The genetic architecture of the divaricate growth form: A QTL mapping approach in *Sophora* (Fabaceae).

A thesis presented in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

in

Plant Biology

At Massey University,

Manawatu,

New Zealand

Kay Margaret Pilkington

2019

Abstract

Divarication is a plant growth form described, in its simplest form, as a tree or shrub with interlaced branches, wide branch angles and small, widely spaced, leaves giving the appearance of a densely tangled shrub. The frequency of this growth form is a unique feature in the New Zealand flora that is present in ~ 10% of the woody plant species, a much higher frequency than that of other regional floras. While several hypotheses have been developed to explain why this growth form has evolved multiple times within New Zealand, to our knowledge, no work has addressed the genetic basis of the divaricating form. *Sophora* is one of several genera in New Zealand that possesses divaricate species. Among the factors making this an ideal system for a genetic investigation of divarication is an existing F_2 population formed from reciprocal crosses between the divaricating *S. prostrata* and the non-divaricating *S. tetraptera*.

Using this segregating population and newly developed molecular markers, the first linkage maps for *Sophora* were generated, providing a new genetic resource in *Sophora*. These linkage maps allowed for quantitative trait locus (QTL) mapping for traits associated with the divaricate form in the segregating population. Multiple QTL were mapped to seven of the divaricate traits with many QTL co-locating for multiple traits, indicating that the divaricate growth form is genetically controlled by many loci, potentially including pleiotropic loci, that each contribute to the overall divaricate phenotype in *Sophora*.

The strigolactone biosynthesis and perception pathway is a good candidate for involvement in control of the divaricate form based on mutant phenotypes in *Pisum* that display similarities to the divaricate growth form, such as increased branching, shorter plant height and smaller leaves. QTL, for multiple traits, were mapped to two candidate genes investigated, *RMSI* and *RMS4*. An amino acid replacement was identified in RMSI,

in *S. prostrata,* and was predicted to be deleterious suggesting RMSI may be non-functional in *S. prostrata.* These results support *RMSI* as a strong candidate gene for future work on divarication. This study is the first to investigate the genetic architecture of the divaricate growth form and contributes to further understanding of this unique feature in the New Zealand flora and of plant architecture generally.

Acknowledgments

Throughout the research and writing of this thesis, I have received a great amount of support and assistance. I would first like to express my sincere gratitude to my supervisor, Dr. Vaughan Symonds, for the invaluable knowledge, patience and guidance in undertaking the research and writing of this thesis. I am also grateful to my co-supervisors, Dr. Jennifer Tate, for the insightful comments, encouragement and guidance and, Dr. Peter Heenen, for the support, knowledge and care of the plants at the study site.

My sincere thanks also go to the Massey Graduate Scholarship and Heseltine Trust Scholarship for the funding that enabled me to undertake this research. Without this support, I would not have been able to take this opportunity.

I would like to thank my fellow labmates and colleagues for all the support and willingness to help me. I would especially like to thank Sofie and Jacob for the help with collecting and measuring of plants and Lesley from the Massey Horticultural Units for the continuous care of the growing plants and seedlings. Likewise, my sincere thanks to the lab technician, Prashant, for the help and knowledge in the lab. Also, I would like to thank all the administrative staff that in one way or another helped and supported me during my PhD study.

In addition, I would like to thank my parents for their support and counsel during the course of my PhD. Finally, my thanks to family and friends for the wonderful support and fun you provided throughout this time.

Table of Contents

List of Figures

List of Tables

List of Appendices

Chapter One - Introduction

1.1 Plant architecture

Plant architecture is defined as the three-dimensional organisation of the plant body (Figure 1.1) and describes the branching pattern, size, shape and position of leaves and reproductive organs (Reinhardt & Kuhlemeier, 2002). Different plant architectures have arisen in response to environmental pressures (Reich et al., 2003; Rowe & Speck, 2005) which can include climate, competition and predation by other species. These different architectures enable plant species to survive in and occupy habitats that would otherwise be unsuitable. In addition to adaptation, plant architecture has been utilised in taxonomic and systematic classification and is still the most common method of identifying a species today. It is also important for agriculture as it can influence plant stability (Quine, 1990; Plourde et al., 2009), yield (Huyghe, 1998; Peng et al., 1999; Dawo et al., 2007; Kaggwa-Asiimwe et al., 2013; Rosati et al., 2013) and harvest efficiency (Takeda et al., 2013).

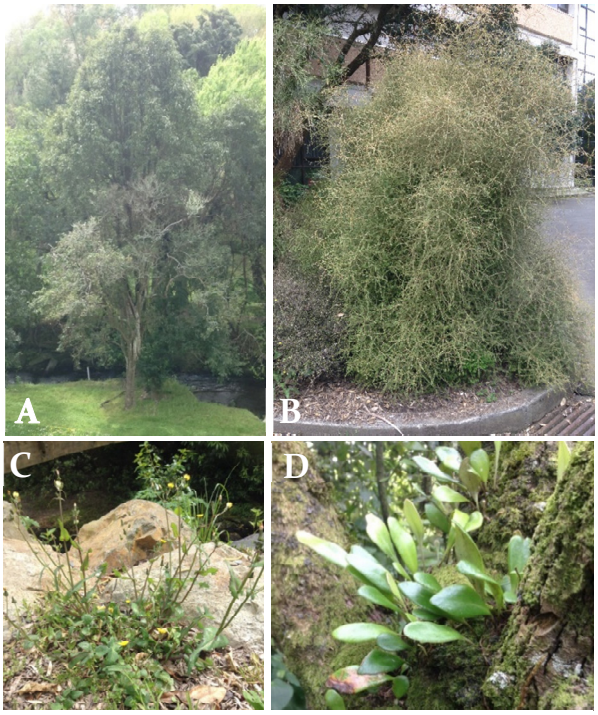


Figure 1.1. Examples of a different plant architectures. A- arborescent tree, B - divaricate shrub, C - Herbaceous plant, D - Epiphytic growth habit.

Plant architecture is 'open and indeterminate' compared to body architecture in animals (Niklas, 2000). Although it is still under strong genetic control, plant architecture also can be strongly influenced by the environment resulting in phenotypic

plasticity (Reinhardt & Kuhlemeier, 2002). The result is an uneven distribution of architectures found in particular habitats across the world (Brunig, 1976; Tranquillini, 1979). Different plant architectures are believed to have evolved from a simple body plan that has differentiated into the diverse plant forms seen today (Sussex & Kerk, 2001). Plant body architectures have often been lost and reverted back throughout the history of many groups as seen in the distribution of plant architectures in plant phylogenies such as *Aloe* (Asphodelaceae), where transitions between Tomlinson's (basal branched shoot with terminal inflorescences) and Corner's (unbranched shoot with lateral inflorescences) architectural models were identified (Chomicki et al., 2017), and *Echium* (Boraginaceae), which gained the tree habit following the invasion of Macaronesia and subsequent diversification (Bohle et al., 1996). The simplest architecture is considered to be a single shoot that terminates in a reproductive structure (Chomicki et al., 2017). Unbranched bryophyte-like plants with this morphology likely gave rise to the diverse and complex plant architectures seen today (Bower, 1908; Chomicki et al., 2017). From the simple architecture, forms that are more complex evolved. Dichotomous branching involves the apical meristem dividing into two meristems. The evolution of this architecture in protracheophytes (early vascular plants) is considered a major event in plant evolution (Chomicki et al., 2017). The evolution of axillary branching, and other complex branching architectures, likely contributed to the reduction in dichotomous branching. However, the dichotomous architecture is still observed in some derived higher plant groups such as the Cactaceae (Mauseth, 2004), where dichotomous branching may be a way to overcome the restriction of axillary branching imposed by large stems (Chomicki et al., 2017).

Plant architecture morphology arises from the complex genetic mechanisms contributing to these different plant forms. Several genes have been identified in the extant moss species, *Physcomitrella patens,* that can result in the formation of a double

sporangium from mutant forms such as *PINB* (Bennett et al., 2014), *TCP5* (Ortiz-Ramirez et al., 2016) and *TEL1* (Vivancos et al., 2012) as well as one gene, *CLF*, that results in a sporophyte-like body forming from the gametophyte that does not develop into a sporangium and later also forms branches (Okano et al., 2009). This suggests bryophytes may have the elements required to develop dichotomous branching (Chomicki et al., 2017). However, the independent dichotomous branching origins in angiosperms result from meristems that are different from bryophytes and therefore could involve different genetic mechanisms (Chomicki et al., 2017). Different phylogenetic analyses have revealed the Charales (Karol et al., 2001), the Coleochaetales (Finet et al., 2010), or more currently the Zygnematophyceae (Timme et al., 2012; Wickett et al., 2014) as sister groups to land plants, suggesting all land plant architectures have arisen from a common ancestor. Plant architecture is likely a result of the development of meristems (Chomicki et al., 2017) as these give rise to the structures making up plant architecture and so studies on meristem initiation and control have been important in understanding the genetics of plant architecture. Axillary meristem initiation in *Arabidopsis* involves a complex, and tightly controlled, pathway including interaction from many genes such as *LAS1* (Greb et al., 2003), the *CUC* genes and microRNA, *miR164* (Raman et al., 2008). Control of the fate of the axillary buds involves a complex interaction between multiple pathways including auxin, cytokinin and strigolactone hormone pathways (Chomicki et al., 2017).

1.1.1 Woody plant architectures

Woody plant architecture has been described as an architectural continuum with a large variety of architectures in many different habitats (Halle et al., 1978). A perennial woody habit is believed to be the ancestral trait to an annual habit (Sinnott, 1916; Arber, 1928; Feild et al., 2004) which evolved independently multiple times. A

reversion to the woody perennial habit is also thought to have occurred multiple times (Carlquist, 1974); for example, insular woodiness developing in island species such as *Echium* (Bohle et al., 1996) and *Argyranthemum* (Asteraceae) (Doria et al., 2018). A study in *Arabidopsis* (Brassicaceae) showed the MADs box genes, *soc1* and *ful* double mutants produce perennial-like phenotypes from the annual *Arabidopsis* including longer lifespan, woody stem, recurrent growth and vegetative buds on floral structures showing a switch in plant architecture and life history traits and indicating annuals retain the possibility to develop a perennial habit (Melzer et al., 2008). Tree architecture has also been of interest, particularly in crop plants. Tree architecture has been described as multiple components (1) the phytomer, including an node, internode and leaf, (2) degree of apical dominance and apical control (3) branch angles (4) reaction wood, specialised wood that forms to cope with mechanical weight of the branches and (5) stem diameter and branch size (Hollender & Dardick, 2015).

Many of the genetic mechanisms involved in plant architecture studied in non woody species, such as *Arabidopsis,* are also involved in determining plant architecture in woody species. For example orthologs of genes involved in the strigolactone pathway of axillary branch control, such as *MAX2*, *CCD8*, have been identified in *Populus* (Salicaceae) (Czarnecki et al., 2014) and willow (Salmon et al., 2014) indicating these pathways are conserved in seed plants. Research into tree architecture has often focused on architecture of crop plants for example, peach and apple which shows changes in plant architecture can result from the action of individual genes as well as multiple genes. Pillar peach, columnar apple, weeping peach and compact peach architectures have been identified as the action of a single gene or locus (Mehlenbacher & Scorza, 1986; Yamazaki et al., 1987; Bassi & Rizzo, 2000; Moriya et al., 2012; Hollender & Dardick, 2015) showing that a large change in architecture can result from a single gene, either from functional or regulatory control. However, the compact apple architecture

indicates a more complex genetic contribution involving a single gene with modifiers, indicating that multiple genes are involved in producing this phenotype (Blazek, 1982).

One area of interest in tree architecture is phase change, the shift between juvenile to adult stages, which can sometimes involve a large change in morphology and architecture. MicroRNAs are involved in a conserved mechanism for the phase change in herbs, such as *Arabidopsis,* and trees, such as *Eucalyptus* (Myrtaceae) (Poethig, 2009; Wu et al., 2009; Jung et al., 2011; Wang et al., 2011; Chomicki et al., 2017). This involves down-regulation of *miR156* by a signal derived from the leaves (Yang et al., 2011) and *miR172* upregulated to promote phase change to adult and reproductive transition (Wu et al., 2009; Zhu & Helliwell, 2011). *miR156* is known to be involved in phase change in *Eucalyptus* species (Chomicki et al., 2017) including *Eucalyptus globulus,* which has a change in phyllotaxy, from decussate juvenile to spiral adult, and in leaf shape (Wang et al., 2011). These genes are conserved across land plants suggesting they may be involved in phase change across many plant species.

Study of plant architectures have often focused on agricultural fields or evolution of early plants (Niklas, 2000) with less focus on the current variety of plant architectures which range from small herbs to large trees. However, research on plant architecture has revealed genetic pathways and processes that may contribute to developing the variety of different architectures seen in plant species. Further studies of plant architecture, especially the genetic basis of different architectures, will help advance our understanding of plant evolution in general as well as increasing knowledge that can contribute toward systematics, plant development, agriculture and plant growth forms of plant species today.

1.2 Defining divarication

Among the variety of plant growth forms seen is the divaricate growth form (Figure 1.1B) which, in its simplest form, is defined as a plant with interlacing branches, wide branch angles and small leaves (Figure 1.2). Divarication is a growth form comprising ~ 10% of the native New Zealand woody flora, which is a much higher frequency than observed in many other regional floras. However, there are reports that the divaricate growth form is common in Argentina (Kelly & Ogle, 1990) and California (Tucker, 1974). The high frequency of divaricates is a long-standing mystery of the evolution of the New Zealand flora, with nothing known about the genetic origins of this form. In New Zealand, the divaricating growth form is found in ~ 50 species across 18 families (Greenwood & Atkinson, 1977; Atkinson & Greenwood, 1989). The prevalence of divarication in disparate plant families suggests that it has evolved independently multiple times.

Figure 1.2. An example of a divaricate *Coprosma* species displaying the interlacing branches, wide branch angles and small leaves of the divaricate form.

As a variety of divaricate morphologies are observed across different plant groups (Bulmer, 1958; Tomlinson, 1978; Kelly, 1994), the definitions of plant divarication vary (Table 1.1) but the divaricate form is typically described as small-leaved shrubs or juvenile trees with wide branch angles, interlacing branches, long internodes and weak apical dominance. In the past, the lack of a consistent divaricate definition did not allow comparisons of divaricate plants from New Zealand to those elsewhere. Such confusion led to the independent development of divarication indices by Atkinson (1992) and Kelly (1994). A divarication index is important to determine the defining characteristics of divarication and provide a quantitative method to measure these. This helps in

identifying divaricate species from non-divaricate species, a necessary process for investigating the evolution of this growth form. Atkinson (1992) developed a divarication index that focused mainly on branching patterns and was developed to be easily applicable in the field, with simple to measure parameters, enabling a quick and effective way to determine the index value for plants. The index value of Atkinson is obtained by adding each of four measured parameters, some of which are scaled so that each individual parameter falls mostly in a range of 1 - 10, as indicated below:

$$Index_{(A)} = Number\ of\ peripheral\ wide\text{-}\ angle\ branches + (mean\ branch\ angle/10) + number\ of\ branching\ orders + (branching\ density/10)$$

Atkinson (1992) noted that this study was preliminary, with a small sample size, but pointed out that (1) the degree of divarication between species can be quantified using this method and (2) divaricate species can be distinguished from non-divaricate species. However, two species sampled, *Pittosporum anomalum* Laing & Gourlay (Pittosporaceae) and *Melicytus alpinus* (Kirk) Garn.-Jones (Violaceae), were not as highly valued by the index as would be expected based on their perceived 'extreme' divaricate form. This indicates a weakness in the index and Atkinson suggested that another parameter, specifically one that reflects the high degree and tightness of branching may be required.

In a separate study, Kelly (1994) identified several variables that best distinguished divaricates from non-divaricates within the New Zealand flora: leaf width, internode length relative to leaf width, and branch angle (Table 1.1) and used these variables to develop an index that can be used to identify divarication:

$$Index_{(K)} = FILIW + (10/LWT) + ANGL/5$$

Table 1.1. Definitions of abbreviations used in Kelly's index (modified from Kelly (1994)).

Abbreviation	Name	Definition
FILIW	Internode/leaf width ratio	Internode length to leaf : width ratio (100/NLT)/LWT
NLT	No. leaves 10 cm, tip	Number of leaves per top 10 cm of main branch
LWT	Leaf/width tip	Mean leaf : width at tip of sample (5 reps)
ANGL	Branch/main stem angle	Angle between branch and main stem, from above (5 reps)

The results obtained from this index were consistent with the perceived degree of divarication for plants measured in the study (Kelly, 1994). Kelly (1994) proposed a formal definition of divarication as "a plant with interlaced branches with wide angles, small leaves widely spaced with larger leaves towards the interior of the plant". Interlacing branches are frequently used to define a divaricate (Table 1.2), however, as Kelly (1994) noted, interlacing branches are difficult to obtain a numerical value to use in an index. The measures attempted by Atkinson (1992) were not suitable for measuring interlacing in the field due to the time required to take these measurements so are also excluded from the index.

These two indices focus on different characteristics of divaricating plants, but correlated well with each other for most of the species included in both New Zealand studies (Kelly, 1994). However, in Patagonia, a comparison of these indices for divaricating plants showed no significant correlation between the indices, suggesting they measure different characteristics that correlate in New Zealand divaricating plants but are not characteristics of the Patagonian divaricate plants (McQueen, 2000). McQueen (2000) observed that most New Zealand divaricates form a shield-like structure from the interlacing of peripheral branches but this form is not present in

Patagonian divaricates. Instead, Patagonian divaricates commonly have spines suggesting different characteristics of divaricates from each of these regions.

Patagonian divaricate plants are only found in open country, unlike New Zealand, where divaricates are common in forests, especially forest edges (McGlone & Webb, 1981). McQueen (2000) suggests that Patagonian divaricates may have evolved in response to small browsing animals where the spines are likely to deter the soft-nosed mammals from browsing. In contrast, the only large browsers in New Zealand were ratites, the moa, which had a horny covering protecting their head (Greenwood & Atkinson, 1977) making spines less effective to deter browsing. However, Patagonia also has an extant small browsing ratite, Darwin's rhea *Pterocnemia pennata pennata* (Rheidae) and the effect of environmental conditions in Patagonian divaricate evolution could not be ruled out. New Zealand and Chile have many plant genera in common (Godley, 1960) but only two genera east of the Andes share the divaricating growth form, providing evidence that divarication may arise in response to local environmental conditions rather than within the history of a lineage (McQueen, 2000). The range of degree of divarication and the lack of a consistent definition make determining the origins of the divaricate growth form problematic, but, ultimately, determining the origin of divarication may be important in developing a universal divaricate definition, and several hypotheses have been suggested for why this form is more predominant in the New Zealand flora.

1.3 Hypotheses for origin of divarication:

The high prevalence of divaricate plants in New Zealand compared to most other floras has raised the question of how and why this growth form has evolved at a higher frequency in New Zealand. Greenwood and Atkinson (1977) identified three evolutionary trends in divaricate plants. The first is development of an increased proportion of wide-angled lateral branches, the second is reduced leaf size, and sometimes leaf/leaflet

Table 1.2. Divaricate definitions from the literature and the variables used in those definitions (Y = yes, S = sometimes, often or not exclusively to divaricate). Table modified from Kelly (1994). Interlacing branches, small leaves and wide branch angles are most often used to describe divaricate forms.

Trait	Cockayne (1912)	Bulmer (1958)	Went (1971)	Greenwood and Atkinson (1977)	Tomlinson (1978)	McGlone and Webb (1981)	McGlone and Wardle 1988)	Dawson (1988)	Atkinson and Greenwood (1989)	Kelly and Ogle (1990)	Wardle (1991)	Wilson (1991)	Atkinson (1992)	McGlone and Clarkson (1993)	Cooper et al. (1993)	Kelly (1994)
Interlacing	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y
Small leaves		Y	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y
Wide branch angles	Y			Y	S	Y		Y		Y			Y	Y		Y
Highly branched	Y	S	Y					Y		Y		Y	Y			
Fewer leaves at tip				Y		Y	Y		Y	Y			S	Y	Y	
Long internodes			Y								Y		S	Y	Y	
Some branches >90°		Y	Y	Y				Y					Y			
Tough stems				Y					Y				S	Y	Y	
Smaller leaves at tip				Y					Y				S	Y	Y	Y
Reduced apical dominance			Y	Y				Y					Y			
Many short shoots					Y			S		Y	Y					
Stiff wiry stems	S	S		S	S					Y						
Spineless		S			S					S	S					
Lateral flowering					Y					Y						
Small buds							Y				Y					
Slender stems							Y				Y					
Entire leaves			S							Y						
Recurved branches					Y						S					
Continuous growth of higher order branches													Y			
Crooked branchlets												Y				
Sympodial branching					S											
Zig-zag stem				S	S			S			S					

number, so that the exterior of the plant becomes woodier, and the third is increasing toughness of the stems compared to the non-divaricating relatives. There have been several hypotheses proposed (Appendix 1.1) for the evolution of divaricating plants in New Zealand but the precise cause(s) remain unclear.

1.3.1 Moa Browsing Hypothesis

Greenwood and Atkinson (1977) proposed the divaricate form evolved to reduce browsing by moa (Emeidae). This hypothesis had been briefly considered earlier (Denny, 1964; Went, 1971; Carlquist, 1974; Livingstone, 1974; Taylor, 1975) but had not been explored in detail prior to Greenwood and Atkinson. Moa were large flightless birds belonging to the ratite group that includes ostriches, emus and kiwi. Ratites are found in many countries including Australia, South America, Africa, New Guinea, Madagascar and New Zealand. Historically, moa were the only large browsing herbivores in New Zealand because New Zealand had no browsing mammals before human colonisation (Greenwood & Atkinson, 1977). Moa lacked teeth and a prehensile tongue that resulted in a clamping, pulling and breaking action important for browsing (Greenwood & Atkinson, 1977) for some species (Attard et al., 2016). Bond et al. (2004) observed the feeding behaviour of extant ratites and determined a clamping and tugging action was the primary method of feeding. However, divaricate twigs have been found in moa gizzards which show clear-cut ends rather than broken ends, suggesting moa may not have fed by tugging and breaking twigs (Burrows, 1980a, 1980b) and that they did feed on some divaricate plants. The increased stem toughness and smaller leaves of divaricate plants means browsing on these plants becomes less profitable for herbivores (Greenwood & Atkinson, 1977; Bond et al., 2004).

Bond et al. (2004) demonstrated that thin, strong, highly branched shrubs with small, widely spaced leaves, like the divaricate growth form, reduce the offtake from the

plants, so that the birds are unable to maintain their daily energy requirements from eating divaricate plants alone. This finding indicates that although they may eat some divaricate plants, these were not likely a large or nutritious source in the diet. A cafeteria-like study to compare feeding preferences between extant ratites (ostriches) and ungulates (deer and goats) observed that all three herbivores ate less of a species with small leaves and that divaricates, and conifers, were generally the least browsed species (Pollock et al., 2007). These results show that although the divaricate growth form reduces browsing from ratites, it also is reduced for mammalian browsers with little difference in feeding preference between them and food mainly selected based on its structure. Therefore, the divaricating growth form has the same effect on reducing browsing from mammals as extant ratites which does not explain why divaricates are more common in New Zealand and not elsewhere if browsing from mammals is also reduced.

Bond and Silander (2007) compared the divaricate growth forms in Madagascar, which had a grazing ratite similar to moa (the now extinct elephant birds), to the New Zealand divaricates. They found that many of the species in Madagascar had traits consistent with the 'wire plant syndrome' of New Zealand, considered to reduce ratite grazing. However, Malagasy divaricates lacked the higher concentration of leaves in the interior of the shrub, which is commonly described for New Zealand divaricates. They suggested that New Zealand plants contain traits that prevent moa browsing and aid in protection from climate, specifically for cold tolerance, whereas the Malagasy divaricates growing in the tropical climate of Madagascar did not need protection from cold climates.

Greenwood and Atkinson (1977) observed that divaricate plants often grew in lowland and fertile soils which are places where moa are thought to have inhabited. As

most of pre-colonial New Zealand was forested, it is likely moa evolved in a forested environment. Consistent with this idea, New Zealand divaricate plants often grow in forested environments. With the extinction of moa, slow-growing divaricates may be at a disadvantage, especially as they often fail to grow on low-fertility soils, compared to other plant growth forms, and this may explain the reduced range seen today (Greenwood & Atkinson, 1977). Greenwood and Atkinson (1977) also suggest heteroblastic divaricate species may have had an advantage, as they are divaricating as juveniles offering protection from browsing moa, but become arborescent as they transition to adults when they are tall enough to be above the range of browsing moa. If divaricates arose in response to moa browsing, it might be expected that divaricates would not be found where moa (or other browsing ratites) were not present, such as the Chatham Islands. However, divaricate species are present on the Chatham Islands, but have smaller branching angles and larger leaves. The lack of moa may contribute to the difference in these divaricates (Kavanagh, 2015), but this could also indicate the divaricate growth form may be influenced by other factors, such as the climate.

1.3.2 Climate Hypothesis

Although the moa browsing hypothesis is often favoured, there are still uncertainties and alternative explanations that have been proposed. One alternative hypothesis suggests this growth form results from adaptation to an earlier climatic period where conditions were windier and drier than at present. Initially proposed by Diels (1897) as a response to the wind, frost and abrasion of earlier glacial climates, the theory was later expanded on (Cockayne, 1912a; McGlone & Webb, 1981) to form the climate hypothesis. This hypothesis suggests the smaller, sparser leaves and interlacing branches of the divaricating growth form provide protection from wind, reduce leaf damage, act as a heat trap and as a frost screen (McGlone & Webb, 1981).

Densely branched shrubs have significantly calmer airflow zones, with less wind velocity, than regular branched shrubs, indicating that dense branching produces suitable environments for growth of new leaves, flowers or fruits. This form may reduce transpiration in these areas, enabling plants to survive in windier conditions. However, this effect still needs to be investigated further (Keey & Lind, 1997). Kelly & Ogle (1990) investigated the difference in temperature on internal leaves to outer leaves that was proposed to result from the interlacing branches of the divaricate form, but found the plants studied had little difference in temperature inside compared to the exterior and concluded there was no heat trap operating. However, a difference in leaf temperature, of 0.4-0.6°C, has been observed for divaricate shrubs on the forest margin compared with non-divaricate species but the significance of these small temperature differences is not clear and there was little difference for species in the understory of the forest (Lusk & Clearwater, 2015).

The climate hypothesis suggests the interlacing branches and internal leaves may act as protection by protecting inner leaves from cold. However, frost tolerance varies with species (Kelly & Ogle, 1990) with some divaricate juveniles no different or even less frost tolerant than the non-divaricate adults (Darrow et al., 2001) Therefore, while the divaricate growth form may provide some protection against climatic conditions, such as wind, many of these advantages vary with the species. The absence of or reduced divarication of plants on islands, such as the Chatham Islands which Kavanagh (2015) suggested resulted from lack of moa, may also be attributed to milder climates often seen on offshore islands (McGlone & Webb, 1981). Milder conditions mean plants require less protection from the climate and could result in a reduced divaricate form.

A model developed by Lusk (2016) examined the correlation between environmental factors and the geographic distribution of divaricates and showed the

proportion of the divaricate growth form was best predicted by the minimum temperature of the coldest month, suggesting a link between environmental conditions and prevalence of the divaricate growth form. However, the environmental conditions suggested for the evolution of divaricates are also found in other parts of the world where divaricate species are less numerous or lacking (Dawson, 1963), raising the question of why this growth form only arose in these conditions at such a high frequency in New Zealand. Divaricate plants also occur in a wide range of habitats in New Zealand, which would have slightly different climatic conditions and could explain the variation seen in the divaricate form. However, the climate hypothesis does not explain the wide branching angles often found in divaricate plants (Greenwood & Atkinson, 1977).

A study of Williamsoniaceae, a family of extinct gymnosperms which are typically depicted as shrubs with a broadly divaricating growth form (Thomas, 1916; Pott, 2014), concluded that these divaricate plants evolved due to the climate-adaptation hypothesis. The fossil species studied were believed to grow under conditions that are today thought to result in the development of the divaricate growth form in modern divaricates (Pott & McLoughlin, 2014). However, they noted the impact of browsers cannot be ruled out as there were many large herbivores in areas occupied by these species which potentially browsed on divaricates.

Another fossil of a shrub fragment, found in the Nevis Valley in New Zealand, displays some of the characteristics of many extant divaricate shrubs (Campbell et al., 2000). Using Kelly's (1994) index this fossil scored 19.6, just above the threshold for divarication of 19, likely due to the larger leaves and their lower density. However, Atkinson's (1992) index could not be fully utilised as only some characters could be measured. The sample was identified as either a fossil of an extant species, of which *Raukaua anomalus* (Hook.) A.D.Mitch., Frodin & Heads (Araliaceae) is considered the

most likely species, or an extinct species, believed to be a divaricate shrub. It was dated as being from the Early Miocene, suggesting the divaricate habit has been present in the New Zealand flora at least since the early Miocene. The authors suggest this fossil is consistent with the moa-browsing theory because the fossil pre-dates the climatic conditions that are considered to have led to the development of the divaricate form (Campbell et al., 2000). However, the dating of the fossil also confirms the divaricate habit was present in the Early Miocene.

In New Zealand, the Miocene period resulted in reduced forested areas to scrubland or grassland and the beginning of the Pleistocene resulted in climate cooling (Pole, 2014) with the onset of glaciations. The New Zealand flora was mainly derived from sub-tropical and warm-temperate climates (Dawson, 1962) which meant it had to adapt to a cooler, drier and increasingly unstable climate. McGlone & Webb (1981) suggest that this was the driving force for the evolution of divarication. However, the divaricate form also comes at a cost to the plant due to the high degree of self-shading (McGlone & Webb, 1981), therefore divaricates may only be expected where there was a higher risk to the plants from the harsh climate. Greenwood & Atkinson (1977) observed divaricate plants to prefer forest-edges, more open habitats and fertile soils which they attributed to being where moa frequently browsed. However, this could also be explained as these areas may be more exposed to harsher climates than other forested areas (McGlone & Webb, 1981) and the cost of the divaricate habit may require more fertile soils and prevent divaricates from competing further in forested environments.

