of two alleles at each ambiguous site.

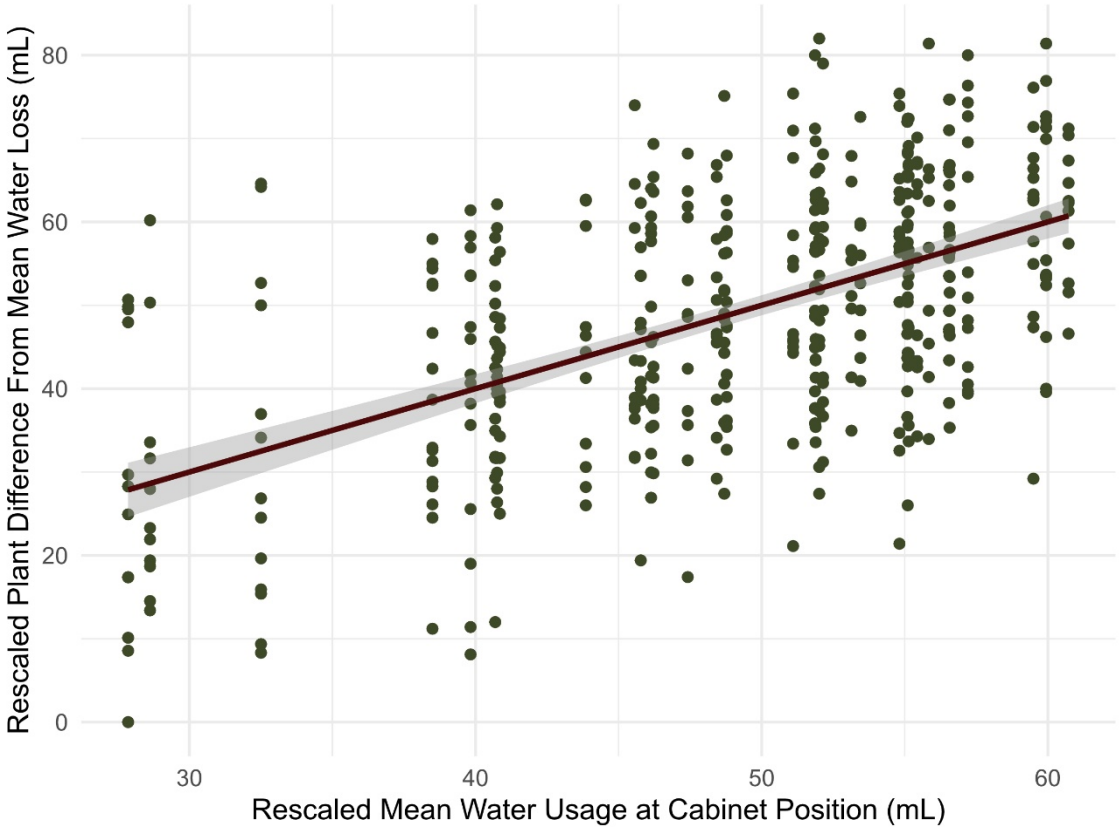


Supplementary Figure 6 Plants, used in the RNA-seq analyses, at the time of potting. From left to right, top to bottom, plant ID: L5, L6, L7, C4 C7, C8, M1, M2, M10, M3, M7, M12. Date: 29 November 2019.

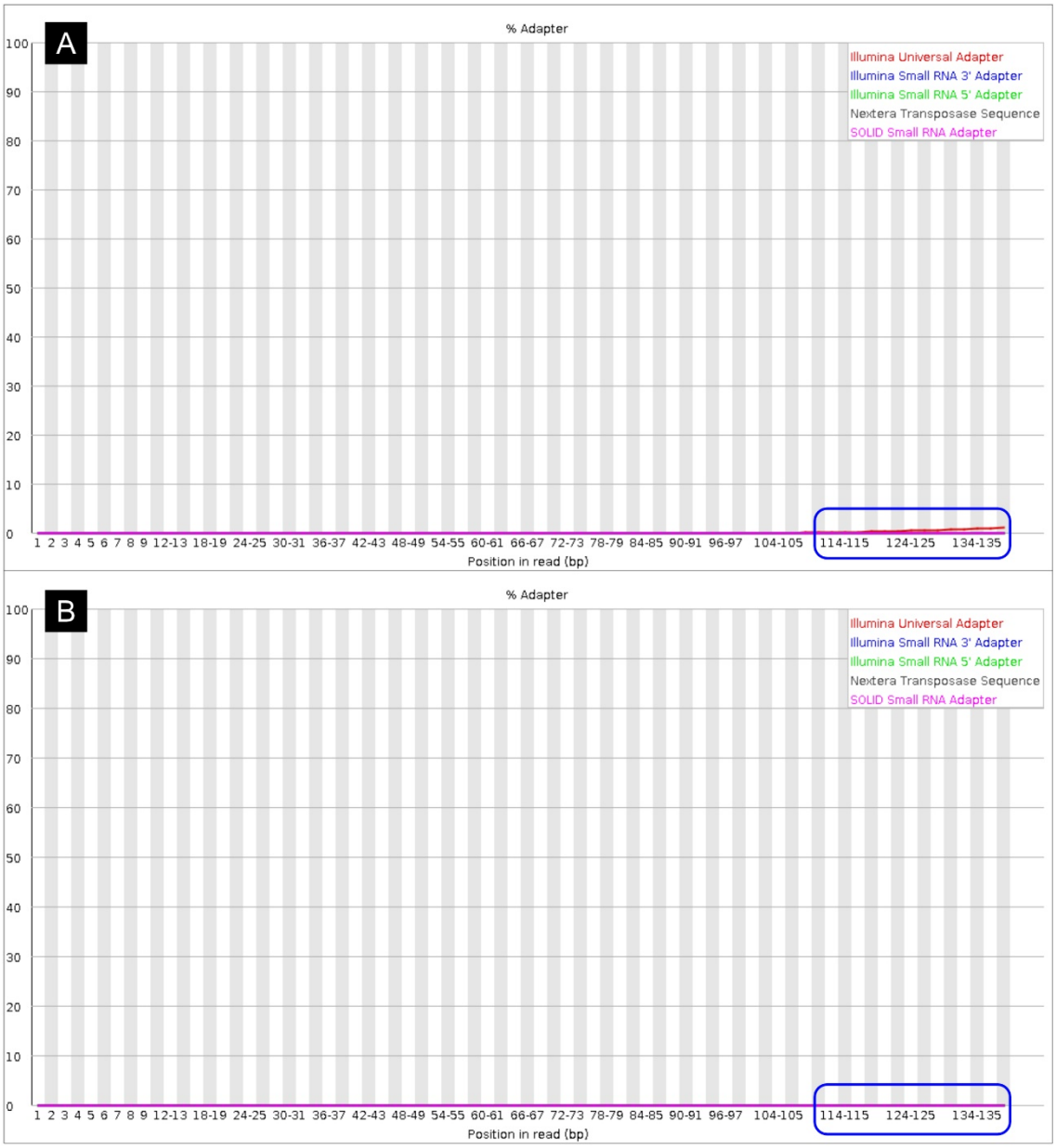
Supplementary Figure 7 Plants, used in the RNA-seq analyses, two days prior to harvest. From left to right, top to bottom, plant ID: L5, L6, L7, C4 C7, C8, M1, M2, M10, M3, M7, M12. Date: 25 February 2020.