It has also been suggested, and is highly possible, that both the moa browsing and climate hypotheses were factors contributing to the divaricate growth form (Wardle, 1985; Cooper et al., 1993). This may have occurred in the Pliocene, when the climate began to develop into harsher frosty and droughty conditions and increasing exposure to

grazing ratites, especially for juvenile plants (Lusk et al., 2016). These two hypotheses have received the most support for the evolution of the divaricate growth form, although there are other hypotheses or factors proposed that may have had an effect on the divaricate growth form.

1.3.3 Alterative Hypotheses

One further hypothesis suggests the divaricate growth form developed to increase light harvesting. This is important in forests where light levels are lower below the canopy. The close packing of leaves at diverse angles, as seen in the divaricate growth form, may increase light harvesting for the plant (Kelly, 1994). Consequently, these plants should have small, narrow-width leaves (possibly with teeth or lobes), internodes greater than the leaf size, and have many leaves in the interior, all of which would enable light to penetrate into the centre. Lobes or teeth on leaf margins were rarely found on divaricate plants, but Kelly (1994) noted leaves are already narrow. A further hypothesis considering the effect of irradiance and photosynthesis on the evolution of divaricates suggests that photoinhibition may be important in divaricate evolution.

Photoinhibition is a light-induced decrease in photosynthetic efficiency as a result of light levels exceeding the photosynthetic requirement and causing photodamage. Exposure to bright sunlight following a cold night promotes cold-induced photoinhibition and this has been observed in non-divaricate tree seedlings (Ball, 1994). Divaricates in New Zealand are frequently found in exposed, frosty inland habitats (McGlone & Webb, 1981) such as basins or valley bottoms where these conditions are common, so we may expect photoinhibition to be important to these divaricate plants. In *Aristotelia fruticosa* Hook. F. (Elaeocarpaceae), *Corokia cotoneaster* Raoul. (Escalloniaceae) and *Coprosma propinqua* A. Cunn. (Rubiaceae) the shielding of leaves with an outer screen of branches reduced the photoinhibition of photosynthesis, but

plants differed in the susceptibility to photoinhibition and the ability to recover from this (Howell et al., 2002) suggesting species differences play a role. Divaricates do possess some traits that may be advantageous to avoid photoinhibition: upper canopy with fewer leaves than lower canopy and smaller leaf and stem measurements, but the species studied showed no distinct response to high light (Christian et al., 2006; Schneiderheinze, 2006) compared to the non-divaricates, with many divaricate traits not typical of plants adapted to high light environments. The costs of leaf area support, in resources expended to develop the branching structure, are so high that benefits to divaricates from avoiding photoinhibition, such as net carbon gain, are offset (Christian et al., 2006), suggesting photoinhibition is unlikely to be the primary factor in the evolution of divaricates as it offers little advantage.

Another proposed hypothesis for the evolution of the divaricate form is seed dispersal through lizard frugivory (Whitaker, 1987). Many divaricates bear fruits within the canopy of densely interlacing branches that are inaccessible to most New Zealand birds, as New Zealand lacks native frugivorous birds smaller than 15 cm (Lord & Marshall, 2001); even extinct frugivorous birds were also larger than this. Conversely lizards are smaller and can easily move through divaricate branches, which in turn, often can provide a relatively safe basking site (Lord & Marshall, 2001), such as has been observed in the shrub *Larrea divaricata* (Zygophyllaceae) in Argentina (Deviana et al., 1994). Lizards have a preference for white or blue coloured fruits (Wotton et al., 2016). A comparison of New Zealand shrubs with fleshy fruits observed a significant correlation between white and blue fruit colours, growth form and small fruit size, supporting the hypothesis that lizard frugivory may be involved in the evolution of fruit colour in these shrubs (Lord & Marshall, 2001). However, whether this is a result of the divaricate form or had influence on the divaricate form is unknown. Lizards are known to browse on divaricate fruits and have been identified as important dispersers in New Zealand,

especially in shrublands, but seeds are not dispersed large distances by lizards and small birds are known to disperse divaricate seeds also (Wotton et al., 2016). Weta have also been documented to consume and disperse fruits from New Zealand shrubs (Burns, 2006; Duthie et al., 2006) and may also have played a role in the evolution of fruits in New Zealand. However, Morgan-Richards et al. (2008) observed that although weta consumed fruits from two species, *Gaultheria depressa* (Ericaceae) and *G. antipoda*, when offered, they did so in a way to avoid most seeds and likely did not act as seed dispersers. However, more work is needed to determine if weta can act as seed dispersers and if they have co-evolved with divaricate plants (Burns, 2008).

1.4 Heteroblasty

Heteroblasty is another unique feature of the New Zealand flora and is often associated with divarication (Philipson, 1964; Zotz et al., 2011). It is defined as a substantial difference between the juvenile and subsequent adult phase (Goebel, 1905). Godley (1975, 1979 & 1985) proposed that heteroblasts arose from hybridization between two species, for example, in the case of divaricate heteroblasts the parents would consist of a divaricate and non-divaricate species to give rise to species with a divaricating juvenile and arborescent adult. This hypothesis was investigated in the heteroblast *Pittosporum turneri* Petrie. (Pittosporaceae) where it was proposed *P. tuneri* arose from a hybrid origin between the divaricate *P. divaricatum* Cockyane. and the non-divaricate *P. colensoi* Hook.f. although further work is needed to support this hypothesis (Carrodus, 2009).

The developmental changes in leaf morphology may help the developing plant cope with a changing environment (McGlone & Webb, 1981; Day, 1998; Winn, 1999; Darrow et al., 2002) especially plants growing in a forested environment. Heteroblasty was proposed to have evolved as a strategy to cope with the change in low light intensity

below the canopy to a high light intensity above the canopy (Clearwater & Gould, 1995; Day, 1998). Alternatively, Day (1998) suggested that the divaricate form may have evolved from the transition into more open habitats and the subsequent loss of the adult plant form. It has also been suggested that adaptation to avoid herbivory may have contributed to the evolution of heteroblasty where the juvenile form discourages browsing (Greenwood & Atkinson, 1977). An example of this is seen in *Elaeocarpus hookerianus* Raoul (Elaeocarpaceae) where juvenile leaves have low chromatic and achromatic contrasts against a leaf litter background, which may make them hard to distinguish for browsing birds (Fadzly & Burns, 2010). However, once the plants grow above 3m, which is believed to be above the reach of browsing birds such as the tallest moa (Greenwood & Atkinson, 1977), the leaves become a more common shape and colour (Allan, 1961).

A comparison between heteroblastic species on the Chatham Islands and New Zealand main islands revealed the heteroblastic species from New Zealand, *Muehlenbeckia australis* Meisn. (Polygonaceae), *Plagianthus regius* (Poit.) Hochr. (Malvaceae), *Pseudopanax chathamica* Kirk. (Araliaceae), and *Pseudopanax crassifoulius* K. Koch. (Araliaceae), had little or no difference in juvenile and adult forms on the Chatham Islands compared to the heteroblastic forms on the main islands. They suggest the reversion of heteroblasty of these species on the Chatham Islands aligns with the moa theory (Burns & Dawson, 2009) as moa never reached the Chatham Islands. However, as McGlone and Webb (1981) suggested, islands often have milder climatic conditions than the larger land masses and a more favourable climate may have resulted in a reversion from heteroblasty.

The different hypotheses for the origins of divarication are still debated. Because moa are extinct and the environmental conditions are different at present, it is difficult

to determine the exact reasons for the evolution of the divaricating growth form. Gaining an understanding of the habitat, traits and genetic bases behind the divaricate form will provide further insights on a unique feature of the New Zealand flora and may help determine whether the genetic architecture of this growth form is shared or unique to each species.

1.5 The genetics of divaricate traits

To our knowledge, no genetic studies on the divaricating growth form have been performed previously therefore nothing is currently known about the genetic basis of this relatively common growth form of the New Zealand flora. However, certain traits indicative of divaricates have been studied in other species and the genetic underpinning of these may be similar to those in the corresponding divaricate analog.

1.5.1 Apical dominance

Weak apical dominance (Figure 1.3) is often a feature of divaricate plants (Went, 1971; Greenwood & Atkinson, 1977; Dawson, 1988; Atkinson, 1992) where lateral shoots are able to develop resulting in a bush-like plant instead of an arborescent plant form. Apical dominance has been thought to be controlled by the flow of auxin from the growing apical bud which inhibits the growth of the axillary (or lateral) buds. If this flow is interrupted, for example the shoot tip is damaged, the axillary buds break dormancy. Auxin is a plant hormone involved in many plant processes including leaf, shoot, root and fruit development (Eichhorn et al., 2005).

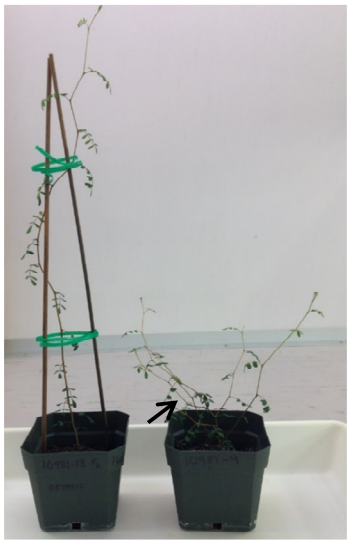


Figure 1.3. F_2 seedlings of Sophora grown in a growth room. The seedling on the left displays high apical dominance with one main shoot and no branching while the seedling on the right displays little or no apical dominance with high levels of axillary branching. The leading shoot is identified with an arrow.

In plants with normal arborescent growth, the lowermost axillary buds of the main shoot are the first to break dormancy. Auxin is formed in the shoot apex and diffuses down the shoot to the older sections where the concentration is lower resulting in potential growth of the axillary buds. A high auxin concentration promotes dormancy of the axillary buds (Thimann & Skoog, 1933). Dawson (1963) observed the juvenile divaricate of *Carpodetus serratus* J. R. Forst. & G. Forst. (Escalloniaceae) develops differently to the adult arborescent growth form. After seven nodes have formed the axillary buds closest to the apex begin to grow and the axillary buds on the new branches start to grow soon after. The buds further from the apex have less growth and the lowermost buds initially don't have any growth at all. Dawson (1963) suggests there may be a mutation in auxin production or its effects that result in growth from the axillary buds closest to the apex.

However, recently sucrose was shown to be important in the control of apical dominance, in *Arabidopsis*. Sucrose rapidly increases in the axillary buds after decapitation, whereas the levels of auxin do not change until after bud release (Mason et al., 2014). Artificially increasing the sugar concentration was also able to repress expression of *BRC1*, which normally acts to repress bud release, indicating that accumulation of sugars in the buds can lead to promotion of bud release. Mason et al. (2014) proposed that the shoot's demand for sugars may inhibit axillary bud growth by limiting the availability of sugars for those buds. Therefore, apical dominance may be controlled by a combination of auxin inhibition and sugar availability as a consequence of the growth from the apex (Van den Ende, 2014; Kebrom, 2017).

Studies of apical dominance in plants have revealed that the potential for axillary bud growth appears to be influenced by the balance between several hormones: auxin, cytokinin, strigolactone and abscisic acid (Shimizu-Sato & Mori, 2001). Cytokinin is

predominantly found in actively dividing tissues of the plant and has a role in promoting cell division (Eichhorn et al., 2005). Application of exogenous cytokinin to lateral buds causes the buds to grow even in the presence of auxin (Wickson & Thimann, 1958; Faiss et al., 1997; Mueller & Leyser, 2011) suggesting cytokinin may play a role in reduced apical dominance and could be important in the divaricating habit. A study of *Elaeocarpus hookerianus* Raoul. (Elaeocarpaceae), a heteroblast with a divaricate juvenile form, by Day, Jameson and Gould (1995) compared cytokinin content in leaves of pot grown and forest grown material of adult and juvenile plants. There was a trend of decreasing active cytokinin between juvenile and adult. Also, the juvenile plants had a lower proportion of storage cytokinins relative to total cytokinin content compared to adult plants (Day et al., 1995). However, they noted further work was required to determine if this is linked to divarication.

Exogenous application of cytokinins has been shown to increase the divaricate form of juvenile *Sophora microphylla* (Fabaceae), a heteroblastic tree, which had more outgrown lateral branches, wider branch angles and narrower node angles than control plants (Carswell et al., 1996) suggesting cytokinins may play a role in regulation of the divaricate form in this species. They also observed a higher ratio of active to storage cytokinins in juvenile plants, consistent with the findings of Day et al. (1995) in *E. hookerianus* (Elaeocarpaceae).

Once an axillary bud is formed, it enters a transition phase. At this phase the buds are sensitive to signals that promote (cytokinin) or inhibit (strigolactone or auxin) bud growth. Dormant buds must pass through the transition stage before they can break dormancy and begin outgrowth. Buds in the transitional stage may switch between being more or less sensitive to the factors either promoting or inhibiting growth (Stafstrom & Sussex, 1988; Napoli et al., 1999; Cline et al., 2001; Shimizu-Sato & Mori, 2001; Beveridge

et al., 2003; Morris et al., 2005). Dormancy of a bud depends on the interaction of conditions such as the characteristics of the bud, including size and maturity, environmental factors, such as temperature and light, and mechanisms of inhibition (Horvath et al., 2003; McSteen & Leyser, 2005).

Apical dominance also has been studied at the genetic level. For example, a gain of function mutant of the *axr3-1* gene in *Arabidopsis* had increased amplification of auxin responses and greater stability of the AXR3-1 protein (Leyser et al., 1996). This mutant presented as smaller in size and with increased apical dominance. The gene mutant *axr3-1* is from the Aux/IAA protein family (Rouse et al., 1998), which acts as transcription regulators, indicating this gene mediates responses to auxin in *Arabidopsis* (Cline et al., 2001). A maize mutant showed increased tiller growth, ear number and elongation of ear branches indicating loss of apical dominance; Grassy – tillers 1 (*gt1*) was identified as the gene responsible for the loss of apical dominance in these maize mutants. The gene mutant *gt1* is a class 1 homeodomain leucine zipper that stimulates dormancy of lateral buds (Whipple et al., 2011).

1.5.2 The Strigolactone Pathway

Tomlinson (1978) studied features of many divaricate species and concluded that this growth form results from a lack of organisational control of branch development. Meristem initiation and development have been considered important in plant architecture and so the mechanisms and pathways involved in these have been of interest in the study of plant architecture. Many genes involved in these pathways, such as the strigolactone, auxin and cytokinin pathways, have been identified and play a role in plant architecture.

One pathway known to be involved in branching is the Ramosus pathway (Figure 1.4). The *ramosus* (*rms*) mutations in pea plants (*Pisum*, Fabaceae) are particularly

relevant to the divaricate growth form. These mutations show reduced apical dominance by increased branching on shoots (Morris et al., 2001). To date seven independent *RAMOSUS* mutations have been identified showing increased branching phenotypes; *rms1, rms2, rms3, rms4, rms5, rms6* and *rms7.* The genes underlying the *rms1, rms2 rms3, rms4* and *rms5* mutants have been identified and form part of a pathway (Figure 1.4, Table 1.3) that inhibits branching from axillary buds.

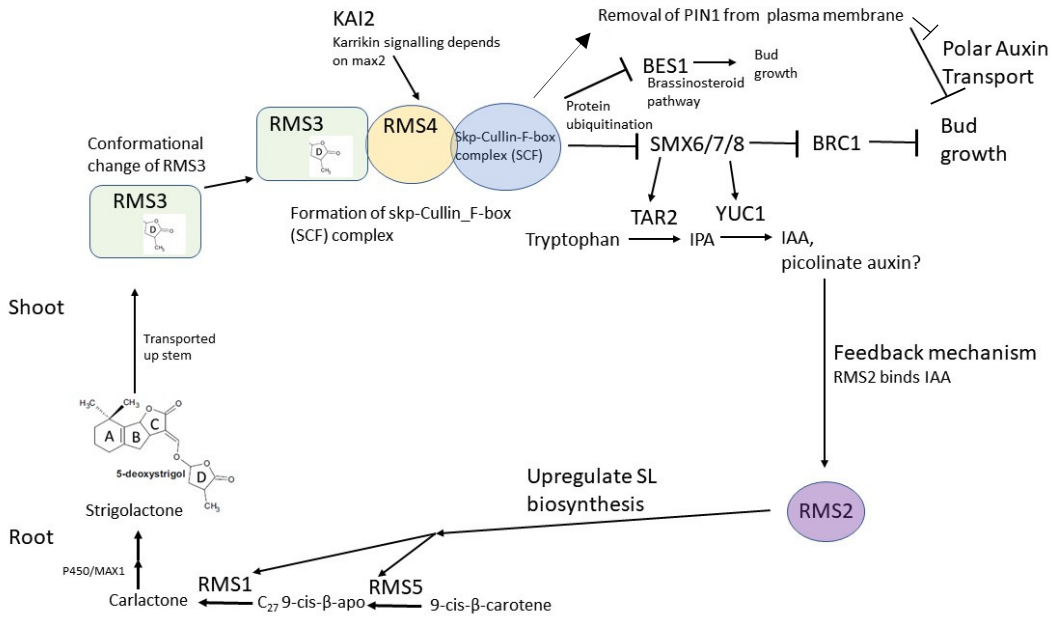


Figure 1.4. The current model of the RMS pathway (modified from Beveridge, 2006; Dun et al., 2009; Braun et al., 2012; Ligerot et al., 2017). *RMS1* and *RMS5* are required together to produce a novel long distance signal, strigolactone, which can only move unidirectional up the shoot. *RMS3* hydrolyses the ABC lactone and D ring of strigolactone and binds to the D ring causing a conformational change. This allows the formation of the SCF complex by binding to *RMS4*. *RMS3* and *RMS4* are both required together and are involved in signal perception The SCF complex targets proteins for degradation including BES1, involved in the Brassinosteroid pathway, and SMX6/7/8. *RMS4* also integrates signals from *KAI2*, involved in the karrikin pathway, and removal of PIN1 is associated with strigolactone signalling. The degradation of *SMX6/7/8* removes the repression of *BRC1,* which acts as a repressor to bud growth. *RMS4* also acts to repress *RMS1,* independently of the long distance signals, by suppressing the feedback signal involving *RMS2* (Quine, 1990; Dun et al., 2009). When signalling from strigolactone is lacking, SMX6/7/8 upregulate biosynthesis of auxin. *RMS2* binds to auxin and acts to upregulate *RMS1* and *RMS5* in the rootstock and control cytokinin levels in the shoot, which inhibits promotion of axillary bud growth as cytokinin in known to be involved in promotion of branching of axillary buds. *RMS3* and *RMS4* also act independently in the root to inhibit *RMS1* and *RMS5* expression.

Table 1.3. Summary of the genes involved in strigolactone signalling.

Orthologs	*RMS1* MAX4, DAD1[11]	*RMS2* AFB4/5[5]	*RMS3* D14[7, 8]	*RMS4* MAX2[11]/D3[12, 13]	*RMS5* MAX3[11]	*BRC1* AtBRC1	*SMX6/7/8*
Protein class	Carotenoid cleavage dioxygenase (CCD8)	F-box (auxin receptor)[5]	α/β hydrolase[7]	F-box protein[11,13, 14]	Carotenoid cleavage dioxygenase (CCD7)		
Mutant phenotype	Increased branching, shorter leaflets/leaf blades, shorter petioles, thinner stems[1,2,3]	Increased shoot branching especially at basal nodes[5] reduction in internode length[23]	Increased branching, shorter plant height[7]	Increased branching[11]	Increased branching, shorter stems, smaller leaves[18, 19]		In *Arabidopsis* larger petioles and leaf blade length[16]
Role	Biosynthesis of Strigolactone - cleaves C_{27} 9-cis-β-apo into intermediate Carlactone containing butenolide ring [4]	Lack of SL response stimulates IAA biosynthesis, RMS2 binds auxin and upregulates RMS1 and RMS5 as a feedback signal[5]	Hydrolyse the bond between ABC lactone and D ring of Strigolactone forming a covalent bond with the d-ring complex on to the histidine of catalytic triad causing a conformational change enabling binding to the RMS4 complex[7,9]	Component of SCF complex[13, 14] involved in protein ubiquitination of SMX6/7/8[5, 15, 16, 17]	Biosynthesis of strigolactone - cleaves 9-cis-β-carotene into two products: C_{13}-β-ionone and C_{27} 9-cis-β-apo[4]	Repress axillary bud growth[20.] BRC1 works in concert with auxin regulation and feed-forward loop with strigolactone where it reduces likelihood of bud release epistatic to SMX6/7/8[21]	Epistatic to BRC1[21] Repress BRC1 to promote branching, GR24 triggers rapid degradation of SMXs[16]. Also upregulate Tar2 and YUC1 to synthesise auxin creating the feedback loop[5]
Regulated by	Upregulated by feedback loop involving RMS2[5]				Upregulated by feedback loop involving RMS2[5]	SMX/6/7/8	SCF complex containing RMS4 and RMS3[5, 15, 16, 17]
Expression	Highest in the roots, low expression in internode and epicotyl[6]		In phloem sap, transported into the axillary buds[10]	Throughout the plant but required at each node to repress bud growth[11, 13]	Highest in the roots[18], low expression in internode, apex and epicotyl[11]	Exclusively in axillary bud[22]	Primarily rosette buds, expanding leaves, vasculature of hypocotyl, cotyledon and mature roots[16]

[1] (Auldridge et al., 2006), [2] (Sorefan et al., 2003), [3] (Beveridge et al., 1997),[4] (Flematti et al., 2016), [5](Ligerot et al., 2017), [6] (Foo et al., 2005), [7] (de Saint Germain et al., 2016), [8](Kameoka et al., 2016), [9] (Yao et al., 2016), [10] (Kameoka et al., 2016), [11] (Johnson et al., 2006), [12] (Dun et al., 2009), [13] (Stirnberg et al., 2007), [14](Woo et al., 2001), [15] (Bennett & Leyser, 2014) , [16](Soundappan et al., 2015), [17] (Jiang et al., 2013), [18] (Booker et al., 2004), [19](Morris et al., 2001), [20](Gao et al., 2016), [21](Seale et al., 2017), [22](Martin-Trillo et al., 2011) [23] (Abbott, 1992)

RMS1 and *RMS5* are orthologous to *MAX4* and *MAX3*, respectively, in *Arabidopsis* (Sorefan et al., 2003; Johnson et al., 2006). These encode carotenoid cleavage dioxygenases, CCD8 and CCD7, respectively, (Sorefan et al., 2003; Booker et al., 2004), which are involved in biosynthesis of the plant hormone strigolactone (SL) (Morris et al., 2001; Gomez-Roldan et al., 2008) (Figure 1.4). These enzymes oxidatively cleave carotenoids at specific positions (Schwartz et al., 2001; Sorefan et al., 2003; Booker et al., 2004) to produce apocarotenoids which are important for many biological functions in plants including growth, through inhibition or promotion of buds, predator repellent and chemoattractants (Beltran & Stange, 2016). *RMS5 Pisum* mutants show a highly branched phenotype with shorter stems and less expanded leaves (Morris et al., 2001). These mutants are restored to wild-type by exogenous application of strigolactone (GR24). *RMS5* is mainly expressed in the root tissue (Booker et al., 2004; Johnson et al., 2006) where it acts to cleave 9-cis-β-carotene into two products: C_{13}-β-ionone and C_{27} 9-cis-β-apo (Flematti et al., 2016). RMS1 then cleaves C_{27} 9-cis-β-apo into the intermediate Carlactone containing a butenolide ring (Flematti et al., 2016) and this product is further modified to form strigolactone. *RMS1* also has highest expression in root tissue (Foo et al., 2005). *Pisum RMS1* mutants are highly branching with shorter internodes, thinner stems and shorter leaflets (Beveridge et al., 1997) and this phenotype is also rescued by application of GR24. The synthesised strigolactone is transported unidirectionally from the roots, up the shoot, where it is perceived by *RMS3*.

RMS3 is orthologous to D14 in rice which is expressed in the phloem sap and subsequently transported to the axillary buds (Kameoka et al., 2016). RMS3 is classed as an α/β hydrolase which possess a conserved catalytic triad of Asp, Ser and His and it functions to hydrolyse the bond between the ABC lactone and the D ring in strigolactone. The D ring is then covalently bound to RMS3, attaching to the His of the catalytic triad (de Saint Germain et al., 2016) and causes a conformational change (Yao et

al., 2016) that enables binding of the RMS3 plus D ring complex to RMS4 (de Saint Germain et al., 2016; Yao et al., 2016). *RMS4* is orthologous to *MAX2* of *Arabidopsis* which is expressed in all tissues of *Arabidopsis* with strongest expression in the vasculature of growing leaves, flowers, siliques, stems and buds and is required at each node for repression of branching (Stirnberg et al., 2007). It is classed as an F-box protein (Johnson et al., 2006). F-box proteins have two domains; one to bind the substrate which is likely the C-terminal leucine rich domain in MAX2 and the F-box domain (N terminus) which is required for assembly of the SCF complex by binding to Skp1 (Woo et al., 2001; Kuroda et al., 2002; Stirnberg et al., 2007). The SCF complex is involved in protein ubiquitination targeting proteins for degradation. The formation of the complex by binding of RMS3 promotes degradation of downstream targets such as SMXLs (Jiang et al., 2013; Bennett & Leyser, 2014; Soundappan et al., 2015). SMX6/7/8 function to repress *BRC1* expression in axillary buds (Soundappan et al., 2015) providing the potential for the buds to develop as BRC1 acts to repress branching. The SCF protein complexes have also been associated with many hormone pathways in *Arabidopsis* such as those of ethylene (Stepanova & Alonso, 2005) and gibberellins (Dill et al., 2004).

After targeted degradation of the SMX's by the SCF complex, BRC1 expression, exclusively expressed in axillary buds, is no longer repressed and therefore acts to repress bud growth, likely integrating signals from different pathways (Martin-Trillo et al., 2011; Gao et al., 2016; Seale et al., 2017). *BRC1* is not the only gene involved in repressing bud growth as strigolactone was able to suppress branching even in buds lacking *BRC1* expression and buds with high *BRC1* expression are still able to develop (Seale et al., 2017). As well as the SMX proteins *RMS4* interacts with *BES1*, which is involved in the brassinosteriod pathway, (Wang et al., 2013) and *KAI2,* involved in the karrikin signalling pathway (Waldie et al., 2014) which is involved in germination, photomorphogenesis and leaf development (Waters et al., 2017). Furthermore, strigolactone signalling results

in the removal of PIN1 from the plasma membrane (Shinohara et al., 2013), affecting auxin flow, indicating there may be many targets downstream of *RMS4* involved in repressing bud growth.

The *SMX6/7/8* genes have a further role in strigolactone signalling as they are involved in a feedback loop where a lack of signal by strigolactone promotes biosynthesis of auxin (Ligerot et al., 2017) and affects the level of cytokinin in the stem (Beveridge et al., 1997; Beveridge, 2000; Morris et al., 2001; Foo et al., 2007). *SMX6/7/8* upregulate the genes *TAR2* and *YUC1* resulting in biosynthesis of auxin which is then bound by *RMS2*. *RMS2* encodes an F box protein, called AFB4/5, an auxin receptor that binds auxin and upregulates *RMS1/5* (Peng et al., 1999; Dun et al., 2009; Braun et al., 2012; Ligerot et al., 2017).

It will not be lost on the reader that the strigolactone pathway and the *rms* mutants in particular have considerable relevance to the divaricating growth form, including branching frequency, and leaf/leaflet size. The pathway is known to be conserved across diverse plant families with orthologs identified in *Pisum, Arabidopsis, Petunia* and maize (Johnson et al., 2006). These factors indicate that this pathway may be relevant to the divaricating growth form and represents a potentially rich source of candidate genes.

1.5.3 Other candidate genes for divaricate traits

Bushy

A dominant single-gene mutant in pea, named *bushy,* is independent of the *ramosus* mutations and results in a phenotype characterised by tiny leaves, short thin stems and increased basal branching (Symons et al., 1999). Other traits were also affected including shorter internodes, an altered root morphology, and slight increase to time of first open flower. This mutant has a reduced level of free auxin of about 12-fold

compared to wild-type (although this only occurs after 10-14 days of growth where the mutant phenotype begins to show) however the total auxin content (free and conjugated forms) is no different from the wild-type (Symons et al., 2002). Many of these traits, such as smaller leaves, short thin stems and increased branching, are also seen in the divaricate growth form. Along with some of the *ramosus* mutants, this mutant demonstrates that many of these traits may be produced from a single mutation, indicating the divaricate growth form (or at least the common traits between divaricates from different species, e.g. thin stems, smaller leaves and interlacing branching) could have arisen from a mutation in primary sequence or expression of a single gene. While mutations such as this are important for the development of genetic hypotheses for the divaricating habit, the gene behind this particular mutation has not yet been identified.

Compact peach

A cultivar of peach tree, Redhaven, gave rise to a mutant growth form, the compact peach growth habit (Van Well, 1974). Compared to the normal peach, this growth form has wider branching angles, shorter internodes, longer branches, smaller but more numerous leaves and a high degree of lateral branching, indicating weak apical dominance (Scorza, 1984; Mehlenbacher & Scorza, 1986; Scorza, 1989). This growth form has a Mendelian segregation ratio that is expected of a single dominant locus, named *ct* (Scorza et al., 2002), but it has not yet been mapped or identified (Hollender & Dardick, 2015). Compact peach results in a dense canopy reducing light penetration and is not suitable for fruit production, which is the main interest for research into these tree architectures, but the similarities to the divaricate growth form make it a locus of interest, particularly as it appears to be due to a dominant allele that affects multiple phenotypes. Hollender and Dardick (2015) suggest that the allele causing this phenotype in peach may be involved in branching inhibition involving the SL pathway (or biosynthesis of SL) as these mutants result in a bushy appearance and are highly

branching like the peach compact growth form. As this growth form has similar features to the divaricate growth form it further suggests that the divaricate form may arise from changes to relatively few genes.

Branch angle, Leaf size

Other divaricate-relevant traits that have been studied in plant species include branch angle and leaf size. For example, Liu et al. (2016) identified four QTL with good candidate genes in *Brassica napus* L. (Brassicacae) involved in branch angle. These genes include *Brassica* orthologs of genes identified as playing key roles in branch development in *A. thaliana, BnaA.Lazy1* and *BnaC.Lazy1*, *SQUAMOSA PROMOTER BINDING PROTEIN LIKE 14* (*SPL14)* and an auxin responsive *GRETCHEN HAGEN 3 (GH3)* gene. In maize, a TCP (TEOSINTE BRANCHED1/CYCLOIDEA/PROFIFERATING CELL NUCLEAR ANTIGEN FACTOR) member BRANCH ANGLE DEFECTIVE 1 (BAD1) has been identified as influencing branch angle by promoting cell proliferation in the pulvinus, leading to a reduction in branch angle in the mutant (Bai et al., 2012). Reduced leaf size is a common trait of divaricating plants. A study in *Arabidopsis* identified five genes, *AVP1, GRF5, JAW, BRI1,* and *Ga20OX1,* that all increased leaf size when overexpressed. Each of the five genes affected leaf size differently but all included an increase in cell number. Along with physiological data this suggests leaf size is controlled by multiple independent pathways indicating leaf size is a genetically complicated trait (Gonzalez et al., 2010).

ZIG

Another potential candidate gene, identified from *Arabidopsis* (Brassicaceae) mutants, is *ZIG* where the mutants show zig-zag patterned shoots, which are a feature seen in some divaricate plants, e.g. *Sophora prostrata* (Figure 1.5). The zig-zag stem in *Arabidopsis* is the result of a mutation in the *ZIG* gene and causes rosette leaves to

appear small and wrinkled as well as having a zig-zag shaped stem (Kato et al., 2002). The genomic region involved in the *zig* mutant was identified on chromosome 5 and the gene was identified as *MUL8.15* (which previously had been identified from a data base search as *AtVTI11a* (Zheng et al., 1999; Kato et al., 2002). The wild type allele, when transformed, returns mutants to wild type morphology, confirming this gene is responsible for the zig-zag stem (Kato et al., 2002). Another study on the ZIG mutant showed that a paralog of *VTI11*, *VTI12*, is capable of supressing the *zig-1* phenotype and identified the mechanism of zig-zag stem production as loss of function of the Qb-SNARE VTI11 complex, which is a complex of four SNARE proteins and is involved in membrane trafficking (Hashiguchi et al., 2010). The zig-zag stem is not seen in all divaricates, as some have recurved branching where there is development of consecutive

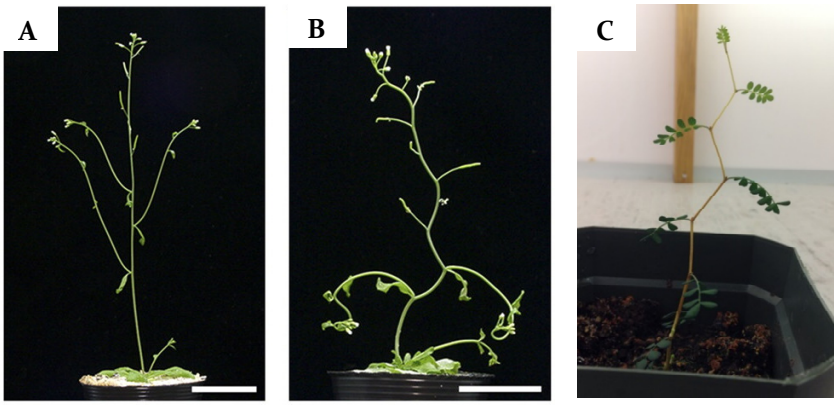


Figure 1.5. An example of *Arabidopsis* wildtype (A), *zig* mutant (B), and the zig-zag stem in the divaricate *S. prostrata* seedling (C). Images for A and B of *Arabidopsis* are from Hashiguchi et al. (2010)

buds on one side of the stem only instead of the opposite bud seen in zig-zag branches. However, VTI11 has the potential to be involved in the production of the zig-zag stem seen in some divaricate species.

These genes identified for traits analogous to those observed in the divaricating growth form, e.g., *rms1-rms5*, *AtVTI11a* and *gt1,* provide many potential candidates for genes that may be involved in the divaricate growth form and a starting point to use in

combination with QTL analysis for studying the divaricating growth form. However, no genes have yet been linked to the divaricating growth form and the involvement of any of these genes in any of the divaricate traits are merely hypotheses. Indeed, *nothing* is known about the genetic basis of the divaricating growth form. As such, this thesis begins with a quantitative genetics approach to understanding the evolution of this unique plant architecture.

1.6 Quantitative Genetics

Quantitative traits, or continuous traits, are those that are measured numerically, rather than as presence or absence as with qualitative traits. Continuous traits, such as height, usually involve multiple genetic loci. However, some of the variation seen in continuous traits is caused by environmental factors, which cause the phenotype to differ from the underlying genetics and contribute to a normal distribution seen for the trait (Lynch & Walsh, 1998). Despite this, each locus that contributes to a continuous trait usually follows traditional Mendelian inheritance. Because of the variation in phenotype, due to genetic and non-genetic factors, detecting the genetic loci involved in continuous traits requires a different approach than for discrete traits, typically using quantitative trait analyses.

1.6.1 Populations for QTL mapping

Quantitative trait loci (QTL) analysis can be used to identify regions of the genome involved in producing the trait of interest and candidate genes can potentially be identified from these regions. QTL are chromosomal locations that are believed to have functional alleles segregating which cause a significant effect on the trait of interest (Geldermann, 1975; Slater et al., 2008). To map QTL, a segregating population, with known genotype and phenotype information, is required. This typically requires a large population produced through the crossing of two different parental phenotypes for the

trait of interest. The population size required varies but generally ranges from 50 – 250 individuals although larger populations are required for high resolution mapping (Mohan et al., 1997; Collard et al., 2005). An example of a segregating population is the F_2 population (Figure 1.6) where parents with distinct phenotypes are crossed to produce an F_1 generation. The F1 individuals will contain half the genetic information from each parent and are then selfed or crossed with each other to produce the F_2 generation. Recombination and independent assortment in the F_1 result in different combinations of the parental genomes in the F_2 individuals and therefore creates associations between the phenotype and genotype which can be utilised for QTL analysis.

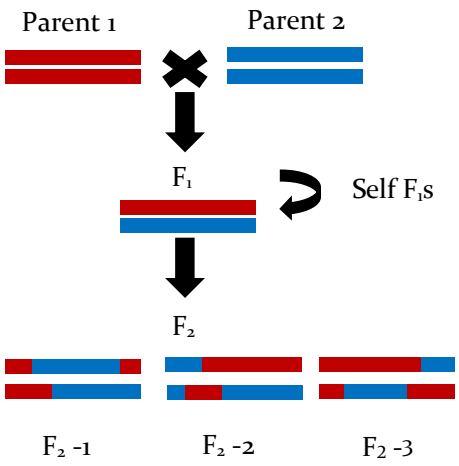


Figure 1.6. Example of the generation of an F_2 population showing one chromosome. Parent 1 genome represented by the red and parent 2 genome by the blue. F_1 inherit one copy of each chromosome from each parent. Recombination and independent assortment in the F_1 result in F_2 individuals with different combinations of the parental genome.

Backcross populations, developed by backcrossing the F_1 with one of the parents, and F_2 populations are the simplest to produce and take relatively less time to develop than other mapping populations (Collard et al., 2005) making them more suitable for species with long generational timespans or where time is restricted. Recombinant inbreed (RI) populations can be developed by inbreeding from F_2 individuals and result in homozygous lines that each contain unique homozygous chromosomal combinations from the parents; however as 6-8 generations are typically required the length of time needed to produce these populations can be a major disadvantage (Collard et al., 2005). Double haploid populations can be produced by inducing chromosomal doubling in pollen grains, however this can only be done in species where tissue culture is possible

(Collard et al., 2005). Though RI and double haploid populations produce homozygous lines that can be multiplied without changes in genetics they are not suitable for many species such as those that have long lifespans or where tissue culture is difficult. In species with long reproductive life spans, the F_1 generation is often used, instead of generating the F_2, reducing the time needed to generate a segregating population. However, using F_1 can limit the ability to detect QTL as it is limited by the amount of heterozygosity in the parents (Sewell & Neale, 2000).

As each individual in a segregating population consists of different combinations of the parental genotypes, the recombination frequencies between different loci can be calculated and this can be used to calculate the genetic distances between markers: the closer markers are physically, the less likely they are to recombine. A recombination frequency of 50% is often used as a threshold, beyond which markers are assumed to be unlinked or far apart on the same chromosome (Collard et al., 2005). When taken together, the recombination frequencies among markers can be used to develop linkage maps for the population.

1.6.2 Linkage maps

A linkage map indicates the order and relative distance between markers (Figure 1.7). Many marker types have been used to develop linkage maps, including RFLPS (restriction fragment length polymorphisms), RAPDs (random amplified polymorphic markers), SSRs (short sequence repeats) and SNPs (single nucleotide polymorphisms). RFLPs are useful markers for the construction of linkage maps but have a complicated hybridization step, typically use radioactivity and are time consuming (Mohan et al., 1997). PCR-based markers, such as RAPDs and SSRs, are relatively inexpensive and less time consuming than RFLPs. SNPs are the most abundant in the genome, but are relatively time consuming and costly (He et al., 2014). However, the advances in next-

generation sequencing have reduced the costs of large-scale analyses down to the point that GBS (genotyping by sequencing) is now possible, allowing for large-scale sequence analyses. GBS methods construct a reduced library, representative of the genome, to run on the Illumina next-generation platform, which generates large amounts of sequence data, from which large numbers of SNPs or other useful markers (e.g., indels), can be identified. Unlike RFLPs, GBS has reduced sample handling, fewer PCR and purification steps, no size fractionation, no reference sequence limitations and reduced costs (He et al., 2014) which make GBS a viable option for many studies.

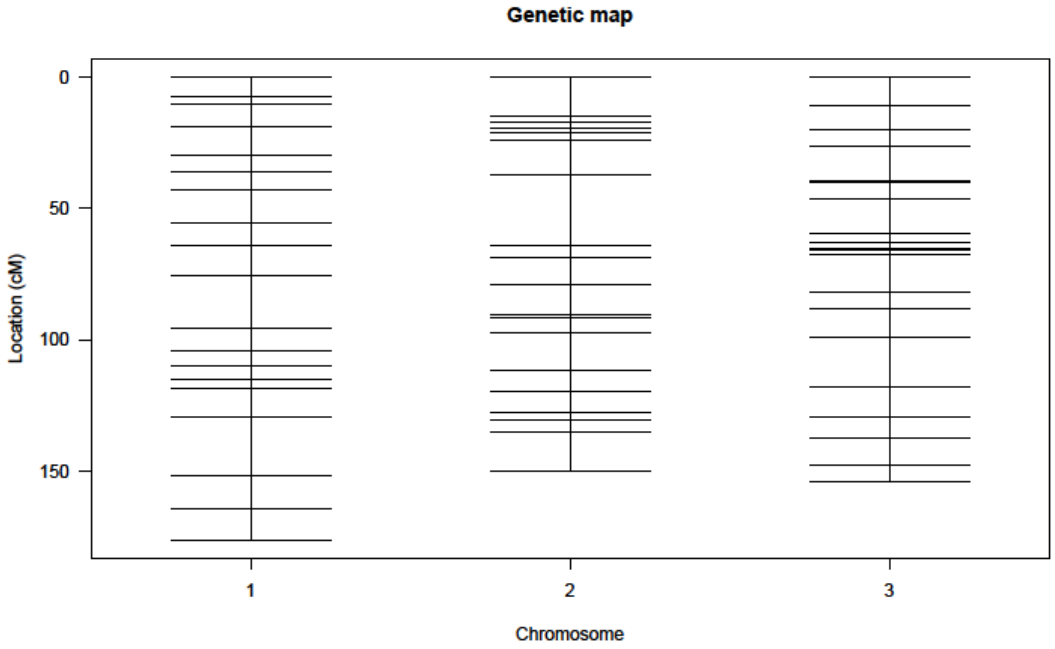


Figure 1.7. Example of a linkage map showing three linkage groups (or chromosomes). Markers are shown by horizontal lines on each linkage group and relative distance (cM) is shown on the Y axis.

The genetic markers that are required for generating linkage maps are those that are polymorphic between the parents. These will be heterozygous in the F_1 and then segregate in the F_2 (Figure 1.8). The entire population needs to be screened for these markers. Linkage analysis is usually conducted using odds ratios (e.g. ratio of linkage between markers vs no linkage between markers) and is typically called the logarithm of odds (LOD) score (Frisch et al., 1999). Values greater than 3, indicate linkage is 1000x

more likely to occur than non-linkage. Genetic distances are typically reported as units of centiMorgans (cM) where a recombination frequency of 0.01 (1%) is equal to 1cM. However, these maps are hypothetical, meaning the map distances may not correspond to actual physical distances on a chromosome as they are based on recombination frequency. Recombination frequency can be affected by factors other than physical distance, including genome size and different frequencies in chromosomal regions, with centromeres and telomeres typically having much reduced recombination (Young, 1994; True et al., 1996; Collard et al., 2005; Myers et al., 2005). Recombination frequency often equals the map distances when they are small (less than 10cM) but this relationship does not apply as distances get larger (Collard et al., 2005).

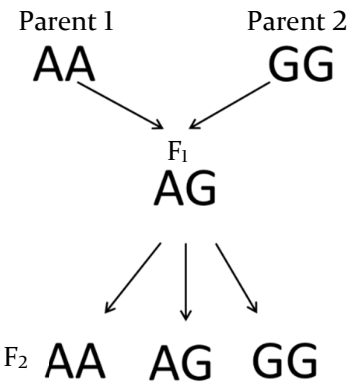


Figure 1.8. Example of a marker identified from sequencing informative for generating linkage maps. Parents are homozygous for different alleles leading to a heterozygous F1. If the marker is segregating, the F2 population will consist of homozygotes of each parental allele and heterozygotes.

The accuracy of the genetic distance of the maps is based on the number of individuals in the mapping (segregating) population. A population of at least 50 individuals is suggested as a minimum for generating linkage maps (Young, 1994; Collard et al., 2005; Singh & Singh, 2015). However, larger populations, 100-200 or more, are often recommended as a larger population has a higher number of recombination events resulting in an increased power to detect QTL (Singh & Singh, 2015). Larger population also generate more accurate maps shown by a simulation study where a drop in population size from 100 to 50 reduces accuracy of marker order from 90% to 60% (Liu, 1998). The number of markers needed varies with the number and length of chromosomes. A marker spacing of 10cM is often suggested appropriate (Singh & Singh, 2015) although simulations suggest there is only a slight decrease in power to detect QTL

with marker spacings of 20cM (Darvasi et al., 1993). However, as the positions and distances of markers are not known prior to generating the map, 100-200 markers (Lynch & Walsh, 1998; Schneider, 2005; Ferreira et al., 2006) are often suggested as a suitable number for quantitative trait analysis. The common issues when generating linkage maps include markers that are not evenly spaced throughout the genome, where some markers are clustered in regions and other regions are without markers, and the frequency of recombination is not equal along the chromosomes (Young, 1994; Collard et al., 2005). Examples of commonly used software in generating linkage maps include R/qtl (Broman et al., 2003), Windows QTL cartographer (Wang, 2010) and MapQTL (Van Ooijen, 2004).

1.6.3 Quantitative trait analysis

In addition to generating and scoring genetic markers, each individual in the segregating population needs to be phenotyped for the traits of interest before QTL analysis can be performed. QTL analysis identifies loci based on associations between phenotype and genotype where a distinct difference in phenotype for different genotypes at a locus indicates this region is involved in variation for the trait of interest. The closer a marker is to a QTL the less likely the chance of recombination, therefore the marker and the QTL will be inherited together creating an association between them (Collard et al., 2005). The most often used methods to detect QTLs are single marker analysis, interval mapping and composite interval mapping.

Single marker analysis involves testing each marker separately for significant differences in phenotype based on genotype, commonly using *t*-tests, linear regression or ANOVA. The major disadvantage is that the further a QTL is from a marker the less likely it is to be detected using single marker analysis as recombination between the marker and QTL causes the magnitude of its effect to be underestimated (Tanksley, 1993;

Collard et al., 2005). Interval mapping utilises the linkage maps analysing pairs of markers along a chromosome. This method is considered a more powerful statistical test than single marker analysis (Lander & Botstein, 1989; Collard et al., 2005) and resolves the problem with recombination between marker and QTL seen in single marker analysis. Composite interval mapping combines interval mapping with linear regression, and includes additional (estimated) markers in the model between the pair of linked markers, making it more precise when mapping QTL compared to the other methods (Jansen, 1993; Zeng, 1993; Jansen & Stam, 1994; Zeng, 1994; Collard et al., 2005). Interval mapping and composite interval mapping both provide a profile with likely positions for the QTL using the linkage map, which is usually based on the LOD score. The peak observed on the profile must be above a certain threshold for the QTL to be declared statistically significant (Collard et al., 2005).

The number of markers required for QTL analysis varies depending on the number and length of the chromosomes in the species studied. Preliminary studies for QTL often contain 100 – 200 markers (Mohan et al., 1997; Schneider, 2005; Ferreira et al., 2006) and is usually considered suitable for detecting QTL with evenly spaced markers (Collard et al., 2005). However, the number required depends on genome size with more markers required when the genome is larger (Collard et al., 2005). QTL are described as major or minor based on the amount of phenotypic variation explained with major QTLs usually describing more than 10% of the variation (Collard et al., 2005). The major factors that can affect QTL mapping include effects of a QTL, environmental factors, population size and sampling error. QTL with small effects may not be detected as they may fall below the significance threshold and closely-linked QTL may be identified as a single QTL (Tanksley, 1993; Collard et al., 2005). The larger the population the more accurate the map and the more likely to detect QTL with smaller effects, therefore, population size can have an effect on the power of QTL mapping

(Lynch & Walsh, 1998; Singh & Singh, 2015). Sampling error is mainly caused by errors in genotyping or phenotyping, including missing data, and is why accurate data collection is important to generate reliable linkage maps and for QTL analysis (Tanksley, 1993; Collard et al., 2005).

QTLs can be mapped to linkage maps as the creation of mapping populations creates associations between the markers and the QTLs. Using an F_2 population has the advantage of allowing the estimation of dominance associated with the QTL as three genotypes can be analysed for each locus compared to the homozygous background of the RILs and NILs (Lynch & Walsh, 1998). Once a QTL has been mapped to a locus, candidate genes can be identified from that region and further analyses carried out to determine the genes involved in producing the trait. Candidate genes can be developed into markers and included in the maps as individual markers. Mapping a QTL to a candidate gene marker is suggestive that the gene is involved in the trait of interest; however, there will be hundreds of other genes also linked to any one locus.

1.7 Sophora

Among the genera in New Zealand with a divaricating growth form is *Sophora* L. (Fabaceae). The genus comprises around 45 species (Hughes, 2002) found in southeast Europe, southern Asia, western South America and Australasia. There are 10 species currently recognised in New Zealand, of which eight are native species: *S. fulvida* (Allan) Heenan & de Lange, *S. godleyi* Heenan & de Lange, *S. tetraptera* J. S. Mill, *S.chathamica* Cockayne, *S. prostrata* Buchanan, *S. longicarinata* G.Simpson & J.S.Thomson, *S. molloyi* Heenan & de Lange and *S. microphylla* Aiton. Two species are introduced exotics: *S. cassioides* (Phil.) Sparre, and *S. howinsula* (W.R.B.Oliv.) P.S.Green. The New Zealand *Sophora* species are not threatened but some are classified as 'at risk' due to revegetation or horticultural planting of hybrids, exotic species and natural species planted outside

their normal range (de Lange, 2004; de Lange et al., 2017). Phylogenetic analysis using ITS sequence supports monophyly of *Sophora* sect. *Edwardsia* (Salisb). Seem., which includes all native New Zealand species of *Sophora,* but relationships among species within this group were not resolved using this marker (Mitchell & Heenan, 2002).

Relationships among New Zealand *Sophora* species have proven difficult to determine due to the low sequence variation seen in typical molecular systematics markers (Hurr et al., 1999; Maich, 2002; Mitchell & Heenan, 2002; Heenan et al., 2004; Song, 2005; Grierson, 2014; Shepherd et al., 2017) and natural hybridization that occurs in these species. *Sophora prostrata* has been identified as the most distinguished of the species in several studies (Song, 2005; Shepherd et al., 2017) but this was contradicted in another study using ISSR markers where there was admixture between *S. prostrata, S. microphylla* and *S. tetraptera* individuals (Grierson, 2014). Recently, a study using microsatellite markers showed some species, such as *S. prostrata, S. chathamica, S. fulvida* and *S. longicarinata* forming distinct groups, however, the other species *S. tetraptera, S. godleyi, S. microphylla* and *S. molloyi* did not (Heenan et al., 2018) signalling that species relationships within New Zealand *Sophora* remain problematic.

Much of the molecular research on New Zealand species of *Sophora* has focused on species relationships while research on other *Sophora* species has included some gene expression and genome work. A few studies have performed transcriptomics on different species, for example, to identify genes involved in the biosynthesis of active compounds from *S. flavescenes* Aiton. (Han et al., 2015), identifying drought stress genes in *S. moorcroftiana* Benth ex. Baker. (Li et al., 2015) and characterization of the transcriptome of *S. japonica* L. (Zhu et al., 2014). The complete mitochondrial genome for *S. japonica* has also been assembled (Shi et al., 2018) and a chloroplast genome also has been completed for *S. japonica* (Lu et al., 2018). However, there is currently no reference

genome sequenced or linkage maps developed for *Sophora* that could be useful resources to use in genetics research.

Among the species of *Sophora* in New Zealand, three are of interest in this study (Table 1.4): *Sophora prostrata* (Figure 1.9 B) is an obligate divaricate and is noted as the only divaricate with consistent compound leaves. It is distinctive compared to other divaricates as it is the only species to correspond most closely with TROLL's tree architectural model which is consistent with *Sophora* as a member of the Fabaceae (Tomlinson, 1978). *Sophora microphylla* is a heteroblastic divaricate, with a juvenile

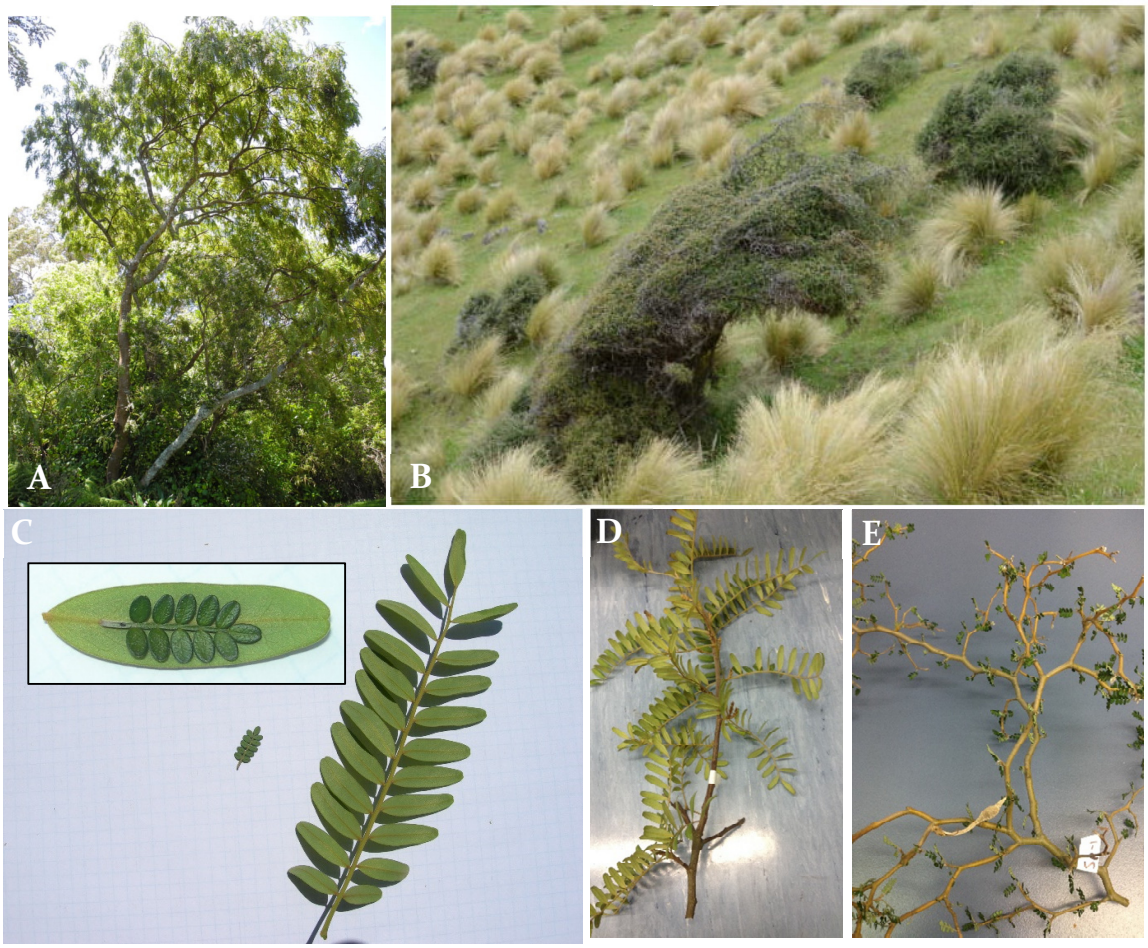


Figure 1.9. Photos of *Sophora* species characteristics. *S. tetraptera* (A) and *S. prostrata* (B). Leaf size differences between *S. tetraptera* (larger) and *S. prostrata* (smallest) are indicated in C with an inset displaying the entire *S. prostrata* leaf is smaller than an *S. tetraptera* leaflet. D displays a branch of *S. tetraptera* and E displays an *S. prostrata* branch showing the increased branching with zig-zag stem and smaller sparser leaves. Photos (A and C): Symonds, Vaughan. *S. tetraptera.* (2017). Photo (B): O2 Landscapes. *S. prostrata.* Image downloaded from: http://www.o2landscapes.com/pages/pp-sophora.php.

divaricating growth form (Heenan et al., 2001) that develops into a 10m tall arborescent adult (Allan, 1961), and *S. tetraptera* (Figure 1.9 A) is a non-divaricate (arborescent) species producing a single leading shoot that develops into the trunk (Philipson, 1964). *Sophora tetraptera* has a natural range from the eastern North Island in East Cape, south to Wairarapa and west to Taihape and Lake Taupo however, has been widely planted outside of this range. The natural habitat of *S. prostrata* is restricted to the eastern South Island while *S. microphylla* naturally ranges throughout the North and South Island (Allan, 1961; Heenan et al., 2001). All three of these species have a chromosome number of 2*n*=18 (Dawson, 2000; Stiefkens et al., 2003). However, karyotype data for *Sophora* species have shown the karyotype to be taxonomically useful as diversification of this genus seems to have involved chromosomal rearrangements (Stiefkens et al., 2001; Stiefkens et al., 2003).

Table 1.4. Features of the three species, *S. prostrata, S. tetraptera* and *S. microphylla* in the genus *Sophora*.

Species	Chromosome number	Habit	Range	Morphological features
Sophora tetraptera	2n = 18[b,d]	Arborescent[a]	Eastern North Island[c]	Tree, no juvenile form, leaves up to 150mm, leaflets up to 35 mm, 10-20 leaflet pairs[a]
Sophora prostrata	2n = 18[d]	Divaricate[a]	Eastern South Island[c]	Shrub appearance, interlacing zig-zag branches, 1-5 leaflet pairs, leaflets up to 4mm long, leaves up to 25mm[a]
Sophora microphylla	2n = 18[d]	Heteroblast[a] (Divaricating juvenile)	North and South Islands[c]	Tree with a juvenile growth form, 30-50 leaflets, leaflets up to 12.5mm petiolule present, leaflets not crowded or overlapping[a,c]

References : [a](Allan, 1961), [b](Dawson, 2000) [c](Heenan et al., 2001) [d](Stiefkens et al., 2003)

Carswell and Gould (1998) compared the development of divaricate and arborescent *Sophora* species by following the growth of potted plants for 1 year. Their work revealed that, despite *S. microphylla* juveniles and *S. prostrata* fitting the traditional definition for the divaricate form compared to the arborescent *S. tetraptera,* the differences between the divaricate and arborescent forms are small when the typical

divaricate properties are measured (morphological values often overlap). However, node angle best distinguished species and was more acute towards the top of the tree canopy. *Sophora prostrata* had significantly shorter internodes but had no significant difference in branch angle or frequency to *S. tetraptera. Sophora prostrata* also had the shortest rachis with oblong leaflets that are significantly shorter and narrower compared to the elliptic-oblong leaflets of *S. tetraptera,* which also had the longest rachis. They also observed a difference in the developmental growth between these forms. The divaricate forms, *S. prostrata* and *S. microphylla,* were characterised by two branch production stages per year: 1) sylleptic growth (growth from buds without a rest period) in spring and 2) proleptic growth (growth after the formation of a bud or period of dormancy) in summer with branches produced from the proximal node. In the arborescent *S. tetraptera,* there is only a single proleptic growth phase and the branches are produced from more distal nodes, suggesting developmental growth patterns may be used to define divaricating habit in *Sophora* (Carswell & Gould, 1998). While only node angle and internode length differed between species in the above study, a comparison between *S. prostrata* and *S. tetraptera* using the two divarication indicies developed by Atkinson (1992) and Kelly (1994), distinguished *S. prostrata* as a divaricate (Table 1.5) suggesting these measurements in these indicies may be useful in comparing these *Sophora* species (Grierson, 2014).

Sophora microphylla juveniles and *S. prostrata* have a higher ratio of active to storage cytokinins than the arborescent *S. tetraptera* and the application of 6-benzylaminopurine (Cytokinin) to arborescent *S. microphylla* resulted in a more strongly divaricating growth form (Carswell et al., 1996). *Sophora prostrata* showed less response to the 6-benzylaminopurine treatment and had lower levels of cytokinins than the other species. These results suggest that in *S. microphylla* cytokinins may play a role in the

Table 1.5. Divarication values obtained from the divarication indices IATK and IKEL for *S. prostrata* and *S. tetraptera*. The divarication threshold for IKEL is 19.2 and for IATK is 14. Table modified from results of Grierson (2014).