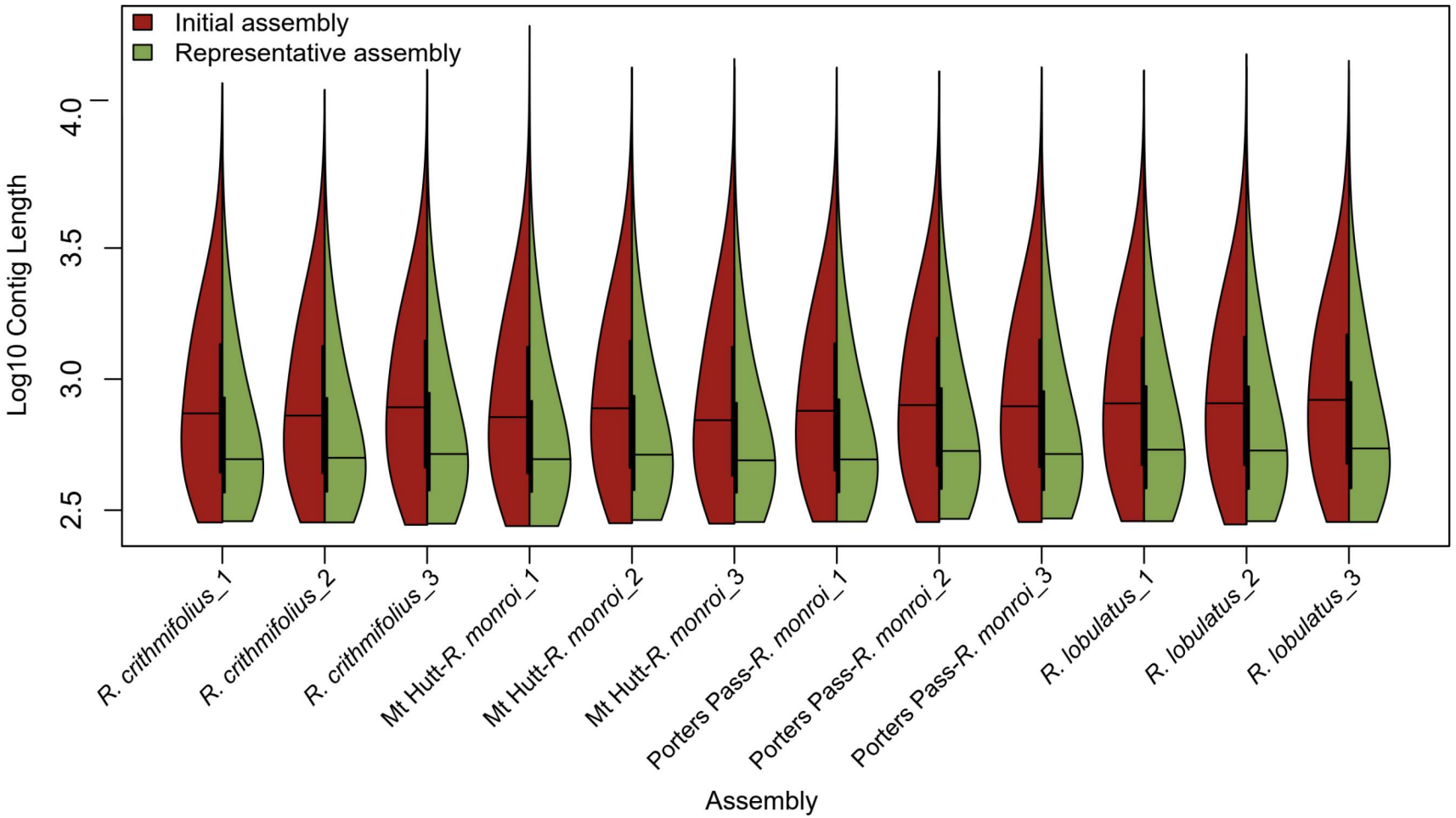
Supplementary Figure 8 Increasing water usage by plants over the course of the growth experiment. Water loss was calculated by measurement of overall plant weight at each watering. Data points of mean water loss were calculated by dividing total mass of water lost by hours since the last watering. Vertical bars indicate standard error (*SE).* The regression line shows a linear model of mean values including the 95% confidence interval (CI) for the predictive value of the model. Increases in plant mass are not accounted for.