	Atkinson (I_{ATK})			Kelly (I_{KEL})		
Species	**Average**	**Minimum**	**Maximum**	**Average**	**Minimum**	**Maximum**
S. prostrata	18.6	15.9	26.1	24.2	20.5	31.2
S. tetraptera	9.9	6.5	14.4	14.5	11.6	18.2

divaricate growth form, but may not for all *Sophora* divaricates (Carswell et al., 1996). A correlation between cytokinin concentration and growth form has also been identified in the species *Elaeocarpus hookerianus* Raoul (Elaeocarpaceae), which also has a divaricating juvenile form and an adult arborescent form. In this species, the juvenile form also contains higher active to storage cytokinin than adults (Day et al., 1995), suggesting this may be a feature of heteroblastic divaricating species, but further research would be required. It is now known that cytokinins and strigolactones are both part of an integrated signal in controlling axillary branching and these may be important pathways to study in these heteroblastic plants to see if they play a role in the differing adult and juvenile stages.

An early hypothesis by Cockayne (1928) considered *S. prostrata* to be a permanent juvenile form of *S. microphylla.* However, the differences between the divaricate and juvenile-divaricate forms where *S. prostrata's* 'suckering' habit, smaller leaves, stouter stems and branches and lack of two phenolic compounds in the leaves compared to the *S. microphylla* juvenile (Godley, 1979, 1985), suggesting these species are too different for *S. prostrata* to be considered a permanent form of juvenile *S. microphylla.* Instead Godley proposed that *S. microphylla* arose from a hybrid origin

between *S. prostrata* and *S. tetraptera* (Godley, 1975, 1979, 1985), yet this has not been substantiated.

1.7.1 An important resource re-discovered

Approximately 45 years ago, the development of an F_2 population from *S. tetraptera* and *S. prostrata* (Figure 1.10) was started by Dr. E J Godley to investigate the origins of the heteroblast *S. microphylla*. This population was made from reciprocal crosses between *S. tetraptera* and *S. prostrata* to produce F_1 plants. Four of the F_1 plants were selfed to create F_2 populations that were planted, on the 20th May 1985, in blocks at the 'East Block' site outside Lincoln, New Zealand. The 'East Block' is comprised of the remaining parental plants and the F_2 plants (arrayed in four blocks, A-D, according to lineage (see Fig. 1.10) and situated alongside a water race. The F_1 plants are at a different site in Lincoln. The exact individuals that were used as the parental plants are not known and some of the original plants are no longer present at the site. However, plants collected from the original source sites are still present to represent the parents and were sampled in place of the exact parental plants. *Sophora prostrata* from Rakaia Gorge (G4928/29) is the only parent that had no representation remaining at the site. All F_1 plants from a particular cross direction look similar to each other but the reciprocal cross F_1s show clear phenotypic differences. The *S. tetraptera* maternal parent cross F_1 (10981 F_1) have darker leaflets and have less leaves than the 10982 F_1 (*S. prostrata* maternal parent F_1). Two F_1 individuals, 10981-23A and 10981-7, produced F_2 blocks A and B respectively. These two F_1 plants are no longer present at the study site, but two F1 plants from the same cross, 10981-3 and 10981-18 were sampled instead. Individuals 10982-7 and 10982-9 produced the F_2 C and D blocks, respectively; these are both still present and were included in our sampling. Since the F_2 population was established, the site has not always been maintained, which resulted in the site becoming overgrown. It was

cleared of overgrowth approximately one and a half years prior to our sampling, but some of the original F_2 have been lost over the decades. Following a survey of the site, the total remaining number of F_2 plants is 88 with 34 plants from blocks A and B (*S. tetraptera* maternal parent) and 54 plants from blocks C and D (*S. prostrata* maternal parent). The total number of plants per block is: A = 23, B = 11, C = 33 and D = 21.

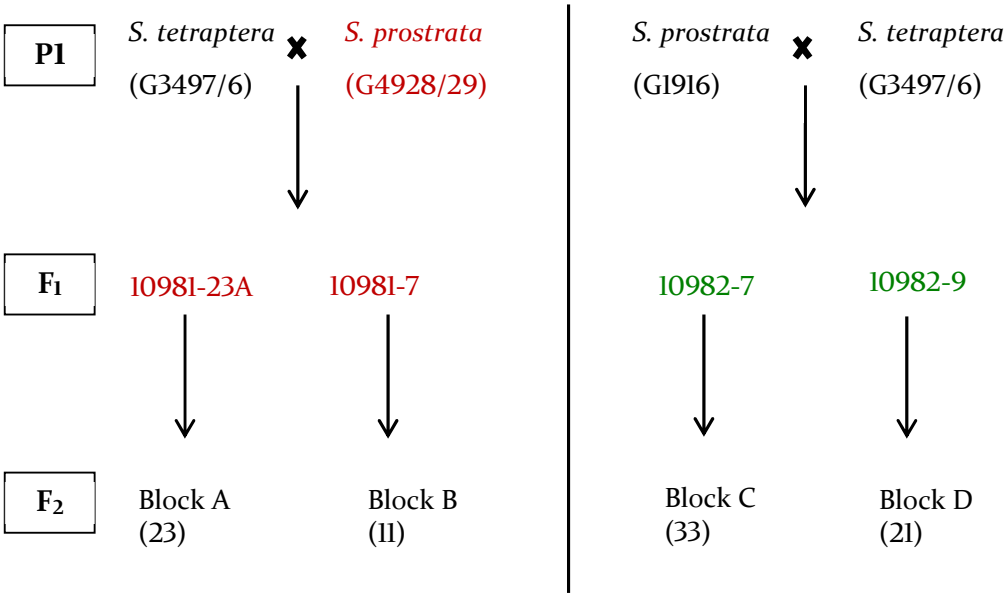


Figure 1.10. Schematic of the crosses used to produce the F2 blocks; A, B C and D. Plants in red are no longer present so could not be used in sampling. Plants in green are still present and have been sampled. For the parental plants (in black) some plants are still present at the site from the original collection sites and these have been sampled in place of the parental plants. F1 were produced from two reciprocal crosses with different *S. prostrata* individuals used in each cross (maternal parent is listed first on the left). Four different F1s were used to produce F2 in each of the four blocks A, B, C and D. Numbers in parentheses indicate the number of plants present in each block.

1.7.2 *Sophora* as a model for divarication

Sophora is an ideal genus to study the unique growth form of divarication as it contains divaricate, non-divaricate and heteroblastic (with a divaricating juvenile) species (Table 3) enabling comparison of these different growth forms among closely related species. *Sophora* species also readily hybridise with each other naturally (Heenan et al., 2001) and can also hybridise between the different growth forms, e.g. divaricate and arborescent, allowing easy creation of hybrids for subsequent experimentation. The

F_2 populations formed from reciprocal crosses between *S. prostrata* and *S. tetraptera* made by Dr. E J Godley in 1985 provide an excellent resource for developing *Sophora* into a model system for the genetic study of the divaricate growth form. Although at the outset of the thesis work there were no molecular genetic resources for *Sophora,* such as linkage maps, a reference genome, or even a good set of molecular markers, the rediscovery of the F_2 populations provided a strong foundation for studying the genetic architecture of the divaricating growth form.

1.8 Objectives

The F_2 population developed by Dr Godley is ideal to develop the first linkage maps for *Sophora* and will provide an excellent resource to study the genetic architecture of the divaricate growth form, which has not been studied previously in any one system let alone compared across the diverse groups that have divaricating species. This study will develop *Sophora* as the first model system to study the genetic basis of the divaricate plant growth form with the following objectives:

1. Develop a new system of molecular markers for use in *Sophora.*
2. Genotype the F_2 populations and develop the first genetic linkage maps for *Sophora.*
3. Isolate, sequence, and design markers from candidate genes identified from literature to incorporate into the linkage maps.
4. Phenotype the F_2 populations and utilise genotype and linkage maps to map QTL for relevant traits to describe the genetic architecture of the divaricating growth form.
5. Develop hypotheses around the evolution of the divaricating growth form based on genetic architecture.

1.9 Thesis structure

Chapter 2 describes the phenotypes utilised to describe the divaricating growth form in *Sophora* from the reciprocal crosses between *S. prostrata* and *S. tetraptera.*

Phenotypic characters measured included leaf size, leaflet number, branch width, internode length, node angle, branch angle and branch number. Trait distributions and correlations are analysed. These data are further used in QTL analysis in later chapters.

Chapter 3 explores the possible relationship between the hormone strigolactone and the divaricate growth form. Sequencing of candidate genes from the strigolactone pathway, identified from literature, in *Sophora* and the development of markers from these sequences is described. The sequencing and marker development for a further candidate gene, *ZIG,* is also described. These markers were later used in the linkage maps and QTL analysis. New phenotypes from the Lincoln F_2 population are analysed and the genotypes of the candidate genes were analysed for this population. The generation of a new F_2 population, Palmerston North (PN) F_2 in a greenhouse under controlled conditions is described. The phenotypic data and genotypic data, for one candidate gene, are described in the PN F_2 population. Lastly, application of the synthetic plant hormone of strigolactone, GR24, to *S. prostrata* seedlings is described.

Chapter 4 describes the generation of single nucleotide polymorphisms (SNP) markers and their use in generating linkage maps for *Sophora.* These maps included the SNP and markers developed from candidate genes described in Chapter 3. The chapter presents the first genetic map for *Sophora.*

Chapter 5 describes QTL mapping for the divaricate growth form in *Sophora.* This is performed using the phenotype data described in chapters 2 and 3 and the genetic maps described in chapter 4. Thirteen phenotypes are analysed for QTL in the Lincoln F_2 and the results presented.

Chapter 6 presents a summary of the thesis, drawing together the results of each chapter in a general discussion. The future directions for research on divarication are suggested and the findings of this thesis are considered relative to all divaricate species.

1.10 References

Abbott, R. J. (1992). Plant invasions, interspecific hybridization and the evolution of new plant taxa. *Trends in Ecology & Evolution, 7*(12), 401-405. 10.1016/0169-5347(92)90020-c

Allan, H. H. (1961). *Flora of New Zealand. Vol. 1. Indigenous Tracheophyta: Psilopsida, Lycopsida, Filicopsida, Gymnospermae, Dicotyledones.* Government Printer, Wellington.

Arber, A. (1928). The tree habit in angiosperms: its origin and meaning. *The New Phytologist, 252*(2)

Atkinson, I. A. E. (1992). *A method for measuring branch divergence and interlacing in woody plants*. Lower Hutt, New Zealand: DSIR Land Resources.

Atkinson, I. A. E., & Greenwood, R. M. (1989). Relationships between moas and plants. *New Zealand Journal of Ecology, 12*, 67-96.

Attard, M. R. G., Wilson, L. A. B., Worthy, T. H., Scofield, P., Johnston, P., Parr, W. C. H., & Wroe, S. (2016). Moa diet fits the bill: virtual reconstruction incorporating mummified remains and prediction of biomechanical performance in avian giants. *Proceedings of the Royal Society B-Biological Sciences, 283*(1822) 10.1098/rspb.2015.2043

Auldridge, M. E., Block, A., Vogel, J. T., Dabney-Smith, C., Mila, I., Bouzayen, M., Magallanes-Lundback, M., DellaPenna, D., McCarty, D. R., & Klee, H. J. (2006). Characterization of three members of the *Arabidopsis* carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family. *Plant Journal, 45*(6), 982-993. 10.1111/j.1365-313X.2006.02666.x

Bai, F., Reinheimer, R., Durantini, D., Kellogg, E. A., & Schmidt, R. J. (2012). TCP transcription factor, BRANCH ANGLE DEFECTIVE 1 (BAD1), is required for normal tassel branch angle formation in maize. *Proceedings of the National Academy of Sciences of the United States of America, 109*(30), 12225-12230. 10.1073/pnas.1202439109

Ball, M. C. (1994). *The role of photoinhibition during tree seedling establishment at low temperatures.* Oxford: BIOS Scientific Publisher.

Bassi, D., & Rizzo, M. (2000). Peach breeding for growth habit. In M. Geibel, M. Fischer, & C. Fischer (Eds.), *Proceedings of the Eucarpia Symposium on Fruit Breeding and Genetics, Vols 1 and 2* (pp. 411-414). Leuven 1: International Society Horticultural Science. 10.17660/ActaHortic.2000.538.72

Beltran, J. C. M., & Stange, C. (2016). Apocarotenoids: A new carotenoid-derived pathway. *Sub-cellular biochemistry, 79*, 239-272. 10.1007/978-3-319-39126-7_9

Bennett, T., & Leyser, O. (2014). Strigolactone signalling: standing on the shoulders of DWARFs. *Current Opinion in Plant Biology, 22*, 7-13. 10.1016/j.pbi.2014.08.001

Bennett, T. A., Liu, M. M., Aoyama, T., Bierfreund, N. M., Braun, M., Coudert, Y., Dennis, R. J., O'Connor, D., Wang, X. Y., White, C. D., Decker, E. L., Reski, R., & Harrison, C. J. (2014). Plasma membrane-targeted PIN proteins drive shoot development in a moss. *Current Biology, 24*(23), 2776-2785. 10.1016/j.cub.2014.09.054

Beveridge, C. A. (2000). Long-distance signalling and a mutational analysis of branching in pea. *Plant Growth Regulation, 32*(2-3), 193-203. 10.1023/a:1010718020095

Beveridge, C. A. (2006). Axillary bud outgrowth: sending a message. *Current Opinion in Plant Biology, 9*(1), 35-40. 10.1016/j.pbi.2005.11.006

Beveridge, C. A., Symons, G. M., Murfet, I. C., Ross, J. J., & Rameau, C. (1997). The *rms1* mutant of pea has elevated indole-3-acetic acid levels and reduced root-sap zeatin riboside content but increased branching controlled by graft-transmissible signal(s). *Plant Physiology, 115*(3), 1251-1258.

Beveridge, C. A., Weller, J. L., Singer, S. R., & Hofer, J. M. I. (2003). Axillary meristem development. Budding relationships between networks controlling flowering, branching, and photoperiod responsiveness. *Plant Physiology, 131*(3), 927-934. 10.1104/pp.102.017525

Blazek, J. (1982). Inheritance and genetic variation of spurred growth habit in apples. *XXIst internat. hortic. congr., Vol. 1.*, 1102-1102.

Bohle, U. R., Hilger, H. H., & Martin, W. F. (1996). Island colonization and evolution of the insular woody habit in Echium L (Boraginaceae). *Proceedings of the National Academy of Sciences of the United States of America, 93*(21), 11740-11745. 10.1073/pnas.93.21.11740

Bond, W. J., Lee, W. G., & Craine, J. M. (2004). Plant structural defences against browsing birds: a legacy of New Zealand's extinct moas. *Oikos, 104*(3), 500-508. 10.1111/j.0030-1299.2004.12720.x

Bond, W. J., & Silander, J. A. (2007). Springs and wire plants: anachronistic defences against Madagascar's extinct elephant birds. *Proceedings of the Royal Society B: Biological Sciences, 274*(1621), 1985-1992. 10.1098/rspb.2007.0414

Booker, J., Auldridge, M., Wills, S., McCarty, D., Klee, H., & Leyser, O. (2004). MAX3/CCD7 is a carotenoid cleavage dioxygenase required for the synthesis of a novel plant signaling molecule. *Current Biology, 14*(14), 1232-1238. 10.1016/j.cub.2004.06.061

Bower, F. (1908). *The origin of a land flora: a theory based upon the facts of alternation*. London: Macmillan & Co.

Braun, N., de Saint Germain, A., Pillot, J.-P., Boutet-Mercey, S., Dalmais, M., Antoniadi, I., Li, X., Maia-Grondard, A., Le Signor, C., Bouteiller, N., Luo, D., Bendahmane, A., Turnbull, C., & Rameau, C. (2012). The Pea TCP Transcription Factor PsBRC1 Acts Downstream of Strigolactones to Control Shoot Branching. *Plant Physiology, 158*(1), 225-238. 10.1104/pp.111.182725

Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics, 19*(7), 889-890. 10.1093/bioinformatics/btg112

Brunig, E. F. (1976). Tree forms in relation to environmental conditions: an ecological viewpoint. *Tree physiology and yield improvement - shoot and cambial growth.*, 139-156.

Bulmer, G. M. (1958). A key to the divaricating shrubs of New Zealand. *Tuatara, 7*, 48-61.

Burns, K. C. (2006). Weta and the evolution of fleshy fruits in New Zealand. *New Zealand Journal of Ecology, 30*(3), 405-406.

Burns, K. C. (2008). When is it coevolution? A reply to Morgan-Richards et al. *New Zealand Journal of Ecology, 32*(1), 113-114.

Burns, K. C., & Dawson, J. W. (2009). Heteroblasty on Chatham Island: a comparison with New Zealand and New Caledonia. *New Zealand Journal of Ecology, 33*(2), 156-163.

Burrows, C. J. (1980a). Diet of New Zeland Dinornithiformes. *Naturwissenschaften, 67*(3), 151-153. 10.1007/bf01073627

Burrows, C. J. (1980b). Some empirical information concerning the diet of moas. *New Zealand Journal od Ecology, 3*, 125 - 130.

Campbell, J. D., Lee, D. E., & Lee, W. G. (2000). A woody shrub from the Miocene Nevis Oil Shale, Otago, New Zealand - a possible fossil divaricate? *Journal of the Royal Society of New Zealand, 30*(2), 147 - 153.

Carlquist, S. (1974). *Island Biology*. New York: Columbia University.

Carrodus, S. K. (2009). *Identification and the role of hybridisation in New Zealand Pittosporum.* ((Unpublished masters thesis)), University of Waikato, Hamilton, New Zealand, Hamilton, New Zealand.
Carswell, F. E., Day, J. S., Gould, K. S., & Jameson, P. E. (1996). Cytokinins and the regulation of plant form in three species of *Sophora*. *New Zealand Journal of Botany, 34*(1), 123-130. 10.1080/0028825x.1996.10412699
Carswell, F. E., & Gould, K. S. (1998). Comparative vegetative development of divaricating and arborescent *Sophora* species (Fabaceae). *New Zealand Journal of Botany, 36*(2), 295-301. 10.1080/0028825x.1998.9512567
Chomicki, G., Coiro, M., & Renner, S. S. (2017). Evolution and ecology of plant architecture: integrating insights from the fossil record, extant morphology, developmental genetics and phylogenies. *Annals of Botany, 120*(6), 855-891. 10.1093/aob/mcx113
Christian, R., Kelly, D., & Turnbull, M. H. (2006). The architecture of New Zealand's divaricate shrubs in relation to light adaptation. *New Zealand Journal of Botany, 44*(2), 171-186.
Clearwater, M. J., & Gould, K. S. (1995). Leaf orientation and light interception by juvenile *Pseudopanas crassifolius* (cunn) C-KOCH in a partially shaded forest environment. *Oecologia, 104*(3), 363-371. 10.1007/bf00328372
Cline, M. G., Chatfield, S. P., & Leyser, O. (2001). NAA restores apical dominance in the axr3-1 mutant of *Arabidopsis thaliana*. *Annals of Botany, 87*(1), 61-65. 10.1006/anbo.2000.1298
Cockayne, L. (1912a). Observations concerning evolution derived from ecological studies in New Zealand. *Transactions of the New Zealand Institute, 44*, 1- 50.
Cockayne, L. (1912b). Observations concerning evolution, derived from ecological studies in New Zealand. *Transactions and Proceedings of the New Zealand Institute, 44*, 1-30.
Cockayne, L. (1928). The vegetation of New Zealand. ed. 2. *Die Vegetation der Erde, 14*
Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica, 142*(1-2), 169-196. 10.1007/s10681-005-1681-5
Cooper, A., Atkinson, I. A. E., Lee, W. G., & Worthy, T. H. (1993). Evolution of the moa and their effect on the New Zealand flora. *Trends in Ecology & Evolution, 8*(12), 433-437. 10.1016/0169-5347(93)90005-a
Czarnecki, O., Yang, J., Wang, X., Wang, S., Muchero, W., Tuskan, G. A., & Chen, J.-G. (2014). Characterization of MORE AXILLARY GROWTH Genes in *Populus*. *Plos One, 9*(7) 10.1371/journal.pone.0102757
Darrow, H. E., Bannister, P., Burritt, D. J., & Jameson, P. E. (2001). The frost resistance of juvenile and adult forms of some heteroblastic New Zealand plants. *New Zealand Journal of Botany, 39*(2), 355-363. 10.1080/0028825x.2001.9512741
Darrow, H. E., Bannister, P., Burritt, D. J., & Jameson, P. E. (2002). Are juvenile forms of New Zealand heteroblastic trees more resistant to water loss than their mature counterparts? *New Zealand Journal of Botany, 40*(2)
Darvasi, A., Weinreb, A., Minke, V., Weller, J. I., & Soller, M. (1993). Detecting marker-qtl linkage and estimating qtl gene effect and map location using a saturated genetic-map. *Genetics, 134*(3), 943-951.
Dawo, M. I., Sanders, F. E., & Pilbeam, D. J. (2007). Yield, yield components and plant architecture in the F3 generation of common bean (*Phaseolus vulgaris* L.) derived from a cross between the determinate cultivar 'prelude' and an indeterminate landrace. *Euphytica, 156*(1-2), 77-87. 10.1007/s10681-007-9354-1

Dawson, J. W. (1962). The New Zealand lowland podocarp forest. Is it subtropical? *Tuatara, 9*, 98-116.
Dawson, J. W. (1963). A comment on divaricating shrubs. *Tuatara, 11*(3)
Dawson, J. W. (1988). *Forest vines to snow tussocks: the story of New Zealand plants*. Wellington: Victoria University Press.
Dawson, M. I. (2000). Index of chromosome numbers of indigenous New Zealand spermatophytes. *New Zealand Journal of Botany, 38*(1), 47-150.
Day, J. S. (1998). Light conditions and the evolution of heteroblasty (and the divaricate form) in New Zealand. *New Zealand Journal of Ecology, 22*(1), 43-54.
Day, J. S., Jameson, P. E., & Gould, K. S. (1995). Cytokinins associated with metamorphic vegetative growth in *Elaeocarpus hookerianus*. *Australian Journal of Plant Physiology, 22*, 67 -73.
de Lange, P. J. (2004). *Sophora tetraptera. Flora of New Zealand.* Retrieved 9/06/2015 2015 from http://www.nzpcn.org.nz/flora_details.aspx?ID=1304
de Lange, P. J., Rolfe, J. R., Barkla, J. W., Courtney, S. P., Champion, P. D., Perrie, L. R., Beadel, S. M., Ford, K. A., Breitwieser, I., Schonberger, I., Hindmarch-Walls, R., Heenan, P. B., & Ladley, K. (2017). *New Zealand Threat Classification Series 22*. Wellington: Department of Conservation. https://www.doc.govt.nz/Documents/science-and-technical/nztcs22entire.pdf
de Saint Germain, A., Clave, G., Badet-Denisot, M.-A., Pillot, J.-P., Cornu, D., Le Caer, J.-P., Burger, M., Pelissier, F., Retailleau, P., Turnbull, C., Bonhomme, S., Chory, J., Rameau, C., & Boyer, F.-D. (2016). An histidine covalent receptor and butenolide complex mediates strigolactone perception. *Nature Chemical Biology, 12*(10), 787-+. 10.1038/nchembio.2147
Denny, G. (1964). *Habit heteroblastism of Sophora microphylla* M.Sc thesis. University of Canterbury.
Deviana, M. L., Jovanovich, C., & Valdes, P. (1994). Density, sex ratio and use of space of *Liolaemus darwinii* (Sauria, Iguanidae) in Valle de Tin Tin, Argentina. *Revista de Biologia Tropical, 42*, 281 - 287.
Diels, L. (1897). Vegetations biologie von Nei-Seeland. *Botanische Jahrbucher fur Systematik, Pflanzengeschichte und Pflanzengeographie, 22*, 202-300.
Dill, A., Thomas, S. G., Hu, J. H., Steber, C. M., & Sun, T. P. (2004). The Arabidopsis F-box protein SLEEPY1 targets gibberellin signaling repressors for gibberellin-induced degradation. *Plant Cell, 16*(6), 1392-1405. 10.1105/tpc.020958
Doria, L. C., Podadera, D. S., del Arco, M., Chauvin, T., Smets, E., Delzon, S., & Lens, F. (2018). Insular woody daisies (*Argyranthemum*, Asteraceae) are more resistant to drought-induced hydraulic failure than their herbaceous relatives. *Functional Ecology, 32*(6), 1467-1478. 10.1111/1365-2435.13085
Dun, E. A., Hanan, J., & Beveridge, C. A. (2009). Computational Modeling and Molecular Physiology Experiments Reveal New Insights into Shoot Branching in Pea. *Plant Cell, 21*(11), 3459-3472. 10.1105/tpc.109.069013
Duthie, C., Gibbs, G., & Burns, K. C. (2006). Seed dispersal by weta. *Science, 311*(5767), 1575-1575. 10.1126/science.1123544
Eichhorn, S. E., Evert, R. F., & Raven, P. H. (2005). *Biology of Plants* (7th ed.). United States of America: W. H. Freeman and Company.
Fadzly, N., & Burns, K. C. (2010). Hiding from the ghost of herbivory past: Evidence for crypsis in an insular tree species. *International Journal of Plant Sciences, 171*(8), 828-833. 10.1086/654850
Faiss, M., Zalubilova, J., Strnad, M., & Schmulling, T. (1997). Conditional transgenic expression of the ipt gene indicates a function for cytokinins in paracrine

signaling in whole tobacco plants. *Plant Journal, 12*(2), 401-415. 10.1046/j.1365-313X.1997.12020401.x
Feild, T. S., Arens, N. C., Doyle, J. A., Dawson, T. E., & Donoghue, M. J. (2004). Dark and disturbed: a new image of early angiosperm ecology. *Paleobiology, 30*(1), 82-107. 10.1666/0094-8373(2004)030<0082:Dadani>2.0.Co;2
Ferreira, A., da Silva, M. F., Silva, L., & Cruz, C. D. (2006). Estimating the effects of population size and type on the accuracy of genetic maps. *Genetics and Molecular Biology, 29*(1), 187-192. 10.1590/s1415-47572006000100033
Finet, C., Timme, R. E., Delwiche, C. F., & Marleta, F. (2010). Multigene Phylogeny of the Green Lineage Reveals the Origin and Diversification of Land Plants. *Current Biology, 20*(24), 2217-2222. 10.1016/j.cub.2010.11.035
Flematti, G. R., Scaffidi, A., Waters, M. T., & Smith, S. M. (2016). Stereospecificity in strigolactone biosynthesis and perception. *Planta, 243*(6), 1361-1373. 10.1007/s00425-016-2523-5
Foo, E., Buillier, E., Goussot, M., Foucher, F., Rameau, C., & Beveridge, C. A. (2005). The branching gene RAMOSUS1 mediates interactions among two novel signals 464 and auxin in pea. *Plant Cell, 17*(2), 464-474. 10.1105/tpc.104.026716
Foo, E., Morris, S. E., Parmenter, K., Young, N., Wang, H., Jones, A., Rameau, C., Turnbull, C. G. N., & Beveridge, C. A. (2007). Feedback regulation of xylem cytokinin content is conserved in pea and *Arabidopsis*. *Plant Physiology, 143*(3), 1418-1428. 10.1104/pp.106.093708
Frisch, M., Bohn, M., & Melchinger, A. E. (1999). Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Science, 39*(5), 1295-1301. 10.2135/cropsci1999.3951295x
Gao, Y. L., Song, Z. B., Li, W. Z., Jiao, F. C., Wang, R., Huang, C. J., Li, Y. P., & Wang, B. W. (2016). NtBRC1 suppresses axillary branching in tobacco after decapitation. *Genetics and Molecular Research, 15*(4) 10.4238/gmr15049320
Geldermann, H. (1975). Investigations on inheritance of quantitative characters in animals by gene markers.I. Methods. *Theoretical and Applied Genetics, 46*(7), 319-330. 10.1007/bf00281673
Godley, E. J. (1960). The botany of Southern Chile in relation to New Zealand and the Subantarctic. *Proceedings of the Royal Society Series B-Biological Sciences, 152*(949), 457-475. 10.1098/rspb.1960.0054
Godley, E. J. (1975). Kowhais. *New Zealand's Nature Heritage, 5*, 1804-1806.
Godley, E. J. (1979). Leonard Cockayne and evolution. *New Zealand Journal of Botany, 17*, 197-215.
Godley, E. J. (1985). Paths to maturity. *New Zealand Journal of Botany, 23*, 687-706.
Goebel, K. (1905). *Organography of plants. I. General Ornography,* (I. B. Balfour, Trans.): Oxford, The Clarendon.
Gomez-Roldan, V., Fermas, S., Brewer, P. B., Puech-Pages, V., Dun, E. A., Pillot, J.-P., Letisse, F., Matusova, R., Danoun, S., Portais, J.-C., Bouwmeester, H., Becard, G., Beveridge, C. A., Rameau, C., & Rochange, S. F. (2008). Strigolactone inhibition of shoot branching. *Nature, 455*(7210), 189-U122. 10.1038/nature07271
Gonzalez, N., De Bodt, S., Sulpice, R., Jikumaru, Y., Chae, E., Dhondt, S., Van Daele, T., De Milde, L., Weigel, D., Kamiya, Y., Stitt, M., Beemster, G. T. S., & Inze, D. (2010). Increased Leaf Size: Different Means to an End. *Plant Physiology, 153*(3), 1261-1279. 10.1104/pp.110.156018
Greb, T., Clarenz, O., Schafer, E., Muller, D., Herrero, R., Schmitz, G., & Theres, K. (2003). Molecular analysis of the LATERAL SUPPRESSOR gene in *Arabidopsis* reveals a conserved control mechanism for axillary meristem formation. *Genes & Development, 17*(9), 1175-1187. 10.1101/gad.260703

Greenwood, R. M., & Atkinson, I. A. E. (1977). Evolution of divaricating plants in New Zealand in relation to moa browsing. *Proceedings of the New Zealand Ecological Society, 24*

Grierson, E. R. P. (2014). *The Development and Genetic Variation of Sophora prostrata - A New Zealand Divaricating Shrub.* (Master's Thesis), The University of Waikato, Hamilton, New Zealand.

Halle, F., Oldeman, R. A. A., & Tomlinson, P. B. (1978). *Tropical trees and forests: an architectural analysis*.

Han, R., Takahashi, H., Nakamura, M., Bunsupa, S., Yoshimoto, N., Yamamoto, H., Suzuki, H., Shibata, D., Yamazaki, M., & Saito, K. (2015). Transcriptome Analysis of Nine Tissues to Discover Genes Involved in the Biosynthesis of Active Ingredients in Sophora flavescens. *Biological & Pharmaceutical Bulletin, 38*(6), 876-883. 10.1248/bpb.b14-00834

Hashiguchi, Y., Niihama, M., Takahashi, T., Saito, C., Nakano, A., Tasaka, M., & Morita, M. T. (2010). Loss-of-function mutations of retromer large subunit genes suppress the phenotype of an *Arabidopsis* zig mutant that lacks Qb-SNARE VTI11. *Plant Cell, 22*(1), 159-172. 10.1105/tpc.109.069294

He, J., Zhao, X., Laroche, A., Lu, Z.-X., Liu, H., & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science, 5* 10.3389/fpls.2014.00484

Heenan, P., Mitchell, C., & Houliston, G. (2018). Genetic variation and hybridisation among eight species of kowhai (*Sophora*: Fabaceae) from New Zealand revealed by microsatellite markers. *Genes, 9*(2) 10.3390/genes9020111

Heenan, P. B., Dawson, M. I., & Wagstaff, S. J. (2004). The relationship of *Sophora* sect. *Edwardsia* (Fabaceae) to *Sophora tomentosa*, the type species of the genus *Sophora*, observed from DNA sequence data and morphological characters. *Botanical Journal of the Linnean Society, 146*(4), 439-446. 10.1111/j.1095-8339.2004.00348.x

Heenan, P. B., de Lange, P. J., & Wilton, A. D. (2001). *Sophora* (Fabaceae) in New Zealand: Taxonomy, distribution, and biogeography. *New Zealand Journal of Botany, 39*(1), 17-53. 10.1080/0028825x.2001.9512715

Hollender, C. A., & Dardick, C. (2015). Molecular basis of angiosperm tree architecture. *New Phytologist, 206*(2), 541-556. 10.1111/nph.13204

Horvath, D. P., Anderson, J. V., Chao, W. S., & Foley, M. E. (2003). Knowing when to grow: signals regulating bud dormancy. *Trends in Plant Science, 8*(11), 534-540. 10.1016/j.tplants.2003.09.013

Howell, C. J., Kelly, D., & Turnbull, M. H. (2002). Moa ghosts exorcised? New Zealand's divaricate shrubs avoid photoinhibition. *Functional Ecology, 16*, 232-240.

Hughes, D. (2002). *Sophora* — The Kowhais of New Zealand. *Combined Proceedings International Plant Propagators' Society, 52*

Hurr, K. A., Lockhart, P. J., Heenan, P. B., & Penny, D. (1999). Evidence for the recent dispersal of *Sophora* (Leguminosae) around the Southern Oceans: molecular data. *Journal of Biogeography, 26*(3), 565-577. 10.1046/j.1365-2699.1999.00302.x

Huyghe, C. (1998). Genetics and genetic modifications of plant architecture in grain legumes: a review. *Agronomie, 18*(5-6), 383-411. 10.1051/agro:19980505

Jansen, R. C. (1993). Interval mapping of multiple quantitative trait loci. *Genetics, 135*(1), 205-211.

Jansen, R. C., & Stam, P. (1994). Hight-resolution of quantitative traits into multiple loci via interval mapping. *Genetics, 136*(4), 1447-1455.

Jiang, L., Liu, X., Xiong, G., Liu, H., Chen, F., Wang, L., Meng, X., Liu, G., Yu, H., Yuan, Y., Yi, W., Zhao, L., Ma, H., He, Y., Wu, Z., Melcher, K., Qian, Q., Xu, H. E.,

Wang, Y., & Li, J. (2013). DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature, 504*(7480), 401-+. 10.1038/nature12870

Johnson, X., Brcich, T., Dun, E. A., Goussot, M., Haurogne, K., Beveridge, C. A., & Rameau, C. (2006). Branching genes are conserved across species. Genes controlling a novel signal in pea are coregulated by other long-distance signals. *Plant Physiology, 142*(3), 1014-1026. 10.1104/pp.106.087676

Jung, J.-H., Seo, P. J., Kang, S. K., & Park, C.-M. (2011). miR172 signals are incorporated into the miR156 signaling pathway at the SPL3/4/5 genes in *Arabidopsis* developmental transitions. *Plant Molecular Biology, 76*(1-2), 35-45. 10.1007/s11103-011-9759-z

Kaggwa-Asiimwe, R., Andrade-Sanchez, P., & Wang, G. Y. (2013). Plant architecture influences growth and yield response of upland cotton to population density. *Field Crops Research, 145*, 52-59. 10.1016/j.fcr.2013.02.005

Kameoka, H., Dun, E. A., Lopez-Obando, M., Brewer, P. B., de Saint Germain, A., Rameau, C., Beveridge, C. A., & Kyozuka, J. (2016). Phloem transport of the receptor DWARF14 protein is required for full function of strigolactones. *Plant Physiology, 172*(3), 1844-1852. 10.1104/pp.16.01212

Karol, K. G., McCourt, R. M., Cimino, M. T., & Delwiche, C. F. (2001). The closest living relatives of land plants. *Science, 294*(5550), 2351-2353. 10.1126/science.1065156

Kato, T., Morita, M. T., Fukaki, H., Yamauchi, Y., Uehara, M., Niihama, M., & Tasaka, M. (2002). SGR2, a phospholipase-like protein, and ZIG/SGR4, a SNARE, are involved in the shoot gravitropism of *Arabidopsis*. *The Plant Cell Online, 14*(1), 33-46. 10.1105/tpc.010215

Kavanagh, P. H. (2015). Herbivory and the evolution of divaricate plants: Structural defences lost on an offshore island. *Austral Ecology, 40*(2), 206-211. 10.1111/aec.12196

Kebrom, T. H. (2017). A growing stem inhibits bud outgrowth - The overlooked theory of apical dominance. *Frontiers in Plant Science, 8* 10.3389/fpls.2017.01874

Keey, R. B., & Lind, D. (1997). Airflow around some New Zealand divaricating plants. *New Zealand Natural Sciences, 23*, 19-29.

Kelly, D. (1994). Towards a numerical definition of divaricate (interlaced small-leaved) shrubs. *New Zealand Journal of Botany, 32*(4), 509-518.

Kelly, D., & Ogle, M. R. (1990). A test of the climate hypothesis for divaricate plants. *New Zealand Journal of Ecology, 13*

Kuroda, H., Takahashi, N., Shimada, H., Seki, M., Shinozaki, K., & Matsui, M. (2002). Classification and expression analysis of *Arabidopsis* F-box-containing protein genes. *Plant and Cell Physiology, 43*(10), 1073-1085. 10.1093/pcp/pcf151

Lander, E. S., & Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics, 121*(1), 185-199.

Leyser, H. M. O., Pickett, F. B., Dharmasiri, S., & Estelle, M. (1996). Mutations in the AXR3 gene of *Arabidopsis* result in altered auxin response including ectopic expression from the SAUR-AC1 promoter. *Plant Journal, 10*(3), 403-413. 10.1046/j.1365-313x.1996.10030403.x

Li, H., Yao, W., Fu, Y., Li, S., & Guo, Q. (2015). De novo assembly and discovery of genes that are involved in drought tolerance in Tibetan *Sophora moorcroftiana*. *Plos One, 10*(1) 10.1371/journal.pone.0111054

Ligerot, Y., de Saint Germain, A., Waldie, T., Troadec, C., Citerne, S., Kadakia, N., Pillot, J.-P., Prigge, M., Aubert, G., Bendahmane, A., Leyser, O., Estelle, M., Debelle, F., & Rameau, C. (2017). The pea branching *RMS2* gene encodes the PsAFB4/5 auxin receptor and is involved in an auxin-strigolactone regulation loop. *Plos Genetics, 13*(12) 10.1371/journal.pgen.1007089

Liu, H. (1998). *Statistical Genomics, Linkage, Mapping and QTL Analysis*. Boca Raton, Florida: CRC.

Liu, J., Wang, W., Mei, D., Wang, H., Fu, L., Liu, D., Li, Y., & Hui, Q. (2016). Characterizing variation of branch angle and genome-wide association mapping in Rapeseed (*Brassica napus* L.). *Frontiers in Plant Science, 7* 10.3389/fpls.2016.00021

Livingstone, D. A. (1974). Extract from INQUA report to the U.S. National Academy of Sciences. *Soil News, 22*, 46-48.

Lord, J. M., & Marshall, J. (2001). Correlations between growth form, habitat, and fruit colour in the New Zealand flora, with reference to frugivory by lizards. *New Zealand Journal of Botany, 39*(4), 567-576.

Lu, Y., Li, W., Xie, X., Zheng, Y., & Li, B. (2018). The complete chloroplast genome sequence of *Sophora japonica* var. *violacea*: gene organization and genomic resources. *Conservation Genetics Resources, 10*(1), 1-4. 10.1007/s12686-017-0748-7

Lusk, C. H., & Clearwater, M. J. (2015). Leaf temperatures of divaricate and broadleaved tree species during a frost in a North Island lowland forest remnant, New Zealand. *New Zealand Journal of Botany, 53*(4), 202-209. 10.1080/0028825x.2015.1086390

Lusk, C. H., McGlone, M. S., & Overton, J. M. (2016). Climate predicts the proportion of divaricate plant species in New Zealand arborescent assemblages. *Journal of Biogeography, 43*(9), 1881-1892. 10.1111/jbi.12814

Lynch, M., & Walsh, B. (1998). *Genetics and Analysis of Quantative Traits*. Sunderland, MA, U.S.A: Sinauer Associates, Inc.

Maich, B. (2002). *A biochemical genetic evaluation of taxonomy of Sophora microphylla (Fabaceae Subfamily Papilionoideae).* (Master's thesis), Victoria University, Wellington, New Zealand.

Martin-Trillo, M., Gonzalez Grandio, E., Serra, F., Marcel, F., Luisa Rodriguez-Buey, M., Schmitz, G., Theres, K., Bendahmane, A., Dopazo, H., & Cubas, P. (2011). Role of tomato BRANCHED1-like genes in the control of shoot branching. *Plant Journal, 67*(4), 701-714. 10.1111/j.1365-313X.2011.04629.x

Mason, M. G., Ross, J. J., Babst, B. A., Wienclaw, B. N., & Beveridge, C. A. (2014). Sugar demand, not auxin, is the initial regulator of apical dominance. *Proceedings of the National Academy of Sciences of the United States of America, 111*(16), 6092-6097. 10.1073/pnas.1322045111

Mauseth, J. D. (2004). Giant shoot apical meristems in cacti have ordinary leaf primordia but altered phyllotaxy and shoot diameter. *Annals of Botany, 94*(1), 145-153. 10.1093/aob/mch121

McGlone, M. S., & Clarkson, B. D. (1993). Ghost stories: moa, plant defences and evolution in New Zealand. *Tuatara, 32*, 1-21.

McGlone, M. S., & Webb, C. J. (1981). Selective forces influencing the evolution of divaricatin plants. *New Zealand Journal of Ecology, 4*, 20-28.

McQueen, D. R. (2000). Divaricating shrubs in Patagonia and New Zealand. *New Zealand Journal of Ecology, 24*(1), 69 - 80.

McSteen, P., & Leyser, O. (2005). Shoot branching. In *Annual Review of Plant Biology* (Vol. 56, pp. 353-374). 10.1146/annurev.arplant.56.032604.144122

Mehlenbacher, S. A., & Scorza, R. (1986). Inheritance of growth habit in progenies of com-pact redhaven peach. *Hortscience, 21*(1), 124-126.

Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A., & Beeckman, T. (2008). Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. *Nature Genetics, 40*(12), 1489-1492. 10.1038/ng.253

Mitchell, A. D., & Heenan, P. B. (2002). *Sophora* sect. *Edwardsia* (Fabaceae): further evidence from nrDNA sequence data of a recent and rapid radiation around the Southern Oceans. *Botanical Journal of the Linnean Society, 140*(4), 435-441. 10.1046/j.1095-8339.2002.00101.x

Mohan, M., Nair, S., Bhagwat, A., Krishna, T. G., Yano, M., Bhatia, C. R., & Sasaki, T. (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding, 3*(2), 87-103. 10.1023/a:1009651919792

Morgan-Richards, M., Trewick, S. A., & Dunavan, S. (2008). When is it coevolution? The case of ground weta and fleshy fruits in New Zealand. *New Zealand Journal of Ecology, 32*(1), 108-112.

Moriya, S., Okada, K., Haji, T., Yamamoto, T., & Abe, K. (2012). Fine mapping of Co, a gene controlling columnar growth habit located on apple (*Malus x domestica* Borkh.) linkage group 10. *Plant Breeding, 131*(5), 641-647. 10.1111/j.1439-0523.2012.01985.x

Morris, S. E., Cox, M. C. H., Ross, J. J., Krisantini, S., & Beveridge, C. A. (2005). Auxin dynamics after decapitation are not correlated with the initial growth of axillary buds. *Plant Physiology, 138*(3), 1665-1672. 10.1104/pp.104.058743

Morris, S. E., Turnbull, C. G. N., Murfet, I. C., & Beveridge, C. A. (2001). Mutational analysis of branching in pea. Evidence that RMS1 and RMS5 regulate the same novel signal. *Plant Physiology, 126*(3), 1205-1213. 10.1104/pp.126.3.1205

Mueller, D., & Leyser, O. (2011). Auxin, cytokinin and the control of shoot branching. *Annals of Botany, 107*(7), 1203-1212. 10.1093/aob/mcr069

Myers, S., Bottolo, L., Freeman, C., McVean, G., & Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. *Science, 310*(5746), 321-324. 10.1126/science.1117196

Napoli, C. A., Beveridge, C. A., & Snowden, K. C. (1999). Reevaluating concepts of apical dominance and the control of axillary bud outgrowth. In R. A. Pedersen & G. P. Schatten (Eds.), *Current Topics in Developmental Biology, Vol 44* (Vol. 44, pp. 127-169).

Niklas, K. J. (2000). The evolution of plant body plans - A biomechanical perspective. *Annals of Botany, 85*(4), 411-438. 10.1006/anbo.1999.1100

Okano, Y., Aono, N., Hiwatashi, Y., Murata, T., Nishiyama, T., Ishikawa, T., Kubo, M., & Hasebe, M. (2009). A polycomb repressive complex 2 gene regulates apogamy and gives evolutionary insights into early land plant evolution. *Proceedings of the National Academy of Sciences of the United States of America, 106*(38), 16321-16326. 10.1073/pnas.0906997106

Ortiz-Ramirez, C., Hernandez-Coronado, M., Thamm, A., Catarino, B., Wang, M., Dolan, L., Feijo, J. A., & Becker, J. D. (2016). A transcriptome atlas of *Physcomitrella patens* provides insights into the evolution and development of land plants. *Molecular Plant, 9*(2), 205-220. 10.1016/j.molp.2015.12.002

Peng, J. R., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., Beales, J., Fish, L. J., Worland, A. J., Pelica, F., Sudhakar, D., Christou, P., Snape, J. W., Gale, M. D., & Harberd, N. P. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature, 400*(6741), 256-261.

Philipson, W. R. (1964). Habit in relation to age in New Zealand trees. *Journal of Indian Botanical Science, 42*, 167-179.

Plourde, A., Krause, C., & Lord, D. (2009). Spatial distribution, architecture, and development of the root system of *Pinus banksiana* Lamb. in natural and planted stands. *Forest Ecology and Management, 258*(9), 2143-2152. 10.1016/j.foreco.2009.08.016

Poethig, R. S. (2009). Small RNAs and developmental timing in plants. *Current Opinion in Genetics & Development, 19*(4), 374-378. 10.1016/j.gde.2009.06.001

Pole, M. (2014). The Miocene climate in New Zealand: Estimates from paleobotanical data. *Palaeontologia Electronica, 17*(2)

Pollock, M. L., Lee, W. G., Walker, S., & Forrester, G. (2007). Ratite and ungulate preferences for woody New Zealand plants: influence of chemical and physical traits. *New Zealand Journal of Ecology, 31*(1), 68-78.

Pott, C. (2014). A revision of *Wielandiella angustifolia,* a shrub-sized bennettite from the rhaetian-hettangian of Scania, Sweden, and Jameson land, Greenland. *International Journal of Plant Sciences, 175*(4), 467-499. 10.1086/675577

Pott, C., & McLoughlin, S. (2014). Divaricate growth habit in Williamsoniaceae (Bennettitales): unravelling the ecology of a key Mesozoic plant group. *Palaeobiodiversity and Palaeoenvironments, 94*(2), 307-325. 10.1007/s12549-014-0157-9

Quine, C. P. (1990). Planting at stump - preliminary evidence of its effect on root architecture and tree stability. *Research Information Note - Forestry Commission Research Division*(169), 3 pp.

Raman, S., Greb, T., Peaucelle, A., Blein, T., Laufs, P., & Theres, K. (2008). Interplay of miR164, CUP-SHAPED COTYLEDON genes and LATERAL SUPPRESSOR controls axillary meristem formation in *Arabidopsis thaliana*. *Plant Journal, 55*(1), 65-76. 10.1111/j.1365-313X.2008.03483.x

Reich, P. B., Wright, I. J., Cavender-Bares, J., Craine, J. M., Oleksyn, J., Westoby, M., & Walters, M. B. (2003). The evolution of plant functional variation: Traits, spectra, and strategies. *International Journal of Plant Sciences, 164*(3), S143-S164. 10.1086/374368

Reinhardt, D., & Kuhlemeier, C. (2002). Plant architecture. *Embo Reports, 3*(9), 846-851. 10.1093/embo-reports/kvf177

Rosati, A., Paoletti, A., Caporali, S., & Perri, E. (2013). The role of tree architecture in super high density olive orchards. *Scientia Horticulturae, 161*, 24-29. 10.1016/j.scienta.2013.06.044

Rouse, D., Mackay, P., Stirnberg, P., Estelle, M., & Leyser, O. (1998). Changes in auxin response from mutations in an *AUX/IAA* gene. *Science, 279*(5355), 1371-1373. 10.1126/science.279.5355.1371

Rowe, N., & Speck, T. (2005). Plant growth forms: an ecological and evolutionary perspective. *New Phytologist, 166*(1), 61-72. 10.1111/j.1469-8137.2004.01309.x

Salmon, J., Ward, S. P., Hanley, S. J., Leyser, O., & Karp, A. (2014). Functional screening of willow alleles in *Arabidopsis* combined with QTL mapping in willow (Salix) identifies SxMAX4 as a coppicing response gene. *Plant Biotechnology Journal, 12*(4), 480-491. 10.1111/pbi.12154

Schneider, K. (2005). *Mapping populations and principles of genetic mapping*. Weiheim: WILEY_VCH Verlag GmbH & Co KGaA.

Schneiderheinze, J. (2006). *Photoinhibition under drought and high light loads in New Zealand's divaricate shrubs.* (PhD dissertation), University of Canterbury, Canterbury,

Schwartz, S. H., Qin, X. Q., & Zeevaart, J. A. D. (2001). Characterization of a novel carotenoid cleavage dioxygenase from plants. *Journal of Biological Chemistry, 276*(27), 25208-25211. 10.1074/jbc.M102146200

Scorza, R., Bassi, D., & Liverani, A. (2002). Genetic interactions of pillar (Columnar), compact, and dwarf peach tree genotypes (vol 127, pg 254, 2002). *Journal of the American Society for Horticultural Science, 127*(4), 710-710.

Scorza, R., Lightner GW., Gilreath LE., Wolf SJ. (1984). Reduced-stature peach tree growth types: pruning and light penetration. *Acta Horticulturae, 146*, 159-164.
Scorza, R., Lightner GW., Liverani A. . (1989). The pillar peach tree and growth habit analysis of compact x pillar progeny. *Journal of the American Society for Horticultural Science, 114*, 991-995.
Seale, M., Bennett, T., & Leyser, O. (2017). *BRC1* expression regulates bud activation potential but is not necessary or sufficient for bud growth inhibition in Arabidopsis. *Development, 144*(9), 1661-1673. 10.1242/dev.145649
Sewell, M. M., & Neale, D. B. (2000). Mapping quantitative traits in forest trees. *Molecular Biology of Woody Plants, Vol 1, 64*, 407-423.
Shepherd, L. D., de Lange, P. J., Perrie, L. R., & Heenan, P. B. (2017). Chloroplast phylogeography of New Zealand *Sophora* trees (Fabaceae): extensive hybridization and widespread Last Glacial Maximum survival. *Journal of Biogeography, 44*(7), 1640-1651. 10.1111/jbi.12963
Shi, Y., Liu, Y., Zhang, S., Zou, R., Tang, J., Mu, W., Peng, Y., & Dong, S. (2018). Assembly and comparative analysis of the complete mitochondrial genome sequence of *Sophora japonica* 'JinhuaiJ2'. *Plos One, 13*(8) 10.1371/journal.pone.0202485
Shimizu-Sato, S., & Mori, H. (2001). Control of outgrowth and dormancy in axillary buds. *Plant Physiology, 127*(4), 1405-1413. 10.1104/pp.010841
Shinohara, N., Taylor, C., & Leyser, O. (2013). Strigolactone Can Promote or Inhibit Shoot Branching by Triggering Rapid Depletion of the Auxin Efflux Protein PIN1 from the Plasma Membrane. *Plos Biology, 11*(1) 10.1371/journal.pbio.1001474
Singh, B. D., & Singh, A. K. (2015). Mapping Populations. In *Marker-Assisted Plant Breeding: Principles and Practices* (pp. 125-150). New Delhi: Springer India. 10.1007/978-81-322-2316-0_5
Sinnott, E. W. (1916). Comparative rapidity of evolution in various plant types. *American Naturalist, 50*, 466-478. 10.1086/279557
Slater, A., Scott, N. W., & Fowler, M. R. (2008). *Plant Biotechnology - the genetc manipulation of plants* (2nd ed.). New York, United States: Oxford University Press Inc.
Song, J. (2005). *Genetic diversity and flowering in Clianthus and New Zealand Sophora (Fabaceae).* (Ph.D. Thesis), Massey University, Palmerston North, New Zealand.
Sorefan, K., Booker, J., Haurogne, K., Goussot, M., Bainbridge, K., Foo, E., Chatfield, S., Ward, S., Beveridge, C., Rameau, C., & Leyser, O. (2003). *MAX4* and *RMS1* are orthologous dioxygenase-like genes that regulate shoot branching in *Arabidopsis* and pea. *Genes & Development, 17*(12), 1469-1474. 10.1101/gad.256603
Soundappan, I., Bennett, T., Morffy, N., Liang, Y., Stang, J. P., Abbas, A., Leyser, O., & Nelson, D. C. (2015). SMAX1-LIKE/D53 Family Members Enable Distinct MAX2-Dependent Responses to Strigolactones and Karrikins in Arabidopsis. *Plant Cell, 27*(11), 3143-3159. 10.1105/tpc.15.00562
Stafstrom, J. P., & Sussex, I. M. (1988). Patterns of protein synthesis in dormant and growing vegatative buds of pea. *Planta, 176*(4), 497-505. 10.1007/bf00397656
Stepanova, A. N., & Alonso, J. M. (2005). Arabidopsis ethylene signaling pathway. *Science's STKE : signal transduction knowledge environment, 2005*(276), cm4-cm4. 10.1126/stke.2762005cm4
Stiefkens, L. B., Bernardello, G., & Anderson, G. J. (2001). The somatic chromosomes of *Sophora fernandeziana* (Fabaceae), an endemic tree from Robinson Crusoe Island. *Pacific Science, 55*(1), 71-75. 10.1353/psc.2001.0007
Stiefkens, L. B., Bernardello, G., & Anderson, G. J. (2003). The karyotype of *Sophora tetraptera* (Fabaceae). *New Zealand Journal of Botany, 41*(4), 731-735.

Stirnberg, P., Furner, I. J., & Ottoline Leyser, H. M. (2007). MAX2 participates in an SCF complex which acts locally at the node to suppress shoot branching. *Plant Journal, 50*(1), 80-94. 10.1111/j.1365-313X.2007.03032.x

Sussex, I. M., & Kerk, N. M. (2001). The evolution of plant architecture. *Current Opinion in Plant Biology, 4*(1), 33-37. 10.1016/s1369-5266(00)00132-1

Symons, G. M., Murfet, I. C., Ross, J. J., Sherriff, L. J., & Warkentin, T. D. (1999). bushy, a dominant pea mutant characterised by short, thin stems, tiny leaves and a major reduction in apical dominance. *Physiologia Plantarum, 107*(3), 346-352. 10.1034/j.1399-3054.1999.100312.x

Symons, G. M., Ross, J. J., & Murfet, I. C. (2002). The bushy pea mutant is IAA-deficient. *Physiologia Plantarum, 116*(3), 389-397. 10.1034/j.1399-3054.2002.1160315.x

Takeda, F., Krewer, G., Li, C. Y., MacLean, D., & Olmstead, J. W. (2013). Techniques for increasing machine harvest efficiency in highbush blueberry. *Horttechnology, 23*(4), 430-436.

Tanksley, S. D. (1993). Mapping polygenes. *Annual Review of Genetics, 27*, 205-233. 10.1146/annurev.ge.27.120193.001225

Taylor, G. M. (1975). Divaricating shrubs. *Nature Heritage, 6*, 2118-2127.

Thimann, K. V., & Skoog, F. (1933). Studies on the growth hormone of plants III The inhibiting action of the growth substance on bud development. *Proceedings of the National Academy of Sciences of the United States of America, 19*, 714-716. 10.1073/pnas.19.7.714

Thomas, H. H. (1916). On *Williamsoniella*, a new type of bennettitalean flower. *Philosophical Transactions of the Royal Society of London Series B-Containing Papers of a Biological Character, 207*, 113-U134. 10.1098/rstb.1916.0003

Timme, R. E., Bachvaroff, T. R., & Delwiche, C. F. (2012). Broad phylogenomic sampling and the sister lineage of land plants. *Plos One, 7*(1) 10.1371/journal.pone.0029696

Tomlinson, P. S. (1978). Some qualitative and quantitive aspects of New Zealand divaricating shrubs. *New Zealand Journal of Botany, 16*, 299-309.

Tranquillini, W. (1979). General features of the upper timberline. In *Physiological Ecology of the Alpine Timberline. Ecological Studies (Analysis and Synthesis)* (Vol. 31). Berlin, Heidelberg: Springer.

True, J. R., Mercer, J. M., & Laurie, C. C. (1996). Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics, 142*(2), 507-523.

Tucker, J. M. (1974). Patterns of parallel evolution of leaf form in new world oaks. *Taxon, 23*(1), 129-154. 10.2307/1218095

Van den Ende, W. (2014). Sugars take a central position in plant growth, development and, stress responses. A focus on apical dominance. *Frontiers in Plant Science, 5* 10.3389/fpls.2014.00313

Van Ooijen, J. (2004). MapQTL 5. Software for the mapping of quantitative trait loci in experimental populations. *Plant & Animal Genome XII Conference*

Van Well, R. (1974). Com-pact redhaven. *Fruit Varieties Journal, 38*, 37.

Vivancos, J., Spinner, L., Mazubert, C., Charlot, F., Paquet, N., Thareau, V., Dron, M., Nogue, F., & Charon, C. (2012). The function of the RNA-binding protein TEL1 in moss reveals ancient regulatory mechanisms of shoot development. *Plant Molecular Biology, 78*(4-5), 323-336. 10.1007/s11103-011-9867-9

Waldie, T., McCulloch, H., & Leyser, O. (2014). Strigolactones and the control of plant development: lessons from shoot branching. *Plant Journal, 79*(4), 607-622. 10.1111/tpj.12488

Wang, J.-W., Park, M. Y., Wang, L.-J., Koo, Y., Chen, X.-Y., Weigel, D., & Poethig, R. S. (2011). MiRNA Control of Vegetative Phase Change in Trees. *Plos Genetics, 7*(2) 10.1371/journal.pgen.1002012

Wang, S. B., J.C. Zeng, Z. B. . (2010). Windows QTL Cartographer 2.5. Raleigh, NC: Department of Statistics, North Carolina State University. Retrieved from http://statgen.ncsu.edu/qtlcart/WQTLCart.htm

Wang, Y., Sun, S., Zhu, W., Jia, K., Yang, H., & Wang, X. (2013). Strigolactone/MAX2-Induced Degradation of Brassinosteroid Transcriptional Effector BES1 Regulates Shoot Branching. *Developmental Cell, 27*(6), 681-688. 10.1016/j.devcel.2013.11.010

Wardle, P. (1985). Environmental influences on the vegetation of New Zealand. *New Zealand Journal of Botany, 23*(4), 773-788. 10.1080/0028825x.1985.10434242

Wardle, P. (1991). *Vegetation of New Zealand*. Cambridge: Cambridge University Press.

Wardle, P., & McGlone, M. S. (1988). Towards a more appropriate term for our divaricating shrubs and juvenile trees. *New Zealand Botanical Society newsletter, 11*, 16-18.

Waters, M. T., Gutjahr, C., Bennett, T., & Nelson, D. C. (2017). Strigolactone Signaling and Evolution. In S. S. Merchant (Ed.), *Annual Review of Plant Biology, Vol 68* (Vol. 68, pp. 291-322). 10.1146/annurev-arplant-042916-040925

Went, F. W. (1971). Parallel evolution. *Taxon, 20*, 197-226.

Whipple, C. J., Kebrom, T. H., Weber, A. L., Yang, F., Hall, D., Meeley, R., Schmidt, R., Doebley, J., Brutnell, T. P., & Jackson, D. P. (2011). grassy tillers1 promotes apical dominance in maize and responds to shade signals in the grasses. *Proceedings of the National Academy of Sciences of the United States of America, 108*(33), E506-E512. 10.1073/pnas.1102819108

Whitaker, A. H. (1987). The roles of lizards in New Zealand plant reproductive strategies. *New Zealand Journal of Botany, 25*(2), 315-328.

Wickett, N. J., Mirarab, S., Nam, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M. S., Burleigh, J. G., Gitzendanner, M. A., Ruhfel, B. R., Wafula, E., Der, J. P., Graham, S. W., Mathews, S., Melkonian, M., Soltis, D. E., Soltis, P. S., Miles, N. W., Rothfels, C. J., Pokorny, L., Shaw, A. J., DeGironimo, L., Stevenson, D. W., Surek, B., Villarreal, J. C., Roure, B., Philippe, H., dePamphilis, C. W., Chen, T., Deyholos, M. K., Baucom, R. S., Kutchan, T. M., Augustin, M. M., Wang, J., Zhang, Y., Tian, Z., Yan, Z., Wu, X., Sun, X., Wong, G. K.-S., & Leebens-Mack, J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America, 111*(45), E4859-E4868. 10.1073/pnas.1323926111

Wickson, M., & Thimann, K. V. (1958). The antagonism of auxin and kinetin in apical dominance. *Physiologia Plantarum, 11*(1), 62-74. 10.1111/j.1399-3054.1958.tb08426.x

Wilson, H. D. (1991). Distribution maps of small-leaved shrubs in Canterbury and Westland. *Canterbury Botanical Society Journal, 25*, 3-81.