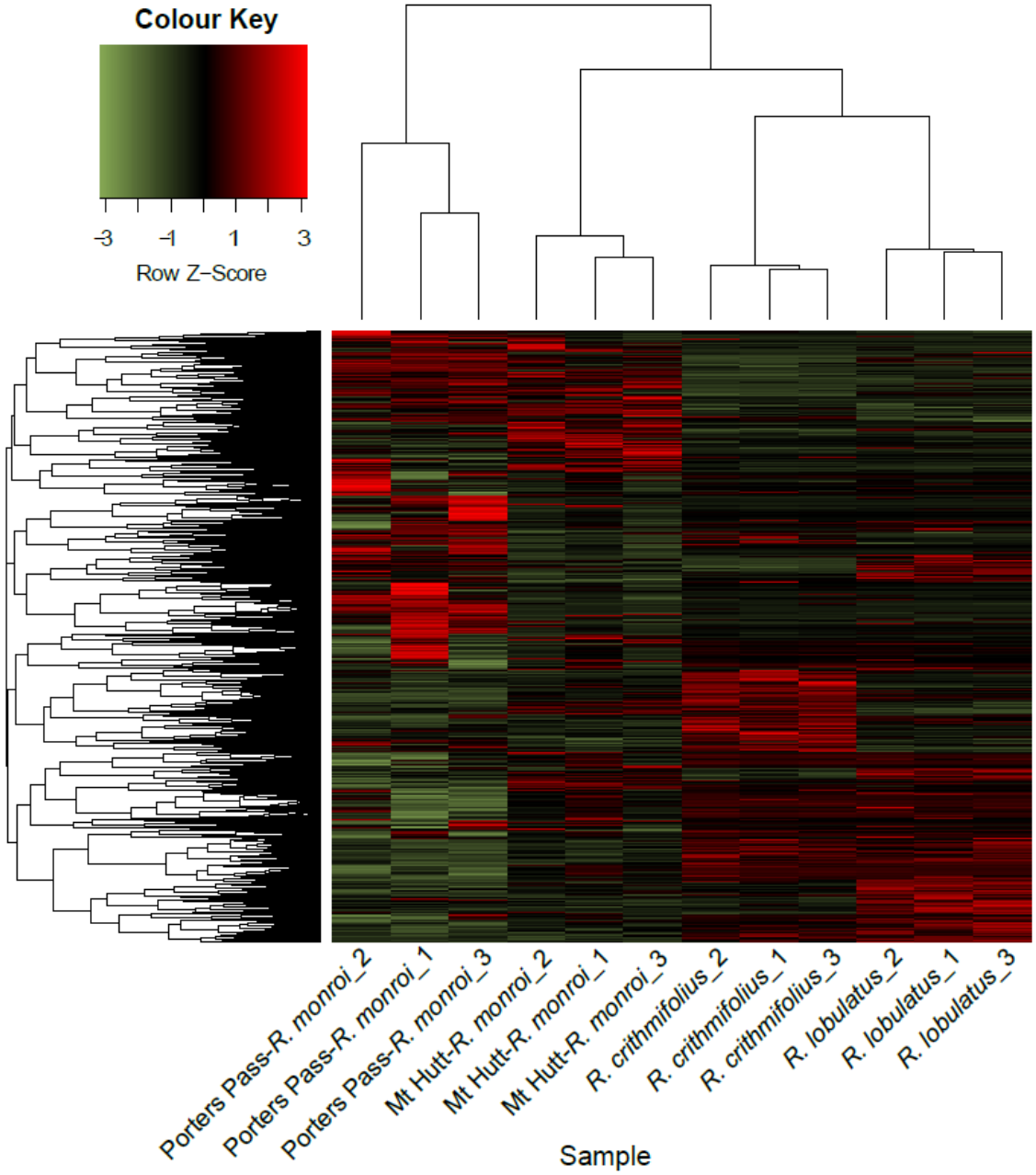
Supplementary Figure 9 Cabinet position influenced plant water usage. To account for variation in watering timing and increasing usage over time, the difference of each plant's water use was subtracted from the mean water use at that time. The results were rescaled so the lowest water use = 0 mL. Mean water use, of these rescaled values, was calculated for each cabinet position and the plot ordered along the x axis from lowest mean use to highest. The water rescaled water-use values (y axis), for each watering period, are plotted for each cabinet position (x axis). The regression line shows a linear model of mean water use including the 95% confidence interval (CI) for the predictive value of the model.