Winn, A. A. (1999). The functional significance and fitness consequences of heterophylly. *International Journal of Plant Sciences, 160*(6), S113-S121. 10.1086/314222

Woo, H. R., Chung, K. M., Park, J. H., Oh, S. A., Ahn, T., Hong, S. H., Jang, S. K., & Nam, H. G. (2001). ORE9, an F-box protein that regulates leaf senescence in Arabidopsis. *Plant Cell, 13*(8), 1779-1790. 10.1105/tpc.13.8.1779

Wotton, D. M., Drake, D. R., Powlesland, R. G., & Ladley, J. J. (2016). The role of lizards as seed dispersers in New Zealand. *Journal of the Royal Society of New Zealand, 46*(1), 40-65. 10.1080/03036758.2015.1108924

Wu, G., Park, M. Y., Conway, S. R., Wang, J.-W., Weigel, D., & Poethig, R. S. (2009). The Sequential Action of *miR156* and *miR172* Regulates Developmental Timing in Arabidopsis. *Cell, 138*(4), 750-759. 10.1016/j.cell.2009.06.031

Yamazaki, K., Okabe, M., & Takahashi, E. (1987). Inheritance of some characteristics and breeding of new hybrids in flowering peaches. *Bulletin of the Kanagawa Horticultural Experiment Station*(34), 46-53.

Yang, L., Conway, S. R., & Poethig, R. S. (2011). Vegetative phase change is mediated by a leaf-derived signal that represses the transcription of miR156. *Development, 138*(2), 245-249. 10.1242/dev.058578

Yao, R., Ming, Z., Yan, L., Li, S., Wang, F., Ma, S., Yu, C., Yang, M., Chen, L., Chen, L., Li, Y., Yan, C., Miao, D., Sun, Z., Yan, J., Sun, Y., Wang, L., Chu, J., Fan, S., He, W., Deng, H., Nan, F., Li, J., Rao, Z., Lou, Z., & Xie, D. (2016). DWARF14 is a non-canonical hormone receptor for strigolactone. *Nature, 536*(7617), 469-+. 10.1038/nature19073

Young, N. D. (1994). *Constructing a plant genetic linkage map with DNA markers*.

Zeng, Z. B. (1993). Theoretical basis for seperation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of the National Academy of Sciences of the United States of America, 90*(23), 10972-10976. 10.1073/pnas.90.23.10972

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics, 136*(4), 1457-1468.

Zheng, H. Y., von Mollard, G. F., Kovaleva, V., Stevens, T. H., & Raikhel, N. V. (1999). The plant vesicle-associated SNARE AtVTI1a likely mediates vesicle transport from the trans-Golgi network to the prevacuolar compartment. *Molecular Biology of the Cell, 10*(7), 2251-2264.

Zhu, L., Zhang, Y., Guo, W., Xu, X., & Wang, Q. (2014b). De novo assembly and characterization of *Sophora japonica* transcriptome using RNA-seq. *BioMed Research International, 2014*, 750961-Article ID 750961.

Zhu, Q.-H., & Helliwell, C. A. (2011). Regulation of flowering time and floral patterning by miR172. *Journal of Experimental Botany, 62*(2), 487-495. 10.1093/jxb/erq295

Zotz, G., Wilhelm, K., & Becker, A. (2011). Heteroblasty-A Review. *Botanical Review, 77*(2), 109-151. 10.1007/s12229-010-9062-8

1.11 Appendix

Appendix 1.1. Summary of the hypotheses for the evolution of the divaricate growth form

Hypothesis	Description	Evidence for	Evidence against
Browsing by moa (Greenwood and Atkinson, 1977)	New Zealand had no browsing mammals but had large flightless moa. The divaricate form may reduce browsing damage from these moa	• Thin, strong, highly branched and small leaved shrubs reduces offtake so birds unable to maintain energy requirements (Bond *et al.*, 2004) • Divaricates thought to grow most often in areas moa were thought to live (lowland, fertile soils) (Greenwood and Atkinson, 1977) • Madagascar divaricates share many traits with New Zealand divaricates (both countries had grazing ratites)(Bond and Silander, 2007) • Spines are typically employed to reduce browsing by mammals but moa have protective horny covering on their head which my make spines ineffective (Greenwood and Atkinson, 1977) • Spines often present in Patagonian divaricates unlike NZ divaricates but Patagonia has browsing mammals as well as a small browsing ratite (McQueen, 2000)	• Moa thought to feed by clamping, breaking and pulling methods but twigs found in moa gizzards show clear cut ends (Burrows, 1980a, Burrows, 1980b) • Browsing of divaricate shrubs by mammalian browsers was also reduced • Divaricates present on the Chatham Islands where moa (or other ratites) never reached (Kavanagh, 2015) however they had smaller branch angles and larger leaves • Madagascar divaricates lacked higher concentration of leaves in centre of plant which has been contributed to protection from climatic conditions (Bond and Silander, 2007)
Climate (McGlone and Webb, 1981)	Divaricate form provides protection against harsh environmental conditions such as wind, frost and cold. Arose when climate began cooling as New Zealand flora derived from sub-tropical and warm-temperate climates so had to adapt to cooler, drier and unstable climates.	• Divaricate form may offer climate protection but comes at a cost due to the high degree of self-shading. Open areas and fertile soils, where divaricates are often found, may be more exposed to harsher climates (McGlone and Webb, 1981) • Island divaricates have smaller branch angles and larger leaves but Islands typically have milder climates so less protection from climate needed (McGlone and Webb, 1981, Kavanagh, 2015) • Densely branched shrubs have significantly calmer wind zones than regular branched shrubs suitable for fruit bearing and growth of new leaves (Keey and Lind, 1997) • The climate hypothesis thought to explain the evolution of an extinct gymnosperm fossil, Williamsoniaceae, as it is thought to have grown under environmental conditions similar to those thought to have contributed to the evolution of divaricates today however large herbivore browsers were also present at this time (Pott and McLoughlin, 2014) • Patagonian divaricate traits do not correlate with NZ divaricates when measured using the indices as they do not form a shield like structure and typically have spines to deter herbivory (McQueen, 2000) • Best predictor or divaricate in NZ is mean minimum coldest temperature	• Frost tolerance among divaricates and related non-divaricates varied with some divaricates less tolerant of frost than non-divaricates (Kelly and Ogle, 1990, Darrow *et al.*, 2001) • A fossil of an early divaricate, the species could not be properly identified, was believed to be from the early Miocene which pre-dates the predicted climatic conditions thought to be involved in divaricate evolution (Campbell *et al.*, 2000) • No evidence of a heat trap operating in species studied so far with little difference between the inside and exterior of the plants (Kelly and Ogle, 1990) • Environmental conditions thought to result in evolution of divaricates are also found in other parts of the world where divaricate species absent (Dawson, 1963) • Divaricate plants in NZ found in a variety of habitats which would have slightly different climatic conditions (Greenwood and Atkinson, 1977) • Wide branching angles of divaricates are not explained by the climate hypothesis (Greenwood and Atkinson, 1977)

Hypothesis	Description	Evidence for	Evidence against
Increased light harvesting (Kelly, 1994)	Close-packing of small, narrow-width leaves possibly with teeth or lobes, internodes greater than leaf size, enable light to penetrate into the centre of the bush and many leaves in the inner of the bush increase light harvesting in divaricates especially under canopy.	• Divaricates have small leaves, typically longer internodes and increased number of leaves in the middle of the bush (Kelly, 1994)	• The divaricate habit comes at a cost of high self-shading (McGlone & Webb, 1981)
Photoinhibition (Howell et al., 2002)	Divaricate plants possess traits that act to reduce photodamage under high irradiance increasing the light harvesting ability of plants	• Photoinhibition is a light-induced decrease in photosynthetic efficiency due to high light levels causing photodamage and can be induced by exposure to cold temperatures. • Exposure to bright light following a cold night promotes photoinhibition and has been observed in non-divaricate tree seedlings (Ball, 1994) • Divaricates in NZ frequently found in exposed, frosty inland habitats (McGlone & Webb, 1981) where conditions for photoinhibtion can be expected • In two plant species studied the shielding of leaves with outer screen of branches reduced photoinhibition (Howell et al., 2002)	• Divaricate species differed in the susceptibility of photoinhibition and their ability to recover from photoinhibition (Howell et al., 2002) • Divaricates possess some traits advantageous to photoinhibition, smaller leaves and stem and upper canopy with fewer leaves, but many traits are not typical of plants adapted to high light environments and divaricate had no distinctive response to high light compared to non-divaricates (Christian et al., 2006) • Cost of leaf area support to divaricates is so high that benefits from photoinhibition are offset offering little advantage to the plant (Christian et al., 2006) • No clear difference was observed in photoinhibition under high irradiance and drought conditions between two species of divaricates and related non-divaricates (Schneiderheinze, 2006)
Lizard frugivory (Lord & Marshall, 2001)	Many divaricates bear fruits on the underside of branches within the interlacing canopy often inaccessible to birds but able to be reached by lizards. The divaricate form provides protection to lizards from predators and fruit colour correlates with lizard preference.	• The divaricate form and lizard frugivory has previously been linked (Whitaker, 1987) • Divaricates often bear fruits on underside of branches and these are inaccessible to most NZ birds as they are larger than 15cm, even extinct species • The smaller and more agile lizards can easily move through divaricates and they provide a relatively safer basking site (Lord & Marshall, 2001) • *Larrea divaricate* (Zygophyllaceae) has been observed to provide a habitat for a species of lizard in Argentina (Deviana et al., 1994) • A significant correlation between white and blue fruit colour (lizard fruit colour preference), small fruit size and shrub or divaricate growth was observed in NZ shrubs (Lord & Marshall, 2001)	• Lizard preference may have resulted in fruit colour evolution in divaricate plants (Lord & Marshall, 2001) but it is not known if this may be a result of an already divaricating form • Weta have also been observed to consume and disperse fruits (Burns, 2006; Duthie et al., 2006) however they mostly did so in a way to avoid most seeds and may not act as seed dispersers (Morgan-Richards et al., 2008) • May only apply to fleshy fruited divaricate species

2 Chapter Two – Phenotyping of the Lincoln *Sophora* F_2

2.1 Abstract

Divarication is a plant growth form described as consisting of small leaved shrubs or tree juveniles with wide branch angles, interlacing branches, long internodes and weak apical dominance, giving the plants an appearance of a shrub with tangled branches. In New Zealand divarication occurs in ~ 50 species in 18 different families (~ 20 genera); however, nothing is known about the genetic architecture of the divaricate growth form. One of the New Zealand genera with a divaricate growth form is *Sophora* (Fabaceae), with the obligate divaricate *S. prostrata,* and the heteroblastic *S. microphylla,* which has a divaricating juvenile form. An existing segregating F_2 population formed from reciprocal crosses between *S. prostrata* and *S. tetraptera,* which has an arborescent tree form, was developed by Dr E.J. Godley ~ 40 years ago and displays segregation for the divaricate form. Nine traits related to the divaricate growth form were measured in 88 F_2 plants with all traits showing continuous distributions. Leaf traits, branch width and node angle were positively correlated, and were all negatively correlated with the other branch traits. These correlations indicated plants with larger leaves had wider branches, straighter stems, less branching and shorter internodes consistent with the phenotype of the parent, *S. tetraptera* and vice versa with the *S. prostrata* parent. Branch angle was not correlated with any of the other traits suggesting it is under control from different loci. Maternal effects were observed in the F_2 population with phenotypes of the F_2 trending toward the phenotypes of their maternal parent. The F_2 population for *Sophora* shows segregation for the divaricate growth form and suggests many traits may be controlled by the same loci because these traits co-vary. Based on our phenotype analyses the F_2 populations will be ideal for the study of the genetic architecture of the

divaricate growth form in *Sophora* and will contribute to increasing knowledge about divarication specifically and plant architecture more generally.

2.2 Introduction

Plant architecture describes the three dimensional organisation of the plant body including branching patterns, size, shape, and position of leaves and reproductive organs (Reinhardt & Kuhlemeier, 2002). The variety of plant architectures, ranging from small herbaceous annuals to vines, bush and tree forms, have evolved in response to differing environmental pressures (Reich et al., 2003; Rowe & Speck, 2005), enabling species to survive in a variety of habitats that previously may not have been habitable. Plant architecture is an important component of taxonomy, where it is used for classification of species, and in agriculture, where it can affect traits such as crop stability (Quine, 1990; Plourde et al., 2009), yield (Huyghe, 1998; Peng et al., 1999; Dawo et al., 2007; Kaggwa-Asiimwe et al., 2013; Rosati et al., 2013) and harvest efficiency (Takeda et al., 2013). Studies of plant architecture have often focused in the agricultural field or on evolution of early plants (Niklas, 2000) with less focus on the origins of the variety of plant architectures currently observed. Among the many plant growth forms observed in nature is the divaricate growth form (divarication). Divarication is described as small-leaved shrubs or tree juveniles with wide branch angles, interlacing branches, long internodes and weak apical dominance (Kelly, 1994). Species with a divaricating form are observed at low frequency throughout the world and some suggest the divaricate growth form may also be common in Argentina (Kelly & Ogle, 1990) and California (Tucker, 1974). However, this growth form is at its greatest frequency in New Zealand where it is found in ~ 50 species across 18 families (Greenwood & Atkinson, 1977; Atkinson & Greenwood, 1989). The prevalence of divarication at a low frequency in diverse plant families suggests that it has evolved independently multiple times.

Multiple hypotheses have been proposed for why the divaricate form has evolved and is present in such a high frequency in the New Zealand flora. The two core hypotheses are (1) the moa hypothesis (Went, 1971; Carlquist, 1974; Livingstone, 1974;

Greenwood & Atkinson, 1977), where it is believed the divaricate growth form reduces the profitability of browsing for moa (New Zealand's only large browsing herbivore pre-colonization) and (2) the climate hypothesis (Diels, 1897; Cockayne, 1912; McGlone & Webb, 1981), which proposes the divaricate growth form provides a shield-like structure protecting the plant from harsh climates thought to be present in New Zealand in the Miocene and Pleistocene periods.

The apparent independent evolution and the difficulty in conclusively determining adaptive explanations for the evolution of divarication led to the development of two independent indices that quantify divarication (Atkinson, 1992 and Kelly, 1994). These indices aimed to determine the main characteristics of the divaricate growth form and enabled comparisons between different species of divaricates. These two indices focused on different characteristics, such as branching or leaf dimensions, but correlated well with each other for most New Zealand species (Kelly, 1994). However, this correlation did not continue with Patagonian divaricate species suggesting that New Zealand divaricating species may have characteristics that differ to Patagonian divaricate plants (McQueen, 2000). Through developing the index, Kelly (1994) proposed a formal definition of divarication as a plant with "many interlaced branches with wide angles (mean >60°), small leaves (<60 mm^2) widely spaced leaves (mean distance >2x leaf width) with larger leaves towards the interior of the plant (inner leaves >1.4x larger than outer leaves)" (p.509).

The genus *Sophora* (Fabaceae) is one of the groups in New Zealand with divaricating species. *Sophora* consists of ~ 45 species with eight native and two exotic species currently recognised in New Zealand including the divaricate species *Sophora prostrata,* a non-divaricate species *S. tetraptera,* and a heteroblastic species with a divaricating juvenile form *S. microphylla*. *Sophora* species also readily hybridize with

each other naturally (Heenan et al., 2001) including among the different growth forms, e.g., divaricate and arborescent. This allows for the creation of hybrids for specific experiments, such as the formation of segregating populations for quantitative trait analysis, making this genus ideal for investigating the genetic architecture behind the divaricate growth form. Indeed, F_2 populations formed from reciprocal crosses between *S. prostrata* and *S. tetraptera* were made by Dr E. J. Godley in 1985 to investigate the hypothesis that heteroblastic *S. microphylla* arose from hybridization between *S. prostrata* and *S. tetraptera* (Godley, 1985). This resource persists today and provides an invaluable resource for the study of the genetics of the divaricate growth form as it has been formed from crosses between a divaricate and non-divaricate and shows segregation for traits that contribute to the divaricating form. This F_2 population is located in Lincoln, Canterbury, New Zealand and includes a total of 92 F_2 plants.

Developing a segregating F_2 population for woody species such as *Sophora* is time consuming as plants have to be grown to reproductive maturity, which can take several years or decades, and with multiple generations required for a segregating population, developing one for species with long lifespans is not often an option for researchers. The existence of a segregating *Sophora* F_2 population makes for a unique opportunity to study the genetic architecture behind the divaricating growth form. This chapter reports on the phenotyping results of the *Sophora* F_2 populations for a suite of nine traits that combine to yield the divaricating growth form.

2.3 Materials and Methods

2.3.1 Study site

The study site populations were initiated approximately 45 years ago by Dr E. J. Godley to investigate the origins of the heteroblast *S. microphylla*. This population was made from reciprocal crosses between *S. tetraptera* and *S. prostrata* to produce F_1 plants (Figure 2.1). Four of the F_1 plants were selfed to create F_2 populations, A, B, C and D that were planted on the 20th May 1985 at the 'East Block' site outside Lincoln, New Zealand. The 'East Block' is comprised of the remaining parental plants and the F_2 plants (arrayed in four blocks, A-D, according to lineage (see Fig. 2.1)) situated alongside a water race. The F_1 plants are located at a different site at the Landcare Research campus in Lincoln. The exact individuals that were used as the parental plants are not known and some of the original plants are no longer present at the site. However, plants collected from the original source sites are still present to represent the parents and were sampled in place

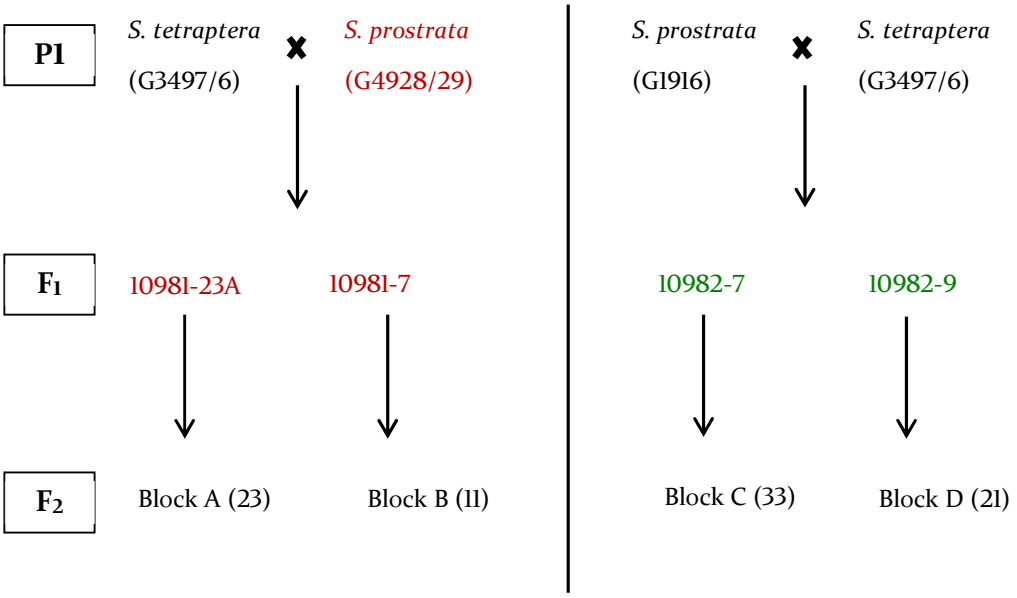


Figure 2.1. Schematic of the crosses used to produce the Lincoln F2 blocks; A, B C and D. Plants in red are no longer present so could not be used in sampling. Plants in green are still present and have been sampled. For the parental plants (in black) some plants are still present at the site from the original collection sites and these have been sampled in place of the parental plants. F1 were produced from two reciprocal crosses with different *S. prostrata* used in each cross and a different species as the maternal parents in each cross. Four different F1 were used to produce F2 in each of the four blocks A, B, C and D.

of the parental plants. *Sophora prostrata* from Rakaia Gorge (G4928/29) is the only parent that had no representation remaining at the source site. Two F_1 individuals, 10981-23A and 10981-7, produced the A and B F_2 blocks, respectively. These two F_1 plants are no longer present at the study site but two F_1 plants from the same cross, 10981-3 and 10981-18 were sampled instead. 10982-7 and 10982-9 produced the C and D F_2 blocks, respectively; these are both still present and were included in our sampling. Since the F_2 population was established, the site has not always been maintained, which resulted in the site becoming overgrown with a variety of other species. It was cleared out in 2014 but some of the original F_2 may have been lost. Following a survey of the site, the total remaining number of plants was 92 with 36 plants from blocks A and B (*S. tetraptera* maternal parent) and 56 plants from blocks C and D (*S. prostrata* maternal parent). The total number of plants identified per block was: A = 25, B = 12, C = 33 and D = 21.

The 92 F_2 plants identified at East Block were genotyped with six microsatellite markers (Table 2.1) previously developed by Van Etten *et al.* (2014) to ensure the sampled F_2 plants were actual F_2. Leaf tissue was collected and dried in silica gel and DNA extracted using the Qiagen DNEasy plant mini kit (Qiagen) as per manufacturer's instructions. DNA was diluted 1:10 in water before microsatellite PCR. PCR amplification was performed in a volume of 10 µL with 1x Standard *Taq* Reaction Buffer (New England Biolabs, Ipswich, Massachusetts), 50 µM of each dNTP, 0.5 U of *Taq* DNA polymerase (New England Biolabs, Ipswich, Massachusetts), 20 nM of M13-tailed forward primer, 450 nM of reverse primer and 450 nM dye-labelled M13 primer. Amplification by PCR was attained by: 95°C for 3 minutes, then 35 cycles of 95°C for 30 seconds, 53° C for 40 seconds and 72°C for 1 minute followed by 72°C for 10 minutes. One of two fluorescent dyes, FAM or VIC, was incorporated into each marker and two markers each containing a different dye were pooled for genotyping. The samples were genotyped on an ABI3730 DNA analyzer at the Massey Genome Service (Palmerston North, New Zealand) with

alleles scored manually in GENEMAPPER (v4.0) using CASS size standard (Symonds & Lloyd, 2004).

Table 2.1. Microsatellite primers used to genotype F_2. Information in the table is from Van Etten et al. (2014).

Locus		Primer sequence (5'-3')	Repeat motif	Size range (bp)	*Ta* (°C)
Sop-42	F	CCATACCTGACACTTGCGG	$(AG)_9$	179 - 187	53
Sop-42	R	TTGAGTCCAACATGAATGGC			53
Sop-248	F	TCCCGGAAATCTCATTCAAAGG	$(GTT)_{13}$	270-290	53
Sop - 248	R	ACTCAAGGAGTTTAGGTAGCG			53
Sop-445	F	CCAAATGGAGGAAGAAGGGTATTC	$(AGG)_7$	178-187	53
Sop-445	R	AGCTTCAACGCCAAACATCC			53
Sop-802	F	ACAAAGCCTCATACAGAGC	$(GTT)_{10}$	303-309	53
Sop-802	R	GAATGACCAAGGTATCGCC*			53
Sop-807	F	AGTGTACCTTGACGATTGTG	$(AT)_9$	319-325	53
Sop-807	R	TCAGTTGGTGAACTATCAAC*			53
Sop-834	F	TTGGGCCTACAATGTATGG	$(TCT)_9$	285-307	53
Sop-834	R	CATGCTCATCTCCCAAGAG			53

Note : Ta = annealing temperature used in PCR, * PIG tail (GTTTCTT) added to the 5' end of each reverse primer

2.3.2 Phenotyping trial

The parents, F_1 and a subset of F_2 individuals, ten from A block and nine from C block, were used to trial possible measurements for phenotyping of divaricate traits. The measurements trialled were chosen from divaricate definitions, divaricate indices and examination of plants at the site. From these criteria suitable traits, and the number of replicates needed for each measurement, were determined. These included those that: (1) were accurately measurable, either in the field or later in the lab, (2) were those typical of the divaricate growth form and (3) could differentiate the parental species. The nine traits measured were rachis length, number of leaflets, leaflet length and width, internode length, branch width, branches per 10 nodes (branch number), node angle and branch angle (Figure 2.2).

2.3.3 Phenotyping of Lincoln F_2

Phenotyping of the F_2 individuals, four F_1 individuals and four parent representatives was carried out at the same time as tissue collection for future genotyping. Four branches from each plant (chosen as representative of the plant) were measured for the nine traits. For each branch, a total of two leaves and two leaflets per leaf were used for the leaf traits. Leaves were chosen at random, but ensuring the leaf was mature, fully expanded and undamaged. Leaflet length was measured from the tip to

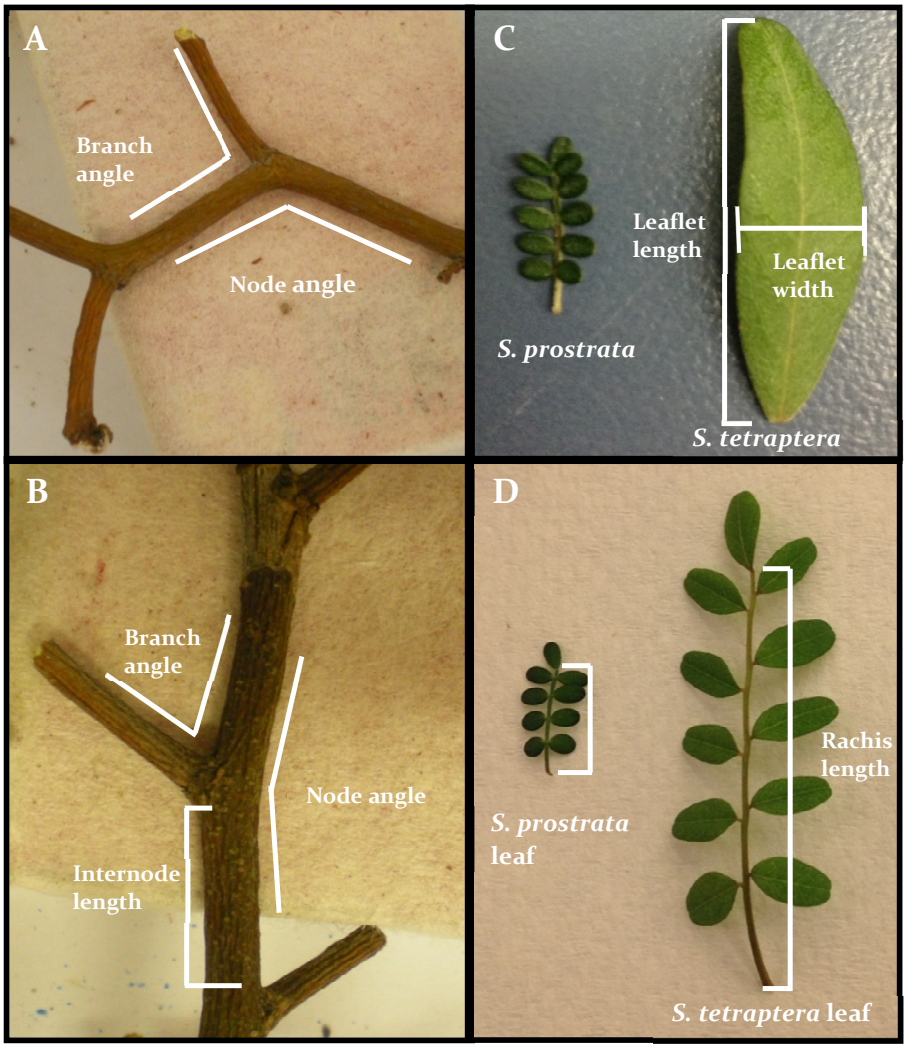


Figure 2.2. Examples of divaricate traits measured in the phenotyping of F_2 individuals. A, *S. prostrata* branch displaying node and branch angle measurements. B, *S. tetraptera* branch displaying node and branch angle measurements and internode length measurements (taken from the base of one node to the top of the subsequent node. C, A mature *S. prostrata* leaf compared to a mature leaflet from *S. tetraptera,* showing the length and width measurements for leaflets. D, Young *S. prostrata* leaf compared to a young *S. tetraptera* leaf displaying rachis length measurement.

the base of the leaflet and width measured from the widest part of the leaflet. The leaflets measured were taken from the middle of the leaf. For branch measurements, the top 10 cm of the branch was excluded and measurements were taken from the next node below this 10 cm. Internode length was measured for ten internodes and averaged and branch width was measured at the middle of the fifth internode. The number of branches per 10 nodes (BP10) count included the ten nodes used for the internode measurements. All measurements were made in mm units using digital callipers and were made on site for leaf measurements and in the lab for branch measurements. Branches were taken back to the lab, trimmed, and node and branch angles photographed using a digital camera (Nikon Coolpix L14). Ten nodes were photographed per branch. Node angle and branch angle measures were taken from these photographs using the angle tool in Fiji (ImageJ) v.1.5.1 (Schindelin et al., 2012).

2.3.4 Phenotyping analysis

Averages for each trait were calculated and analysis of phenotypic data was performed using R v. 3.4.2 (R Development Core Team, 2016). Descriptive statistics were calculated using the psych v. 1.8.3.3 (Revelle, 2016) package and the corrplot v. 0.84 package (Wei & Simko, 2016) was used to generate a heatmap of correlation coefficients. Scatterplots generated with the lattice v. 0.20-35 package (Sarkar, 2008) and histograms were generated with graphics v.3.4.2 package in R. The *t*-tests were performed using the *t.test* function of the basic stats v.3.4.2. The PCAs were calculated using the R stats v. 3.4.2 package with all data log transformed before running the calculation. The PCA was visualised with ggbiplot v. 0.55 (Vincent, 2011). One sample was removed before running analyses as it had no branches measured, resulting in missing data for branch angle, leaving 87 F_2 remaining. The Lincoln F_2 population was also divided into the maternal

cross directions and analysed using the described methods above to examine the potential maternal effect in the F_2.