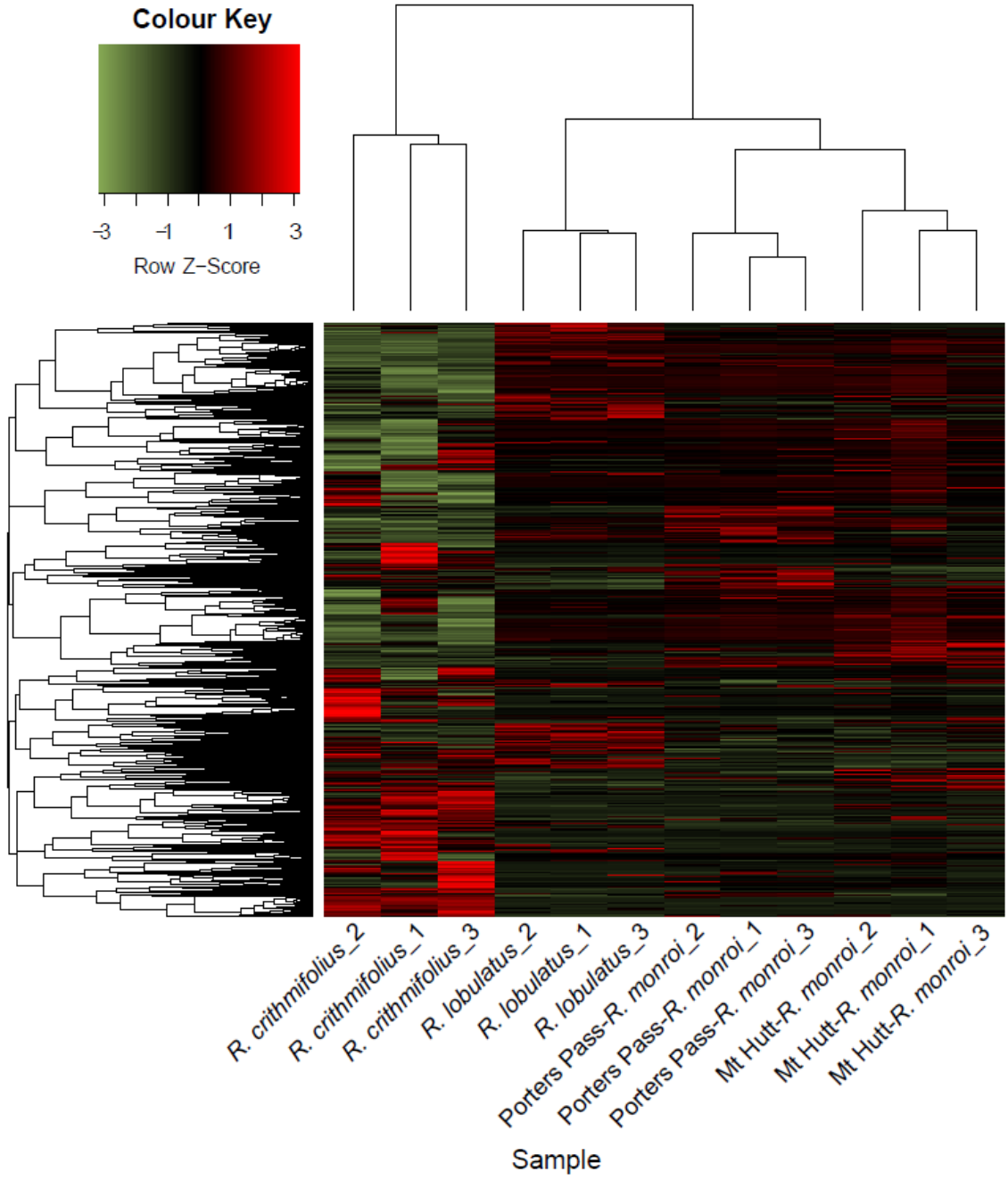
Supplementary Figure 10 Adaptor content in RNA-seq reads before and after trimming and quality filtering. A, FastQC report of adapter contamination for *R. crithmifolius*_1 before trimming. B, FastQC report of adapter contamination for *R. crithmifolius*_1 after trimming. In A, a small amount of adapter contamination is visible at the end of the reads (highlighted in blue) but that contamination is no longer evident in B.



Supplementary Figure 11 Individual transcriptome assembly statistics. The violin plots show the distribution of log10-normalised contig lengths for initial and representative individual assemblies. Median contig length is represented by horizontal bars, and interquartile range by thick vertical bands. Widths of the plots indicate the proportion of contigs at that length.



Supplementary Figure 12 Gene expression analysis by alignment of reads against the Porters Pass-*R. monroi* aggregated assembly. The heatmap shows row-scaled log10-transformed transcripts per million (TPM) normalised values for 16,498 genes in which at least one taxon was significantly differentially expressed. Red represents upregulation of genes relative to the other samples, and green represents downregulation. Each row represents one gene. Columns represent sample RNA-seq reads, and the weighted dendrogram generated using hierarchical clustering with Spearman's rho.



Supplementary Figure 13 Gene expression analysis by alignment of reads against the *R. crithmifolius* aggregated assembly. The heatmap shows row-scaled log10-transformed transcripts per million (TPM) values for 19,932 genes in which at least one taxon was significantly differentially expressed. Red represents upregulation of genes relative to the other samples, and green represents downregulation. Each row represents one gene. Columns represent sample RNA-seq reads, and the weighted dendrogram generated using hierarchical clustering with Spearman's rho.

Appendix II: Tables

Supplementary Table 1 Sample site locations		
Site	Latitude[a]	Longitude
Lake Tennyson	-42.20487	172.73891
Mount Hutt	-43.51657	171.54289
Porters Pass	-43.29531	171.74416

[a] Coordinates given in decimal degrees using the NZGD2000 geodetic datum

Supplementary Table 2 Soil particle size statistics

Particle Sizes (mm)	Site[a]	Depth (cm)	*M*	*SD*	*SE*	95 % CI	Particle Sizes (mm)	Site	Depth (cm)	*M*	*SD*	*SE*	95 % CI
16–8	HC	0–5	0.14	0.05	0.02	0.06	16–8	PP	0–5	0.00	0.00	0.00	0.00
16–8	HC	5–10	0.21	0.07	0.03	0.09	16–8	PP	5–10	0.07	0.08	0.04	0.10
8–5.6	HC	0–5	0.06	0.02	0.01	0.03	8–5.6	PP	0–5	0.00	0.00	0.00	0.01
8–5.6	HC	5–10	0.08	0.02	0.01	0.02	8–5.6	PP	5–10	0.03	0.04	0.02	0.05
5.6–4	HC	0–5	0.07	0.02	0.01	0.02	5.6–4	PP	0–5	0.00	0.01	0.00	0.01
5.6–4	HC	5–10	0.08	0.01	0.01	0.01	5.6–4	PP	5–10	0.02	0.03	0.01	0.03
4–2	HC	0–5	0.20	0.05	0.02	0.06	4–2	PP	0–5	0.01	0.01	0.00	0.01
4–2	HC	5–10	0.18	0.03	0.01	0.04	4–2	PP	5–10	0.04	0.03	0.01	0.04
2–1	HC	0–5	0.16	0.04	0.02	0.04	2–1	PP	0–5	0.01	0.01	0.00	0.01
2–1	HC	5–10	0.15	0.03	0.02	0.04	2–1	PP	5–10	0.02	0.01	0.01	0.01
1–0.5	HC	0–5	0.10	0.02	0.01	0.02	1–0.5	PP	0–5	0.02	0.02	0.01	0.02
1–0.5	HC	5–10	0.09	0.02	0.01	0.03	1–0.5	PP	5–10	0.03	0.01	0.00	0.01
0.5–0.25	HC	0–5	0.05	0.01	0.00	0.01	0.5–0.25	PP	0–5	0.09	0.05	0.02	0.06
0.5–0.25	HC	5–10	0.04	0.01	0.00	0.01	0.5–0.25	PP	5–10	0.05	0.02	0.01	0.02
< 0.25	HM	0–5	0.21	0.05	0.02	0.06	< 0.25	PP	0–5	0.86	0.05	0.02	0.07
< 0.25	HM	5–10	0.16	0.06	0.03	0.08	< 0.25	PP	5–10	0.72	0.18	0.08	0.22
16–8	HM	0–5	0.05	0.05	0.02	0.06	16–8	LT	0–5	0.43	0.27	0.12	0.33
16–8	HM	5–10	0.24	0.20	0.09	0.25	16–8	LT	5–10	0.59	0.17	0.08	0.22
8–5.6	HM	0–5	0.03	0.02	0.01	0.02	8–5.6	LT	0–5	0.13	0.07	0.03	0.08
8–5.6	HM	5–10	0.07	0.03	0.01	0.03	8–5.6	LT	5–10	0.17	0.06	0.03	0.08
5.6–4	HM	0–5	0.05	0.02	0.01	0.02	5.6–4	LT	0–5	0.04	0.03	0.01	0.04
5.6–4	HM	5–10	0.07	0.02	0.01	0.02	5.6–4	LT	5–10	0.07	0.05	0.02	0.06
4–2	HM	0–5	0.18	0.04	0.02	0.05	4–2	LT	0–5	0.04	0.02	0.01	0.03
4–2	HM	5–10	0.14	0.03	0.01	0.04	4–2	LT	5–10	0.06	0.04	0.02	0.05
2–1	HM	0–5	0.17	0.03	0.01	0.03	2–1	LT	0–5	0.10	0.07	0.03	0.09
2–1	HM	5–10	0.12	0.03	0.01	0.04	2–1	LT	5–10	0.03	0.02	0.01	0.02
1–0.5	HM	0–5	0.18	0.03	0.02	0.04	1–0.5	LT	0–5	0.11	0.09	0.04	0.11
1–0.5	HM	5–10	0.13	0.05	0.02	0.06	1–0.5	LT	5–10	0.02	0.02	0.01	0.03
0.5–0.25	HM	0–5	0.12	0.03	0.01	0.03	0.5–0.25	LT	0–5	0.07	0.08	0.04	0.10
0.5–0.25	HM	5–10	0.08	0.03	0.01	0.04	0.5–0.25	LT	5–10	0.02	0.03	0.01	0.04
< 0.25	HM	0–5	0.20	0.05	0.02	0.06	< 0.25	LT	0–5	0.07	0.08	0.04	0.10
< 0.25	HM	5–10	0.15	0.05	0.02	0.07	< 0.25	LT	5–10	0.03	0.06	0.03	0.07

[a] HC = Mt Hutt *R. crithmifolius* site, HM = Mt Hutt *R. monroi* site,

PP = Porters Pass, LT = Lake Tennyson

Supplementary Table 3 Soil chemical principal component variable loadings with all samples