2.4 Results

2.4.1 Parents and F_1

Compared with *S. tetraptera, S. prostrata* had smaller leaves, fewer leaflets per leaf, smaller node angles (resulting in the zig zag stem), smaller branch widths, wider branch angles, more branching per node and slightly longer internodes than *S. tetraptera* (Appendix 2.1). As expected, this combination of traits fits the description of a divaricate plant provided by Kelly (1990). The F_1 were intermediate to the parents (Figure 2.3) for node angle and leaf size (e.g. rachis length, leaflet number). The distributions of the branch traits, such as branch angle, number of branches and branch width (Appendix 2.2) were not intermediate between the parents, but this may be the result of the age difference in plants and F_1 that are frequently trimmed due to their location. All F_1 plants

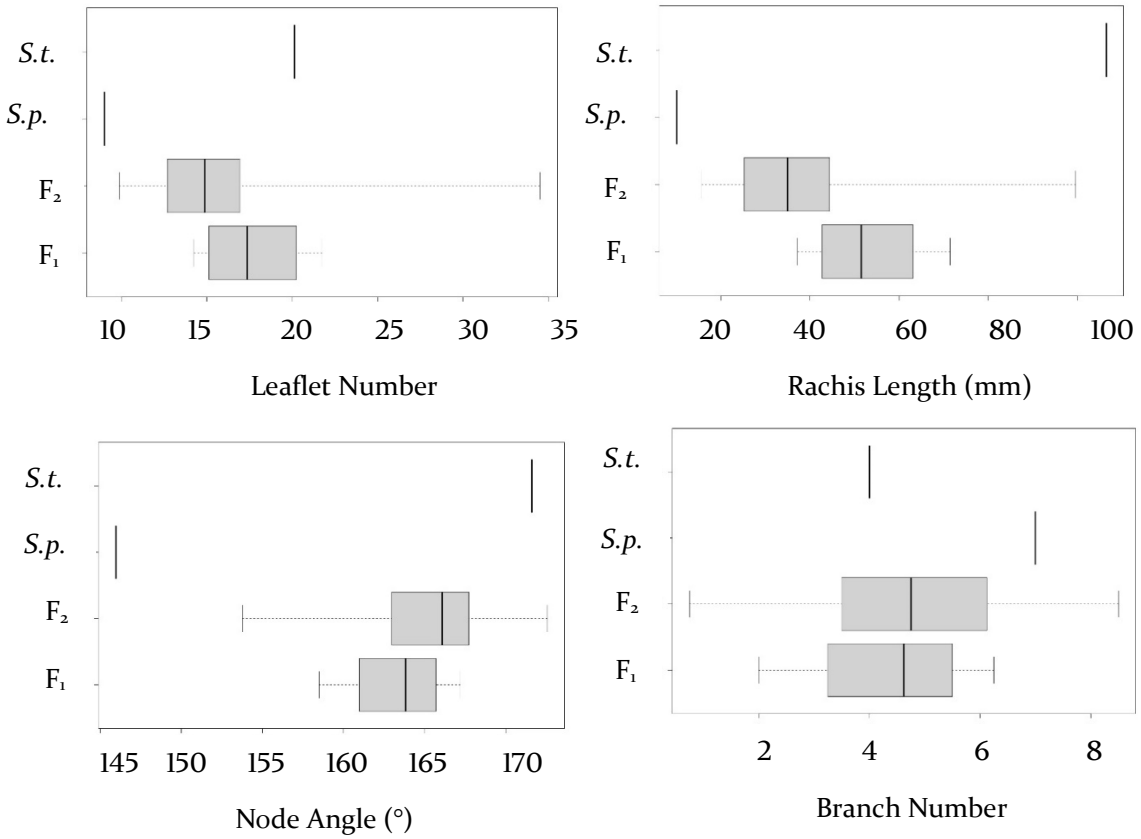


Figure 2.3. Example boxplots for distribution of traits measured in the F_1, F_2 and parents.

from a particular cross direction look similar to each other but F_1 from the reciprocal cross show phenotypic differences. When F_1 were separately grouped base on maternal parent species, the 10981 F_1 (*S. tetraptera* maternal parent) had larger leaves, more leaflets per leaf, larger node angles, shorter internodes, wider branches, and less branching than the 10982 F_1 (*S. prostrata* maternal parent) (Figure 2.4, Appendix 2.4 & 2.5), which is consistent with the phenotype of *S. tetraptera* and indicates that there are maternal effects influencing the F_1. The F_1 with *S. tetraptera* as maternal parent (10981 F_1) also have darker leaflets and fewer leaves than the 10982 F_1 (*S. prostrata* maternal parent) (personal observation, October, 12, 2015).

2.4.2 Divaricate Trait Variation

Microsatellite markers were used to confirm that plants sampled in the F_2 block were actually F_2 and not recruiters. Parents and F_1 were also genotyped to identify alleles that could possibly be inherited in the F_2 individuals. The analysis identified four plants that possessed alleles at three or more markers different from parental or F_1 alleles and were also distinctly different from other F_2 in that block. These plants were likely recruiters and not F_2 individuals and were therefore excluded from future analyses reducing the number of F_2 from 92 to 88. This reduced the total number of F_2 from blocks A and B (*S. tetraptera* maternal parent) to 34 plants and to 54 plants from blocks C and D (*S. prostrata* maternal parent). The total number of plants per block is: A = 23, B = 11, C = 33 and D = 21. Nine traits (Table 2.2) were measured from F_2 plants and each of the parental species, *S. tetraptera* and *S. prostrata*. The total number of F_2 plants was 88 with 34 plants from blocks A + B (*S. tetraptera* maternal parent) and 54 plants from blocks C + D (*S. prostrata* maternal parent). Descriptive statistics are summarised in Table 2.2. Frequency distributions show continuous variation for all traits (examples in Figure 2.5, Appendix 2.8) and suggest that most traits have a relatively normal

distribution. The F_2 ranges for internode length, leaflet number, branch width, branch number and branch angle distributions showed transgressive segregation exceeding parental phenotype values (Figure 2.5, Appendix 2.1, Table 2.2).

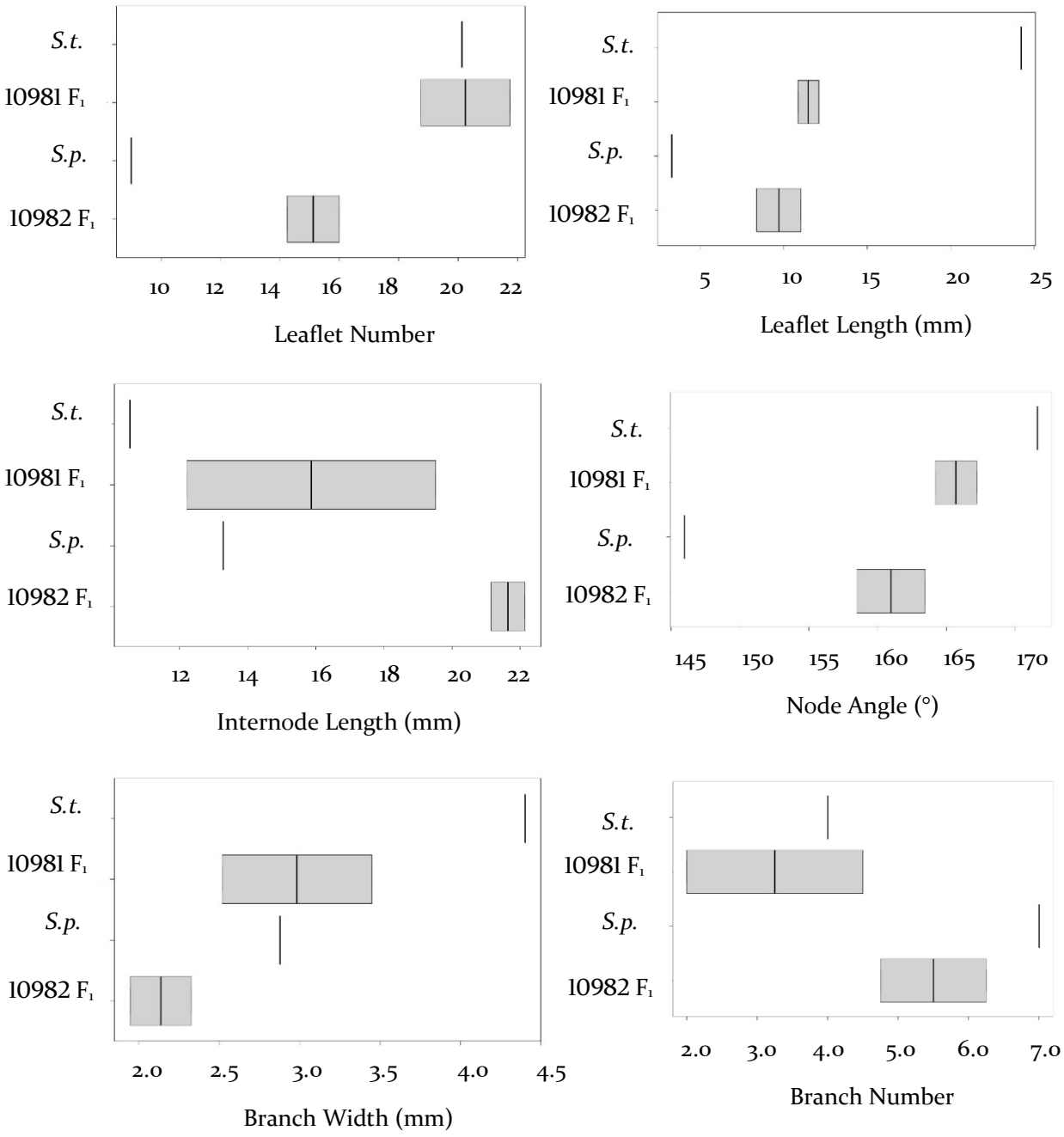


Figure 2.4. Examples of boxplot distributions for traits measured in the F_1 from each maternal cross. 10981 F_1 are F_1 with the maternal parent *S. tetraptera* (*S. t)* and 10982 F_1 are F_1 with the maternal parent *S. prostrata* (*S. p).*

Table 2.2. Descriptive statistics for the Lincoln F_2 *Sophora* population

Trait	n	mean	sd	median	min	max	range	skew	kurtosis	se
Leaflet number	88	15.27	3.74	14.88	9.88	34.50	24.62	1.81	6.68	0.40
Rachis Length (mm)	88	37.76	15.55	35.26	15.70	99.62	83.92	1.30	2.51	1.66
Leaflet Length (mm)	88	9.15	3.22	8.77	4.72	20.84	16.12	1.30	1.61	0.34
Leaflet Width (mm)	88	3.85	0.73	3.73	2.63	5.97	3.34	0.81	0.18	0.08
Internode Length (mm)	88	13.90	5.52	12.99	13.57	34.04	32.02	0.70	0.93	0.59
Branch Width (mm)	88	2.63	0.47	2.71	1.38	3.52	2.14	-0.49	-0.35	0.05
Branch number*	88	4.80	2.00	5.00	0.00	9.00	9.00	-0.12	-0.67	0.21
Node Angle (°)	88	165.30	4.03	166.19	153.77	174.27	20.50	-0.74	0.25	0.43
Branch Angle (°)	88	63.25	7.98	62.40	45.53	84.28	38.75	0.31	-0.30	0.86

n = number of samples, sd = standard deviation, min = minimum value, max = maximum value, skew = skewness, se = standard error, * number of branches per 10 nodes

Skewness and kurtosis give a more accurate description of the distribution of the data than frequency histograms. The limits suggested by Bulmer (1979) are often in use today to interpret skewness where skewness values close to zero (between ½ and –½) indicate an approximately symmetric distribution, values between 1 and ½ or -1 and – ½ indicate moderately skewed distributions and values less than -1 or greater than 1 indicate highly skewed distributions. These limits are also used for the Kurtosis measurement which measures the tails of the distribution in relation to the rest of the distribution. Skewness and kurtosis in the F_2 (Table 2.2) suggest most traits are moderately skewed but rachis length (Figure 2.5), leaflet length and leaflet number show high positive skew (longer tail to the right). Branch width, branch angle (Figure 2.5) and branch number show approximately symmetric distributions. However, sample size has been shown to have an effect on skewness (Cox, 2010) and branch age may influence many of these traits although, where possible, young branches were excluded from the measurements. Skewness was not associated with the observation of transgressive segregation in some traits.

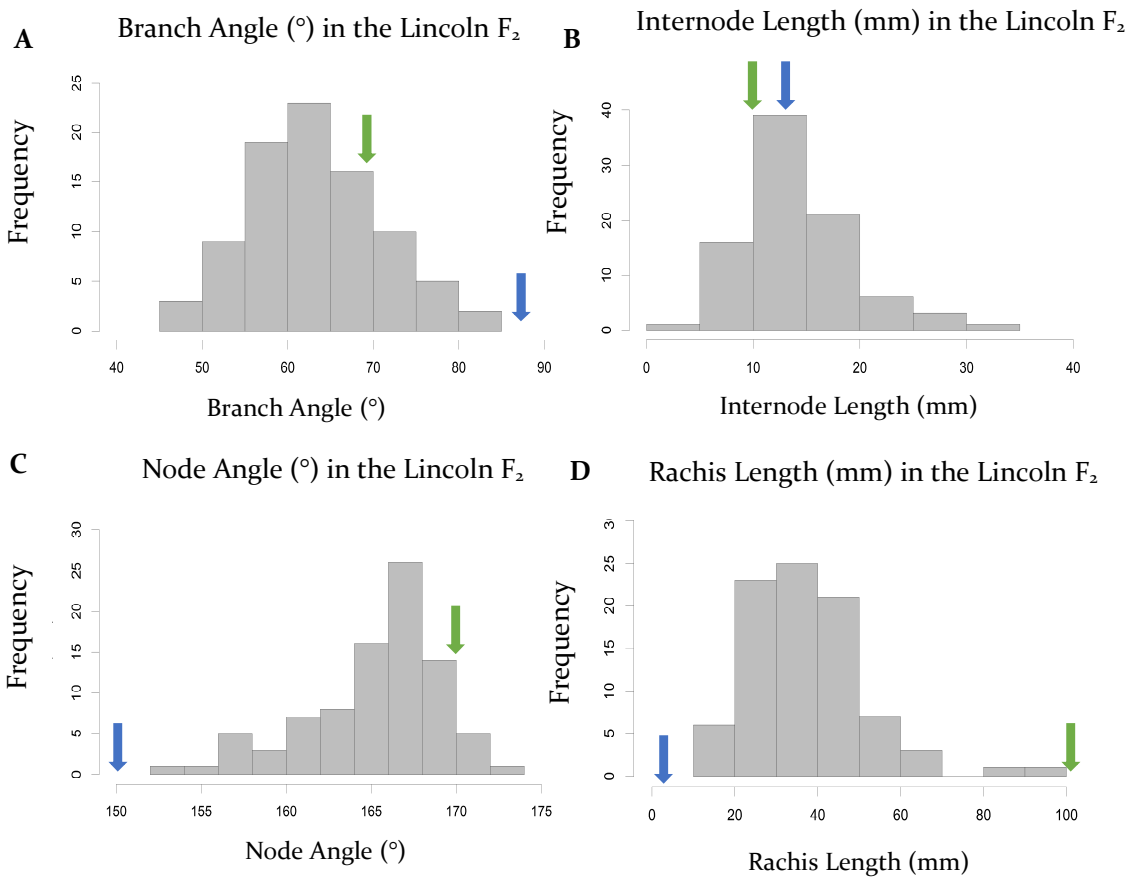


Figure 2.5. Example frequency distributions for the F_2 population. Green arrows represent *S. tetraptera* phenotype values and blue arrows represent *S. prostrata* phenotype values. A, Branch angle with a symmetric distribution, displaying transgressive segregation beyond *S. tetraptera*. B, Internode length with right skew and displaying transgressive segregation beyond both parental values. C, Node angle distribution displaying a negative skew and D, Rachis length with a positive skew.

2.4.3 Divaricate Trait Correlation

Correlations between trait distributions can indicate trait combinations that are inherited together and may have the same genetic underpinnings. Correlation coefficients (Figures 2.6 & 2.7) show that all leaf traits are positively correlated with each other suggesting that as the leaf size increases (rachis length), leaflet size and number tend to increase. Most branch traits showed some correlation with the other traits, except for branch angle which had no or very weak correlation (Figures 2.6 & 2.7). Among the branch traits, node angle has the highest positive correlation with leaf traits and branch width and is also negatively correlated with internode length, branch angle

and branches per 10 nodes (Figures 2.6 & 2.7). Node angle also is negatively correlated with internode length indicating that as node angle decreases internode length increases resulting in straighter branches with shorter internodes. This trend is also seen in branch width which is positively correlated with most traits and negatively correlated with internode length indicating branches with larger widths have larger leaves and straighter branches but shorter internodes. The number of branches is negatively correlated with most traits except internode length (Figures 2.6 & 2.7). These results suggest that plants with larger leaves generally have smaller internodes and less branching but straighter node angles and larger branch widths and vice versa which is consistent with the typical suite of characters that describe divaricates. Similarity between traits was also analysed by PCA (Figure 2.8) which identified PC1 as explaining 50% of the variance and shows the leaf traits, branch width and node angle as positively correlated and all of those are negatively correlated with internode length. Branch angle and branch number (BP10) were essentially distinct from all other traits.

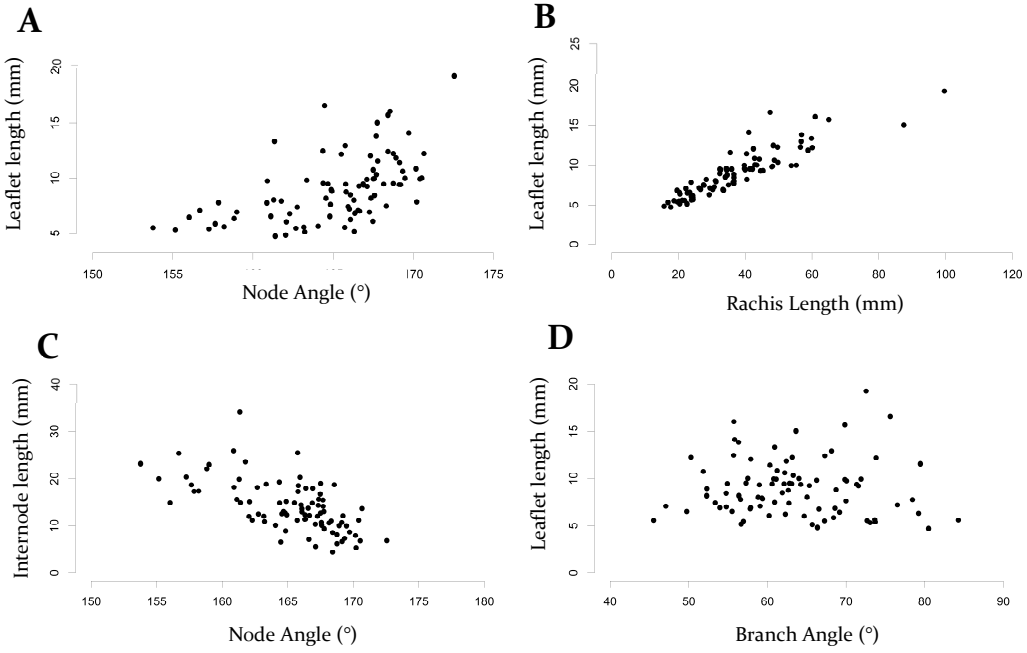


Figure 2.6. Example scatterplots for correlations between traits. A, Leaflet length and Node Angle displaying a positive correlation R=0.54. B, Leaflet length and rachis length with positive correlation R = 0.78. C, Internode length and node angle showing negative correlation R = -0.64 and D, Leaflet length and branch angle showing no correlation R = 0.02.

Figure 2.7. Heatmap showing correlation between traits. Leaf traits show high positive correlation with each other. Node angle shows the strongest correlations of the branch traits and is negatively correlated with branch angle, internode length and branches per 10 nodes and positively correlated with all other traits. The other branch traits tend to show negative correlations with the leaf traits except branch width which has positive correlation with the leaf traits. Branch angle shows very low or no correlation with other traits. Size and colour of the circle represent strength of correlation and R values for the circles are represented in the bottom half of figure.

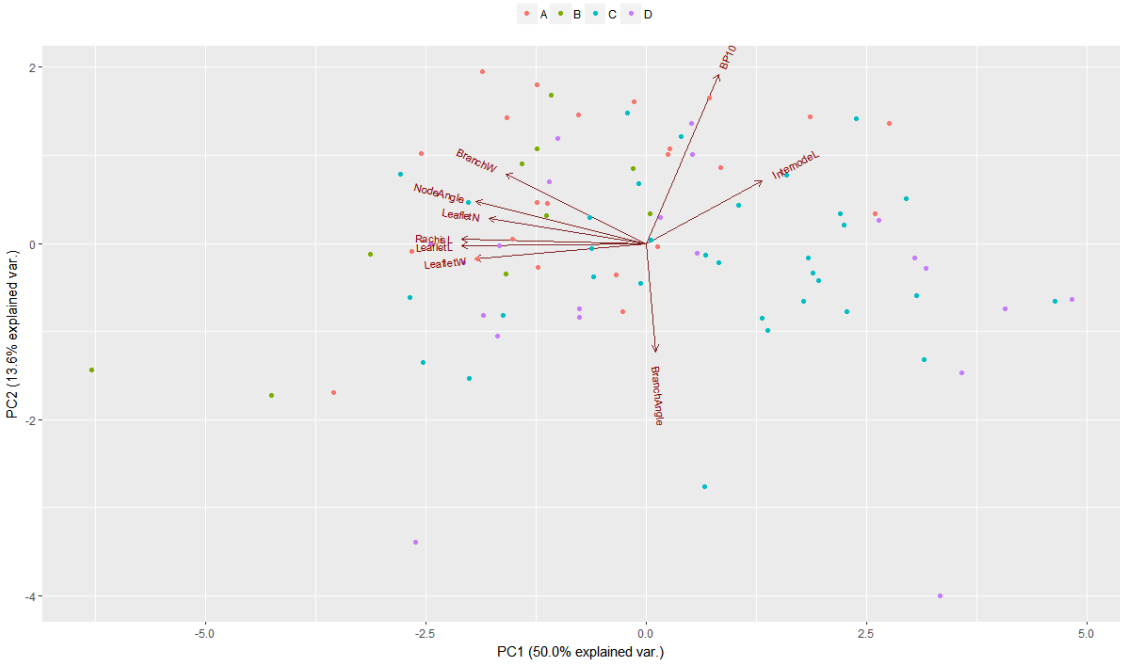


Figure 2.8. PCA plot based on the nine different traits measured in the Lincoln F_2. The different F_2 blocks are represented by the different colours in the plot as described by the legend.

2.4.4 Cross Direction Effects

The study site consisted of four blocks of F_2 plants generated from four separate F_1 plants (Figure 2.1). These four F_1 were generated from reciprocal crosses between *S. tetraptera* and *S. prostrata* with a different *S. prostrata* parent for each cross direction, *S. prostrata* from Rakaia Gorge as paternal parent in the first cross and *S. prostrata* from Waitahi as the maternal parent in the second cross (Figure 2.1). Of the four F_2 blocks; the A and B blocks descend from *S. tetraptera* as a maternal parent and the C and D blocks descend from *S. prostrata*, from Waitahi, as a maternal parent. To test for differences between crosses *t*-tests were carried out between each cross direction to determine if significant differences between the maternal parents were present (Table 2.3). The number of leaflets, rachis length, leaflet length, branch width, branch number and node angle were all significantly different between crosses with p values less than 0.05. Internode length, leaflet width and branch angle were not significantly different between cross direction (Table 2.3, Figure 2.9). Boxplots display distinct differences in distribution of data in

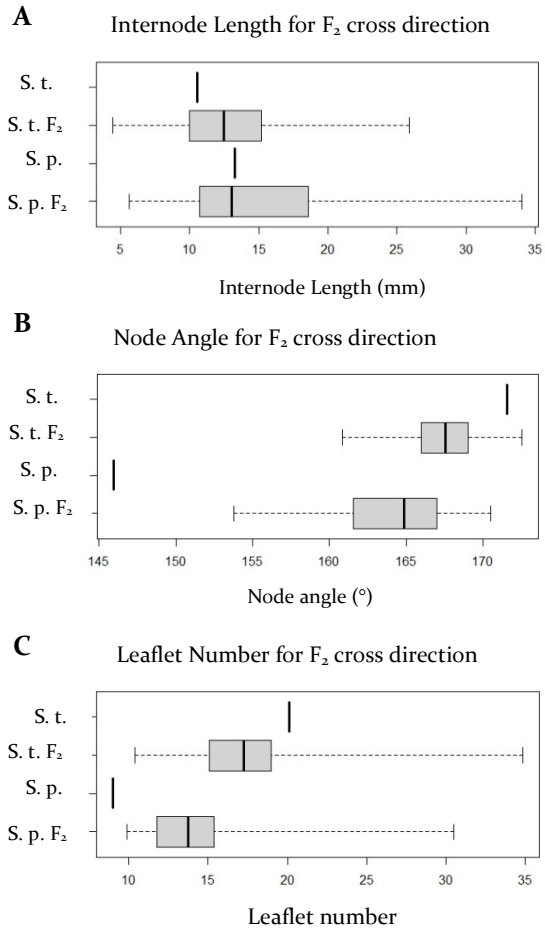


Figure 2.9. Example boxplots of traits with significant differences between cross direction. A, Internode length with transgressive segregation for both cross directions. B, Node angle and C, Leaflet number both displaying significant differences between cross direction. S.t represents the *S. tetraptera* parent and S.p the *S. prostrata* parent. S.t F_2 represents the F_2 with *S. tetraptera* as a maternal parent and S.p F_2 represents F_2 with *S. prostrata* as maternal parent.

significantly different traits, such as node angle and leaflet number, compared to non-significant traits such as internode length (Examples in Figure 2.9).

When F_2s from each cross direction are treated separately, internode length (Figure 2.9), branch width and number of branches per ten nodes continue to show transgressive segregation (Appendix 2.6). In the A and B blocks, leaflet number, rachis length and leaflet length all showed high positive skewness with other traits showing moderate skewness (Table 2.4). In the C and D blocks, leaflet number was the only trait to show high positive skewness (Table 2.5).

Table 2.3. *t*-Test results comparing the maternal cross directions of the Lincoln F_2

Trait	Mean		95% confidence interval		Sample size		df	tstat	P(T<=t) two tailed
	A + B	C + D			A + B	C + D			
Number leaflets*	17.64	13.69	-5.55	-2.35	33	54	44.28	-4.978	1.023e-05
Rachis length*	43.34	33.55	-16.6	-2.98	33	54	52.48	-2.8848	0.005671
Leaflet length*	9.92	8.46	-2.74	-0.17	33	54	67.35	-2.2626	0.02689
Leaflet width	4.00	3.73	-0.57	0.04	33	54	69.69	-1.7122	0.09132
Internode length	13.03	14.65	-0.59	3.84	33	54	80.46	1.454	0.1498
Branch width*	2.81	2.52	-0.48	-0.11	33	54	83.62	-3.227	0.001787
Branch number *	5.28	4.33	-1.81	-0.08	33	54	60.82	-2.1895	0.03241
Node angle*	167.1	164.0	-4.57	-1.70	33	54	84.98	-4.3448	3.841e-05
Branch angle	62.07	63.97	-1.43	5.23	33	54	78.496	1.1363	0.2593

* denotes significance < 0.05, 95% confidence intervals calculated by difference in means between the two groups + or – the error, negative confidence intervals occur when the difference in mean between the two groups is negative. Zero within the confidence intervals means the null hypothesis cannot be rejected.

The correlations between traits were slightly variable for each cross direction (Figures 2.10, 2.11 & 2.12). Leaf traits, branch width, and node angle had similar correlations between cross direction. Branch angle was not correlated in either cross direction except in the A & B blocks with leaflet number. Branch number and internode length had higher correlations with all other traits in the A & B blocks compared to the C & D blocks. The PCA analysis shows PC1 still explains ~ 50% of the variance and the

same traits group together when F_2 are grouped by maternal parent (Figure 2.12). The means of leaflet number, node angle and rachis length were higher for A and B blocks than for C and D blocks (Tables 2.3, 2.4 & 2.5). Internode length and branch angle means were higher in the C and D blocks than the A and B blocks (Tables 2.3, 2.4 & 2.5). These results show that the A and B group, on average, had more leaflets per leaf, longer rachis length, leaflet length and larger node angles (straighter branches) consistent with the phenotype of the original maternal parent species; *S. tetraptera,* which is larger and has straighter branches than *S. prostrata*. Alternatively, C and D F_2 plants typically had fewer leaflets, smaller rachis length, leaflet length and smaller node angles. However, there was an overlap in the distribution of traits (Examples in Figure 2.9) between these groups, clearly indicating that cross-direction alone does not determine trait distributions.

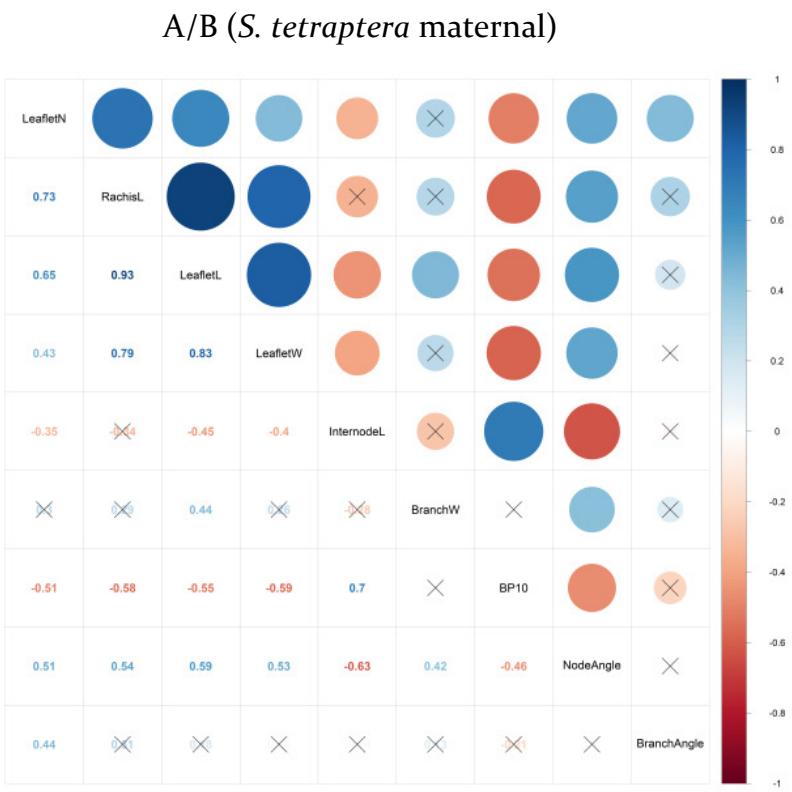


Figure 2.10. Correlation heatmap for the block AB cross direction. Non-significant results represented with the black X. Traits are indicated in the centre diagonal. Size and colour of the circles represent strength and direction of correlation.