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
pH	-0.31244567	0.27820784	-0.29477615	-0.31184244	0.372791909	-0.11130038	-0.3142464	0.42445998	-0.41300998	0.02603612	0.19825651
Olsen P	-0.06873667	0.36586274	0.56410678	-0.57980852	-0.10177266	-0.10519325	0.394123268	0.04060179	-0.03681183	0.1166185	-0.11689915
OM	0.41499902	0.08118782	0.06240613	0.03475813	-0.09873649	-0.02870793	-0.3873614	-0.05320282	-0.26267823	0.70945937	-0.28336691
N	0.3834379	0.13127874	0.31166548	0.069827077	0.23440148	0.22916146	-0.1194783	0.28193883	0.36053256	0.11066083	0.62472149
Total P	-0.20586425	0.36017189	0.34276411	0.568628854	-0.00624688	-0.55064776	-0.18190674	0.14057299	0.08997861	-0.10201253	-0.11164494
K	0.36109138	0.17396783	-0.07134844	0.006685148	0.771010355	-0.0970249	0.173267715	-0.27717902	0.03581321	-0.15906402	-0.31130091
Ca	0.14781583	0.47392151	-0.19580477	-0.2116263	-0.26933815	0.25448954	-0.40804605	0.04639855	0.37908593	-0.34614502	-0.31770868
Mg	0.11519912	0.455709	-0.3126215	0.378974823	-0.1540554	0.28557951	0.552518789	0.26429468	-0.20753855	0.11832635	-0.01461388
Na	0.36355896	-0.11194474	-0.39365977	-0.21710965	-0.180152	-0.65305562	0.195614274	0.26804296	0.27464529	0.05314777	0.07671746
CEC	0.39920522	0.15368256	0.07180374	-0.01799537	-0.25013231	-0.15559269	-0.09142184	-0.32876728	-0.54261549	-0.42937206	0.35989468
TBS	-0.28102813	0.3694705	-0.27402067	-0.03522671	-0.01351153	-0.11648774	0.008881397	-0.62236384	0.25112812	0.33679868	0.36127584

Supplementary Table 4 Soil chemical principal component variable loadings with Lake Tennyson samples and Porters Pass outlier excluded

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
pH	-0.3363331	-0.13683316	-0.08124297	-0.42829763	0.32150777	-0.06124359	0.14507394	-0.37096232	0.635377427	0.09201738	-0.01297231
Olsen P	-0.2779055	-0.15147069	-0.62466957	0.2684404	0.2409074	0.58072177	-0.13065129	0.10015489	-0.04367568	-0.09595323	-0.06374426
OM	0.3520575	-0.18910675	-0.0688991	0.20912402	0.20369869	-0.10854196	-0.07272036	0.07196093	0.117968213	0.64787253	-0.54260275
N	0.3518885	-0.1754667	0.08616178	0.21187146	0.22531473	-0.07060862	-0.15044584	0.48237474	0.548153174	-0.30843205	0.29393656
Total.P	-0.300938	-0.07250918	0.3856227	0.5085224	0.42061645	-0.0719779	0.52863888	-0.02258138	-0.12192985	-0.11251094	-0.08535356
K	0.2429586	-0.41007397	0.25162969	-0.44969587	0.4782464	0.18678833	-0.17309048	-0.04000938	-0.41157792	-0.19534685	-0.06737376
Ca	-0.2162263	-0.52112336	-0.31020459	-0.0330289	-0.21452557	-0.57817598	0.02789962	0.1574484	-0.11143131	-0.30590727	-0.26923143
Mg	-0.196757	-0.48594823	0.46297466	0.11418284	-0.4908778	0.41344581	-0.14453593	-0.05198342	0.209128813	0.04489986	-0.12851275
Na	0.3350169	-0.20750126	-0.19702875	-0.22710924	-0.21563847	0.25645854	0.77187459	0.17312141	0.004436647	0.07403781	0.10379947
CEC	0.3023285	-0.33755892	-0.16923361	0.34893993	-0.02606989	-0.12309062	-0.04535155	-0.65291065	-0.06574649	0.04237269	0.44040244
TBS	-0.3509701	-0.23161845	0.03100712	-0.08712522	0.08085082	-0.12659637	-0.07089463	0.35864453	-0.18772906	0.56014751	0.55623535

Supplementary Table 5 Soil chemistry values

Site	pH		Total P (%)		Olsen P (mg/L)		Total N (%)		C:N		Ca (me/100 g)[a]	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Lake Tennyson	5.5	0.1	792	53	18.5	0.5	0.66	0.08	16.9	1.6	8.9	2.9
Mt Hutt *R. crithmifolius*	6.1	0.03	712	14	12.6	0.5	0.059	0.005	8	0.7	7.18	0.69
Mt Hutt *R. monroi*	5.8	0.07	754	21	6.1	0.3	0.128	0.007	12.5	0.3	5.25	0.19
Porters Pass	5.4	0.12	438	37	5	0.7	0.351	0.064	27.5	1.3	5.83	1.93

	Mg (me/100 g)		K (me/100 g)		Na (me/100 g)		CEC (me/100 g)		TBS (%)		OM (%)	
	M	*SE*	*M*	*SE*	*M*	*NA*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Lake Tennyson	2.48	0.72	1.23	0.22	0.06	0	30	3	41	6	19.45	4.05
Mt Hutt *R. crithmifolius*	2.05	0.17	0.286	0.013	0.054	0.003	13.4	0.7	71	2	0.84	0.13
Mt Hutt *R. monroi*	2.49	0.11	0.505	0.044	0.055	0.003	13.6	0.3	62	2	2.74	0.21
Porters Pass	1.9	0.54	0.894	0.157	0.121	0.008	26.1	3.9	31	5	17.41	3.72

[a] milliequivalents per 100 g

Supplementary Table 6 Plant weights at potting

Taxon	Plant ID	Initial plant weight (g)	Taxon	Plant ID	Initial plant weight (g)
R. lobulatus	L1	10	*R. crithmifolius*	C6	19
R. lobulatus	L2	6	*R. crithmifolius*	C7	14
R. lobulatus	L3	9	*R. crithmifolius*	C8	27
R. lobulatus	L4	20	Mount Hutt *R. monroi*	M1	10
R. lobulatus	L5	10	Mount Hutt *R. monroi*	M2	14
R. lobulatus	L6	9	Mount Hutt *R. monroi*	M10	7
R. lobulatus	L7	13	Porters Pass *R. monroi*	M4	5
R. lobulatus	L8	11	Porters Pass *R. monroi*	M5	14
R. lobulatus	L9	9	Porters Pass *R. monroi*	M6	11
R. lobulatus	L10	9	Porters Pass *R. monroi*	M7	8
R. crithmifolius	C1	15	Porters Pass *R. monroi*	M8	5
R. crithmifolius	C2	10	Porters Pass *R. monroi*	M9	19
R. crithmifolius	C3	6	Porters Pass *R. monroi*	M3	9
R. crithmifolius	C4	10	Porters Pass *R. monroi*	M11	11
R. crithmifolius	C5	22	Porters Pass *R. monroi*	M12	11

Supplementary Table 7 Volumetric soil moisture at plant harvest

Taxon	Plant ID	Soil moisture (m^3/m^3)	Taxon	Plant ID	Soil moisture (m^3/m^3)
R. lobulatus	L1	0.237	*R. crithmifolius*	C8	0.234
R. lobulatus	L2	0.294	Mount Hutt *R. monroi*	M1	0.232
R. lobulatus	L3	0.268	Mount Hutt *R. monroi*	M2	0.234
R. lobulatus	L4	0.254	Mount Hutt *R. monroi*	M10	0.222
R. lobulatus	L5	0.263	Porters Pass *R. monroi*	M4	0.231
R. lobulatus	L6	0.268	Porters Pass *R. monroi*	M6	0.284
R. lobulatus	L7	0.218	Porters Pass *R. monroi*	M7	0.223
R. lobulatus	L8	0.201	Porters Pass *R. monroi*	M8	0.231
R. lobulatus	L9	0.258	Porters Pass *R. monroi*	M9	0.219
R. lobulatus	L10	0.263	Porters Pass *R. monroi*	M3	0.222
R. crithmifolius	C4	0.255	Porters Pass *R. monroi*	M11	0.221
R. crithmifolius	C6	0.227	Porters Pass *R. monroi*	M12	0.233
R. crithmifolius	C7	0.222			

Supplementary Table 8 RNA extraction statistics

Taxon	Plant ID	RNA Concentration (ng/mL)	NanoDrop 260/280	NanoDrop 260/230	LabChip RQS
R. crithmifolius	C4	112	2.01	2.11	5.8
R. crithmifolius	C7	30	1.87	2.14	5.8
R. crithmifolius	C8	110	2.00	2.05	5.7
Mount Hutt *R . monroi*	M1	54.8	1.97	2.28	5.5
Mount Hutt *R. monroi*	M2	32	1.98	1.91	6.4
Mount Hutt *R . monroi*	M10	227	2.07	2.04	6.1
Porters Pass *R. monroi*	M3	37.7	1.86	1.86	6.3
Porters Pass *R. monroi*	M7	70.7	1.96	2.18	5.6
Porters Pass *R. monroi*	M12	56.1	1.93	1.77	7.3
R. lobulatus	L5	38.8	1.93	2.20	6.2
R. lobulatus	L6	47.1	1.93	2.18	6.8
R. lobulatus	L7	80.7	2.05	2.24	6

Supplementary Table 9 Illumina read numbers prior to, and after, quality filtering

Analysis ID	FASTQ ID	Raw reads [a]	After correction trimming and filtering	(%) [b]
R. crithmifolius _1	CLH1FLL_C4_1_1	69142270	67377604	97.45
R. crithmifolius _2	CLH1FLL_C7_2_2	85253816	83461272	97.90
R. crithmifolius _3	CLH1FLL_C8_3_2	88132961	86332620	97.96
Mt Hutt-*R. monroi* _1	LLT1FLL_L7_4_2	75589288	73558633	97.31
Mt Hutt-*R. monroi* _2	LLT1FLL_L5_5_1	84275957	82053449	97.36
Mt Hutt-*R. monroi* _3	LLT1FLL_L6_6_2	72102774	70138934	97.28
Porters Pass-*R. monroi* _1	MLH1FLL_M2_53_2	84139592	81787209	97.20
Porters Pass-*R. monroi* _2	MLH1FLL_M10_55_2	83446573	80915976	96.97
Porters Pass-*R. monroi* _3	MLH1FLL_M1_58_2	66875853	64558184	96.53
R. lobulatus _1	MLP3FLL_M3_8_2	97736494	95713866	97.93
R. lobulatus _2	MLP3FLL_M7_9_2	83387623	81292797	97.49
R. lobulatus _3	MLP3FLL_M12_10_2	78575536	76841169	97.79

[a] Numbers given are for both forward and reverse reads

[b] Percentage of reads retained

Supplementary Table 10 Trinity assembly statistics

Assembly	Contigs	Clusters	%GC	N50	Median contig length	Average contig length	Longest contig	Assembled bases
R. crithmifolius	339382	162975	41.46	1391	691	1002	15491	340145038
Mt Hutt-*R. monroi*	359976	170719	41.06	1377	677	994	22812	357747494
Porters Pass *R. monroi*	310862	131148	40.73	1446	741	1047	16783	325612100
R. lobulatus	341220	140165	40.62	1459	754	1060	16741	361839591
R. crithmifolius _1	187296	90847	41.66	1446	740	1043	13455	195321971
R. crithmifolius _2	218273	111277	41.77	1417	726	1024	12687	223456551
R. crithmifolius _3	190608	85949	41.02	1468	781	1070	15132	204000910
Mt Hutt-*R. monroi* _1	182066	90933	41.23	1417	716	1021	22238	185979014
Mt Hutt-*R. monroi* _2	190851	85851	41.02	1480	774	1073	15437	204720286
Mt Hutt-*R. monroi* _3	240520	125130	41.6	1461	697	1032	16650	248132947
Porters Pass-*R. monroi* _1	173843	78413	41.93	1447	757	1053	15423	183010265
Porters Pass-*R. monroi* _2	198255	89279	41.15	1510	796	1093	14919	216789724
Porters Pass-*R. monroi* _3	191626	86663	41.12	1491	788	1083	15467	207548880
R. lobulatus _1	202534	92254	40.93	1505	808	1099	15043	222503503
R. lobulatus _2	206174	92935	41.04	1526	809	1107	17353	228322323
R. lobulatus _3	184684	82268	41.22	1546	833	1124	16377	207548335

Supplementary Table 11 Representative assembly statistics

Assembly	Contigs	%GC	N50	Median contig length	Average contig length	Longest contig	Assembled bases	TransDecoder Predicted Proteins
R. crithmifolius	162975	42.44	811	484	699	15491	113950405	51866
Mt Hutt-*R. monroi*	170719	41.05	804	484	697	14620	118991729	50082
Porters Pass *R. monroi*	131148	40.56	884	504	735	16783	96353662	28581
R. lobulatus	140165	40.37	922	511	752	15582	105454274	29749
R. crithmifolius _1	90847	42.16	904	495	733	13455	66604833	28953
R. crithmifolius _2	111277	42.62	886	501	729	12687	81125888	42161
R. crithmifolius _3	85949	40.95	931	518	751	15132	64549309	24313
Mt Hutt-*R. monroi* _1	90933	40.93	877	495	729	22238	66291358	24593
Mt Hutt-*R. monroi* _2	85851	40.92	902	515	744	15421	63897160	24177
Mt Hutt-*R. monroi* _3	125130	42.12	866	490	723	15508	90436756	44984
Porters Pass-*R. monroi* _1	78413	41.18	887	494	730	15423	57213527	23243
Porters Pass-*R. monroi* _2	89279	41.04	982	532	779	14919	69553516	25018
Porters Pass-*R. monroi* _3	86663	40.95	960	518	765	15467	66321367	24485
R. lobulatus _1	92254	40.74	992	538	781	15043	72007971	25080
R. lobulatus _2	92935	40.78	1004	534	787	15506	73133615	25155
R. lobulatus _3	82268	40.97	1051	544	804	15213	66171419	23962

Supplementary Table 12 Gene expression and nucleotide analyses results

Aggregated Assembly Standard Pairwise Gene Expression			
Mount Hutt *R. monroi* grouped with:	DEGs	Functionally Annotated[a]	GO Annotated[b]
Porters Pass *R. monroi*	5773	1874	3899
R. crithmifolius	2807	836	1971
R. lobulatus	2196	1862	334

Aggregated Assembly Pairwise Gene Expression (genes with all samples mapping)			
Mount Hutt *R. monroi* grouped with:	DEGs	Functionally Annotated	GO Annotated
Porters Pass *R. monroi*	3600	1568	1405
R. crithmifolius	1704	803	701
R. lobulatus	633	181	163

Aggregated Assembly Gene Expression Ranking			
Mount Hutt *R. monroi* grouped with:	DEGs	Functionally Annotated	GO Annotated
Porters Pass *R. monroi*	1036	301	262
R. crithmifolius	295	115	97
R. lobulatus	198	81	77

Aggregated SCO Gene Expression Ranking			
Mount Hutt *R. monroi* grouped with:	DEGs	Functionally Annotated	GO Annotated
Porters Pass *R. monroi*	176	151	137
R. crithmifolius	98	91	80
R. lobulatus	82	64	59

Aggregated Assembly SCO Nucleotide Variation Phylogeny			
Mount Hutt *R. monroi* grouped with:	Retained Trees	Functionally Annotated	GO Annotated
Porters Pass *R. monroi*	893	746	642
R. crithmifolius	454	362	309
R. lobulatus	481	391	331

Individual Assembly Nucleotide Variation Phylogeny			
Mount Hutt *R. monroi* grouped with:	Retained Trees	Functionally Annotated	GO Annotated
Porters Pass *R. monroi*	74	73	73
R. crithmifolius	14	14	14
R. lobulatus	14	14	14

[a] Number of DEGs or phylogenetic trees annotated with *A. thaliana* gene function

[b] Number of DEGs or phylogenetic trees annotated with GO terms

Supplementary Table 13 agriGO v2 output

GO Term	Ontology	Description	Number in Input List	Number in BG/Ref	*p* value	FDR[c]
GO:0043038	P[a]	amino acid activation	5	46	3.4e-05	0.0086
GO:0043039	P	tRNA aminoacylation	5	46	3.4e-05	0.0086
GO:0006418	P	tRNA aminoacylation for protein translation	5	43	2.5e-05	0.0086
GO:0016875	F[b]	ligase activity, forming carbon-oxygen bonds	5	48	4.2e-05	0.0026
GO:0004812	F	amino acyl-tRNA ligase activity	5	48	4.2e-05	0.0026
GO:0016876	F	ligase activity, forming amino acyt-tRNA and related compounds	5	48	4.2e-05	0.0026

[a] Biological Process

[b] Molecular Function

[c] False Discovery Rate (Yekutieli-corrected *p* value)