C/D (*S. prostrata* maternal)

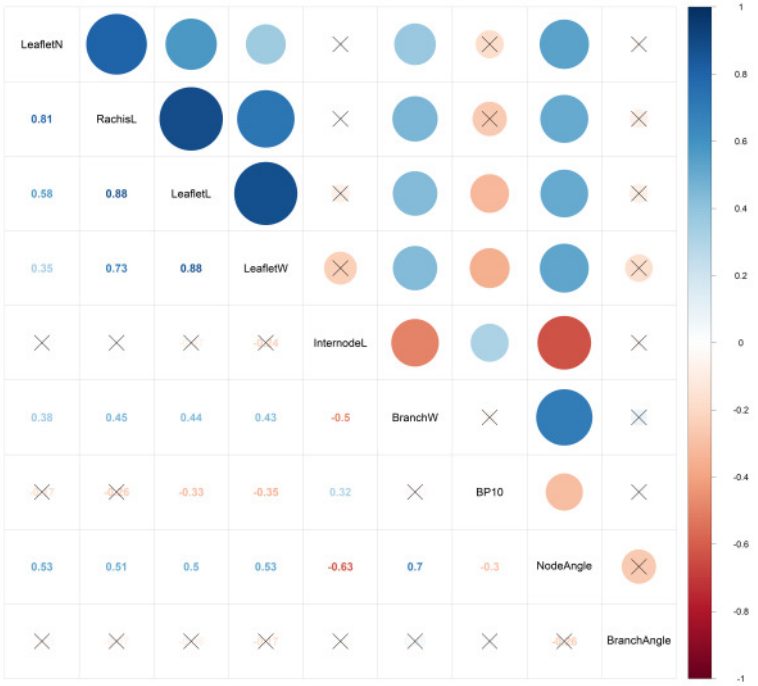


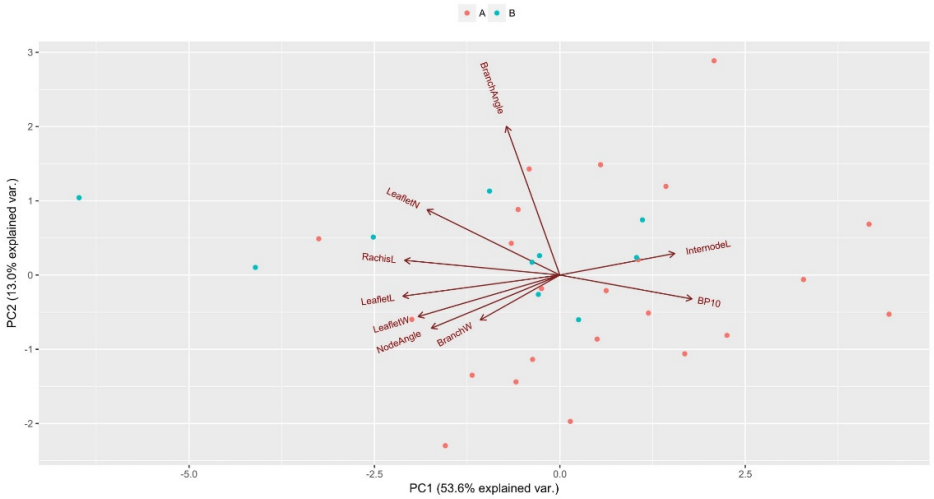
Figure 2.11. Correlation heatmaps for the block CD cross direction. Non-significant results represented with the black X. Traits are indicated in the centre diagonal. Size and colour of the circles represent strength and direction of correlation.

Table 2.4. Descriptive statistics for A and B blocks (with maternal parent *S. tetraptera*)

	n	mean	sd	median	min	max	range	skew	kutosis	se
Leaflet number	33	17.64	4.18	17.12	10.38	34.50	24.12	2.21	7.96	0.73
Rachis length (mm)	33	43.34	16.97	39.73	20.19	99.62	79.43	1.65	3.50	2.95
Leaflet length (mm)	33	9.92	2.92	9.48	5.46	19.27	13.81	1.28	2.27	0.51
Leaflet width (mm)	33	4.02	0.69	3.85	2.90	5.97	3.06	0.94	0.97	0.12
Internode length (mm)	33	13.03	4.49	12.71	4.45	25.90	21.45	0.48	0.99	0.78
Branch width (mm)	33	2.81	0.35	2.79	1.83	3.38	1.55	-0.47	0.57	0.06
Branch number	33	5.28	2.05	5.25	1.50	8.50	7.00	-0.20	-1.03	0.36
Node angle (°)	33	167.14	2.57	167.51	160.87	172.56	11.69	-0.61	0.79	0.45
Branch angle (°)	33	62.07	6.90	61.11	50.28	78.45	28.17	0.35	-0.46	1.20

n = number of samples, sd = standard deviation, min = minimum value, max = maximum value, skew = skewness, se = standard error

S. tetraptera maternal population (AB)



S. prostrata maternal population (CD)

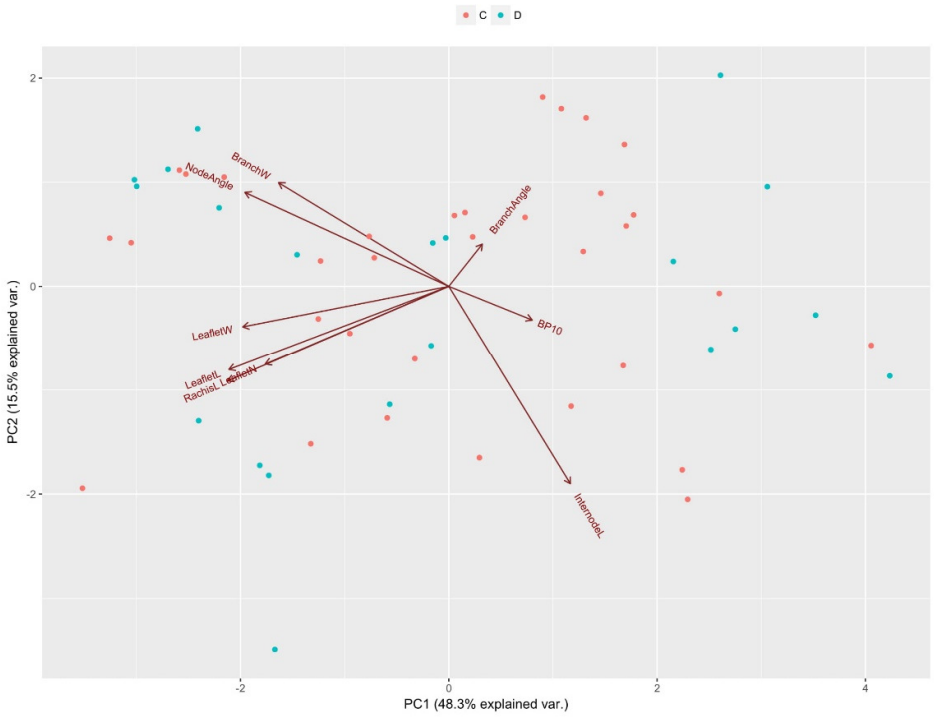


Figure 2.12. PCA plot based on nine traits for each maternal cross direction in the Lincoln F_2.

Table 2.5. Descriptive statistics of C and D blocks (with maternal parent *S. prostrata*)

	n	mean	sd	median	min	max	range	skew	kutosis	se
Leaflet number	54	13.69	2.33	13.69	9.88	19.50	9.62	0.44	-0.39	0.32
Rachis length (mm)	54	33.55	12.27	31.12	15.70	60.95	45.25	0.56	-0.61	1.67
Leaflet length (mm)	54	8.46	2.90	7.87	4.72	16.58	11.87	0.94	0.39	0.39
Leaflet width (mm)	54	3.73	0.72	3.63	2.63	5.79	3.16	0.94	0.56	0.10
Internode length (mm)	54	14.65	5.84	13.36	5.64	34.04	28.40	0.82	0.91	0.79
Branch width (mm)	54	2.52	0.51	2.62	1.38	3.52	2.14	-0.22	-0.63	0.07
Branch number	54	4.33	1.79	4.25	0.75	8.00	7.25	0.07	-0.52	0.24
Node angle (°)	54	164.0	4.17	164.84	153.77	170.53	16.76	-0.65	-0.35	0.57
Branch angle (°)	54	63.97	8.55	62.77	45.53	84.28	38.75	0.23	-0.22	1.16

n = number of samples, sd = standard deviation, min = minimum value, max = maximum value, skew = skewness, se = standard error

2.5 Discussion

Although working definitions for the divaricating growth form have not been entirely consistent or agreed upon, typically definitions consist of small-leaved shrubs or tree juveniles with wide branch angles, interlacing branches, long internodes and weak apical dominance (Kelly, 1994). A formal definition of divarication has been proposed by Kelly (1994) as a plant with interlaced branches, wide branch angles, small leaves, leaves widely spaced with larger leaves towards the interior of the plant (Kelly, 1994). Of the parental plants phenotyped here, *S. prostrata* is described as a divaricate and has smaller leaves, fewer leaflets per leaf, smaller node angles resulting in a zig-zag stem, smaller branch widths, wider branch angles, more branching per node and slightly longer internodes than *S. tetraptera* (Appendix 2.1) showing it fits the description of the divaricate growth form.

2.5.1 Divaricate traits in the *Sophora* F_2

The segregating populations, developed by Dr E. J. Godley, provide an opportunity to investigate the genetic architecture behind the divaricate form as F_2 plants show a range in phenotype from more divaricate-like to less divaricate-like. Here, we focus on nine traits commonly associated with divarication. The results show these traits are segregating in the F_2 populations and have relatively normal distributions (Figure 2.5, Table 2.2) in a common garden setting, suggesting a polygenic basis to trait variation. The PCA plot shows PC1 explains ~50% of the variance (Figure 2.8) and suggests three main groupings among traits further supporting the presence of independent genetic control for several traits but indicates that there may only be a few major loci involved. Likewise, correlations between traits (Figures 2.6 & 2.7) are indicative of a few major loci involved in divarication as most traits show correlations with other traits. Trait correlations occur largely due to pleiotropy, where one gene

influences multiple and often unrelated phenotypes, as seen in QTL analyses when correlated traits often map to the same loci (Xu et al., 2015; Chunlian et al., 2016; Wang et al., 2018), or by linkage disequilibrium (Chen & Luebberstedt, 2010). Therefore, correlations between traits can give an indication of traits that are often inherited together and which may be controlled by some of the same genes. This has been observed in rice where correlations between traits, for example spikelet density and total number of spikelets were highly correlated and QTLs were identified in the same regions for each of the traits indicating a role for pleiotropy (Lin et al., 1996). For divarication, leaf traits are expected to be correlated, as leaf phenotypes involve multiple converging genetic pathways and the final shape and size of leaves are tightly controlled (Gonzalez et al., 2010), but branch traits may be expected to be under different genetic control from leaf traits. Correlation coefficients (Figures 2.6 & 2.7) show all leaf traits are positively correlated with each other, as expected, suggesting that as the leaf size (rachis length) increases, leaflet size and number tend to also increase. Most branch traits showed some correlation with the other traits, except for branch angle which had no or very weak correlations (Figures 2.6 & 2.7). This suggests that branch angle is likely under different genetic control than the other traits. Branch angle is a key element of the divaricate growth form as the wider branch angles result in the densely interlacing growth form unique to divaricates. Branch angle is involved in the efficient gathering of resources, for example by positioning leaves to capture light for photosynthesis, an important factor for divaricates due to the high degree of self-shading (McGlone & Webb, 1981). Many genes have been linked to branch angle (reviewed by Roychoudhry and Kepinski (2015)), indicating this is a genetically complex trait. Based on pairwise correlations branch angle appears to be under different genetic control than the other traits associated with divarication, providing evidence that, at some level, multiple genes are involved in the evolution of the divaricating form in *Sophora*.

2.5.2 Transgressive segregation in the Lincoln F_2

Many of the traits examined here demonstrated transgressive segregation (Figure 2.5). Transgressive segregation occurs when some individuals in a segregating population (e.g. F_2), show phenotypes beyond the parental phenotypes. The traits showing this trend in the F_2 include internode length, leaflet number, branch width, number of branches and branch angle. Many hypotheses have been proposed to explain transgressive segregation such as overdominance, complementary gene action (Devicente & Tanksley, 1993), breakdown of linkage (Hagiwara et al., 2006), increased mutation rate, exposure of recessive alleles (Rick & Smith, 1953), and epistasis (Brem & Kruglyak, 2005). Transgressive segregation in plant hybrids is often suggested to be caused by complementary gene action usually due to the effects from additive alleles (Rieseberg et al., 2003), but epistasis and overdominance have been observed to play a role also (Rieseberg et al., 1999). Complementary gene action has been observed to be involved in transgressive segregation in rice tiller angle (Xu et al., 1998), barley seedling emergence and heading date (Castro et al., 2008), and in eight QTL studied in tomato, including plant height, stem diameter and number of branches, although other factors such as overdominance (Devicente & Tanksley, 1993) were involved in other traits. Epistasis also has been observed for transgressive segregation in nematode resistance in cotton (Wang et al., 2008). Complementary gene action results from recombination and independent assortment occurring in the F_1 to produce different combinations of alleles in the F_2 where some of those combinations result in phenotypes more extreme than the parents. This also indicates that these traits are under the control of multiple genes. Transgressive segregation is observed relatively frequently in segregating populations (Rieseberg et al., 1999), particularly in inbred populations and has been suggested as an evolutionary mechanism that enables rapid adaptation (Stebbins, 1959; Lewontin & Birch, 1966; Rieseberg et al., 1999) and can explain niche divergence (Stebbins, 1959;

Abbott, 1992; Rieseberg, 1997; Rieseberg et al., 1999) through hybridization in natural populations.

2.5.3 Maternal effects

Maternal effects are defined as the causal influence of the maternal phenotype or genotype on the offspring phenotype (Wolf & Wade, 2009). These maternal effects can be due to (1) nutrients, mRNAs or proteins supplied from the mother to the progeny (Byers et al., 1997; Wolf, 2000), (2) effects caused by maternal inheritance of plastids, (3) the endosperm, which is triploid (3*n*) with a 2*n* majority contribution from the maternal parent, (4) effects from the seed coat which is maternal tissue (Donohue, 2009) and (5) the environment the maternal parent is growing in. Maternal effects have been proposed as a mechanism of plasticity where the environment experienced by the maternal parent influences the phenotype in the offspring (Mousseau & Fox, 1998) and can increase the fitness of hybrids, leading to hybrid swarm stability. For example, in *Senecio* (Asteraceae) the offspring from the maternal parent *S. jacobaea* L. had higher fitness than offspring from *S. aquaticus* Hill as maternal parent (Kirk et al., 2005). These fitness differences could ultimately lead to speciation.

Our analysis of the reciprocal F_2 populations show that the A and B group (Figure 2.9, Tables 2.4 & 2.5), on average, had more leaflets per leaf, longer rachis length, leaflet length and larger node angles (straighter branches) which is consistent with the phenotypes of the original maternal parent species; *S. tetraptera (*Appendix 2.1). Alternatively, C and D F_2 plants typically had fewer leaflets, smaller rachis length, leaflet length and smaller node angles. However, there was an overlap in the distribution of traits (Examples in Figure 2.9) between these groups indicating that even though they trended towards the maternal parent's phenotype there is still considerable phenotypic variation in these traits as expected for polygenic traits.

Data from the reciprocal *Sophora* F_1 show that individuals from the same cross are more similar to one another than to those from the reciprocal cross (Figure 2.4), suggesting that there is a phenotypic effect from cross direction occurring, likely to be a maternal effect. In the F_2, the difference in cross direction is likely due to maternal effect where the maternal parent's contribution to the phenotype is more than expected from each parent individually (Roach & Wulff, 1987). As different *S. prostrata* plants were used as maternal and paternal parent in each cross (Figure 2.1), it is difficult to know if the cross effect is due to maternal effect or due to the use of different individual plants. However, the F_2 plants of a particular cross direction show a trend towards the original maternal parental species. For example, C and D block F_2 plants displayed, on average, fewer leaflets, smaller leaves and smaller node angles than the A and B block (Tables 2.3, 2.4, 2.5) showing a closer resemblance to the original maternal parent, *S. prostrata,* although the trait distributions for these F_2 groups often still show overlap (Examples in Figure 2.9). Correlations between traits show some differences in the F_2 blocks from different cross directions with the A & B blocks showing more significant correlations in internode length and number of branches (Figure 2.10). This could result from different maternal effects conveyed by *S. tetraptera,* and *S. prostrata*. Although the decreased number of samples in each cross direction, particularly the A & B blocks with 34 compared to 54 samples in the C & D blocks, also may have contributed to the difference in correlations between traits. There is also the possibility of block effect, as all F_2 from a certain cross have been grown in one block rather than being randomly distributed. This cannot be ruled out and may have resulted in differing environmental conditions acting on the plants from each block. As the plants have been growing at the site for many years and, at times, have been left without maintenance, potential block effects may have been accentuated.

2.6 Conclusion

The F_2 plants show variation in growth forms ranging from a divaricate-like growth form, similar to the *S. prostrata* parental phenotype, to a non-divaricate growth form, similar to the *S. tetraptera* parental phenotype. All the divaricate traits used in this study had continuous distributions with some showing transgressive segregation, suggesting multiple genes underlie each trait distribution. However, the correlation between traits suggests there may be a few loci of major effect for the divaricate growth form as the correlations indicate that these traits are inherited together. Comparing cross directions showed some traits with significantly different distributions between cross direction suggesting maternal effects may be present in *Sophora,* however there are other factors that could also be playing a role. The *Sophora* F_2 population shows segregation for divaricate-related traits and indicates this is an ideal resource for further study investigating the genetic architecture of the divaricate growth form and suggests there may be a few loci of major effect involved in the divaricate growth form, which may be identified by QTL analyses using these data.

2.7 References

Abbott, R. J. (1992). Plant invasions, interspecific hybridization and the evolution of new plant taxa. *Trends in Ecology & Evolution, 7*(12), 401-405. 10.1016/0169-5347(92)90020-c

Atkinson, I. A. E. (1992). *A method for measuring branch divergence and interlacing in woody plants*. Lower Hutt, New Zealand: DSIR Land Resources.

Atkinson, I. A. E., & Greenwood, R. M. (1989). Relationships between moas and plants. *New Zealand Journal of Ecology, 12*, 67-96.

Brem, R. B., & Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences of the United States of America, 102*(5), 1572-1577. 10.1073/pnas.0408709102

Bulmer, M. G. (1979). *Principles of Statistics* (3rd ed.). Cambridge: Cambridge University Press.

Byers, D. L., Platenkamp, G. A. J., & Shaw, R. G. (1997). Variation in seed characters in *Nemophila menziesii*: Evidence of a genetic basis for maternal effect. *Evolution, 51*(5), 1445-1456. 10.2307/2411197

Carlquist, S. (1974). *Island Biology*. New York: Columbia University.

Castro, A. J., Hayes, P., Viega, L., & Vales, I. (2008). Transgressive segregation for phenological traits in barley explained by two major QTL alleles with additivity. *Plant Breeding, 127*(6), 561-568. 10.1111/j.1439-0523.2008.01520.x

Chen, Y., & Luebberstedt, T. (2010). Molecular basis of trait correlations. *Trends in Plant Science, 15*(8), 454-461. 10.1016/j.tplants.2010.05.004

Chunlian, L., Guihua, B., Shiaoman, C., Carver, B., & Zhonghua, W. (2016). Single nucleotide polymorphisms linked to quantitative trait loci for grain quality traits in wheat. *Crop Journal, 4*(1), 1-11. 10.1016/j.cj.2015.10.002

Cockayne, L. (1912). Observations concerning evolution, derived from ecological studies in New Zealand. *Transactions and Proceedings of the New Zealand Institute, 44*, 1-30.

Cox, N. J. (2010). Speaking Stata: The limits of sample skewness and kurtosis. *Stata Journal, 10*(3), 482-495.

Dawo, M. I., Sanders, F. E., & Pilbeam, D. J. (2007). Yield, yield components and plant architecture in the F3 generation of common bean (*Phaseolus vulgaris* L.) derived from a cross between the determinate cultivar 'prelude' and an indeterminate landrace. *Euphytica, 156*(1-2), 77-87. 10.1007/s10681-007-9354-1

Devicente, M. C., & Tanksley, S. D. (1993). QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics, 134*(2), 585-596.

Diels, L. (1897). Vegetations biologie von Nei-Seeland. *Botanische Jahrbucher fur Systematik, Pflanzengeschichte und Pflanzengeographie, 22*, 202-300.

Donohue, K. (2009). Completing the cycle: maternal effects as the missing link in plant life histories. *Philosophical Transactions of the Royal Society B-Biological Sciences, 364*(1520), 1059-1074. 10.1098/rstb.2008.0291

Godley, E. J. (1985). Paths to maturity. *New Zealand Journal of Botany, 23*, 687-706.

Gonzalez, N., De Bodt, S., Sulpice, R., Jikumaru, Y., Chae, E., Dhondt, S., Van Daele, T., De Milde, L., Weigel, D., Kamiya, Y., Stitt, M., Beemster, G. T. S., & Inze, D. (2010). Increased Leaf Size: Different Means to an End. *Plant Physiology, 153*(3), 1261-1279. 10.1104/pp.110.156018

Greenwood, R. M., & Atkinson, I. A. E. (1977). Evolution of divaricating plants in New Zealand in relation to moa browsing. *Proceedings of the New Zealand Ecological Society, 24*

Hagiwara, W. E., Onishi, K., Takamure, I., & Sano, Y. (2006). Transgressive segregation due to linked QTLs for grain characteristics of rice. *Euphytica, 150*(1-2), 27-35. 10.1007/s10681-006-9085-8

Heenan, P. B., de Lange, P. J., & Wilton, A. D. (2001). Sophora(Fabaceae) in New Zealand: Taxonomy, distribution, and biogeography. *New Zealand Journal of Botany, 39*(1), 17-53. 10.1080/0028825x.2001.9512715

Huyghe, C. (1998). Genetics and genetic modifications of plant architecture in grain legumes: a review. *Agronomie, 18*(5-6), 383-411. 10.1051/agro:19980505

Kaggwa-Asiimwe, R., Andrade-Sanchez, P., & Wang, G. Y. (2013). Plant architecture influences growth and yield response of upland cotton to population density. *Field Crops Research, 145*, 52-59. 10.1016/j.fcr.2013.02.005

Kelly, D. (1994). Towards a numerical definition of divaricate (interlaced small-leaved) shrubs. *New Zealand Journal of Botany, 32*(4), 509-518.

Kelly, D., & Ogle, M. R. (1990). A test of the climate hypothesis for divaricate plants. *New Zealand Journal of Ecology, 13*

Kirk, H., Vrieling, K., & Klinkhamer, P. G. L. (2005). Maternal effects and heterosis influence the fitness of plant hybrids. *New Phytologist, 166*(2), 685-694. 10.1111/j.1469-8137.2005.01370.x

Lewontin, R. C., & Birch, L. C. (1966). Hybridization as a source of variation for adaptation to new environments. *Evolution, 20*(3), 315-&. 10.2307/2406633

Lin, H. X., Qian, H. R., Zhuang, J. Y., Lu, J., Min, S. K., Xiong, Z. M., Huang, N., & Zheng, K. L. (1996). RFLP mapping of QTLs for yield and related characters in rice (Oryza sativa L). *Theoretical and Applied Genetics, 92*(8), 920-927. 10.1007/bf00224031

Livingstone, D. A. (1974). Extract from INQUA report to the U.S. National Academy of Sciences. *Soil News, 22*, 46-48.

McGlone, M. S., & Webb, C. J. (1981). Selective forces influencing the evolution of divaricatin plants. *New Zealand Journal of Ecology, 4*, 20-28.

McQueen, D. R. (2000). Divaricating shrubs in Patagonia and New Zealand. *New Zealand Journal of Ecology, 24*(1), 69 - 80.

Mousseau, T. A., & Fox, C. W. (1998). The adaptive significance of maternal effects. *Trends in Ecology & Evolution, 13*(10), 403-407. 10.1016/s0169-5347(98)01472-4

Niklas, K. J. (2000). The evolution of plant body plans - A biomechanical perspective. *Annals of Botany, 85*(4), 411-438. 10.1006/anbo.1999.1100

Peng, J. R., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., Beales, J., Fish, L. J., Worland, A. J., Pelica, F., Sudhakar, D., Christou, P., Snape, J. W., Gale, M. D., & Harberd, N. P. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature, 400*(6741), 256-261.

Plourde, A., Krause, C., & Lord, D. (2009). Spatial distribution, architecture, and development of the root system of *Pinus banksiana* Lamb. in natural and planted stands. *Forest Ecology and Management, 258*(9), 2143-2152. 10.1016/j.foreco.2009.08.016

Quine, C. P. (1990). Planting at stump - preliminary evidence of its effect on root architecture and tree stability. *Research Information Note - Forestry Commission Research Division*(169), 3 pp.

R Development Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reich, P. B., Wright, I. J., Cavender-Bares, J., Craine, J. M., Oleksyn, J., Westoby, M., & Walters, M. B. (2003). The evolution of plant functional variation: Traits, spectra,

and strategies. *International Journal of Plant Sciences, 164*(3), S143-S164. 10.1086/374368
Reinhardt, D., & Kuhlemeier, C. (2002). Plant architecture. *Embo Reports, 3*(9), 846-851. 10.1093/embo-reports/kvf177
Revelle, W. (2016). psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, Illinois: Northwestern University.
Rick, C. M., & Smith, P. G. (1953). Novel variation in tomato species hybrids. *American Naturalist, 87*(837), 359-373. 10.1086/281796
Rieseberg, L. H. (1997). Hybrid origins of plant species. *Annual Review of Ecology and Systematics, 28*, 359-389. 10.1146/annurev.ecolsys.28.1.359
Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity, 83*, 363-372. 10.1038/sj.hdy.6886170
Rieseberg, L. H., Widmer, A., Arntz, A. M., & Burke, J. M. (2003). The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 358*(1434), 1141-1147. 10.1098/rstb.2003.1283
Roach, D. A., & Wulff, R. D. (1987). Maternal effects in plants. *Annual Review of Ecology and Systematics, 18*, 209-235. 10.1146/annurev.ecolsys.18.1.209
Rosati, A., Paoletti, A., Caporali, S., & Perri, E. (2013). The role of tree architecture in super high density olive orchards. *Scientia Horticulturae, 161*, 24-29. 10.1016/j.scienta.2013.06.044
Rowe, N., & Speck, T. (2005). Plant growth forms: an ecological and evolutionary perspective. *New Phytologist, 166*(1), 61-72. 10.1111/j.1469-8137.2004.01309.x
Roychoudhry, S., and Kepinski, S. (2015). Shoot and root branch growth angle control - the wonderfulness of lateralness. *Current Opinion in Plant Biology, 23*, 124-131. 10.1016/j.pbi.2014.12.004
Sarkar, D. (2008). Lattice: Multivariate Data Visualization with R. Springer, New York.
Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods, 9*(7), 676-682. 10.1038/nmeth.2019
Stebbins, G. L. (1959). The role of hybridization in evolution. *Proceedings of the American Philosophical Society, 103*, 231-251.
Symonds, V. V., & Lloyd, A. M. (2004). A simple and inexpensive method for producing fluorescently labelled size standard. *Molecular Ecology Notes, 4*(4), 768-771. 10.1111/j.1471-8286.2004.00763.x
Takeda, F., Krewer, G., Li, C. Y., MacLean, D., & Olmstead, J. W. (2013). Techniques for increasing machine harvest efficiency in highbush blueberry. *Horttechnology, 23*(4), 430-436.
Tucker, J. M. (1974). Patterns of parallel evolution of leaf form in new world oaks. *Taxon, 23*(1), 129-154. 10.2307/1218095
Van Etten, M. L., Houliston, G. J., Mitchell, C. M., Heenan, P. B., Robertson, A. W., & Tate, J. A. (2014). *Sophora microphylla* (Fabaceae) microsatellite markers and their utility across the genus. *Applications in Plant Sciences, 2*(3) 10.3732/apps.1300081
Vincent, Q. V. (2011). A ggplot2 based biplot (Version 0.55).
Wang, C. L., Ulloa, M., & Roberts, P. A. (2008). A transgressive segregation factor (*RKN2*) in *Gossypium barbadense* for nematode resistance clusters with gene *RKN1* in *G. hirsutum*. *Molecular Genetics and Genomics, 279*(1), 41-52. 10.1007/s00438-007-0292-3

Wang, X., Chen, Z., Li, Q., Zhang, J., Liu, S., & Duan, D. (2018). High-density SNP-based QTL mapping and candidate gene screening for yield-related blade length and width in *Saccharina japonica* (Laminariales, Phaeophyta). *Scientific Reports, 8* 10.1038/s41598-018-32015-y

Wei, T., & Simko, V. (2016). Visualization of a Correlation Matrix. R package version 0.77: https://CRAN.R-project.org/package=corrplot.

Went, F. W. (1971). Parallel evolution. *Taxon, 20*, 197-226.

Wolf, J. B. (2000). Gene interactions from maternal effects. *Evolution, 54*(6), 1882-1898.

Wolf, J. B., & Wade, M. J. (2009). What are maternal effects (and what are they not)? *Philosophical Transactions of the Royal Society B-Biological Sciences, 364*(1520), 1107-1115. 10.1098/rstb.2008.0238

Xu, Y., Huang, L., Ji, D., Chen, C., Zheng, H., & Xie, C. (2015). Construction of a dense genetic linkage map and mapping quantitative trait loci for economic traits of a doubled haploid population of *Pyropia haitanensis* (Bangiales, Rhodophyta). *Bmc Plant Biology, 15* 10.1186/s12870-015-0604-4

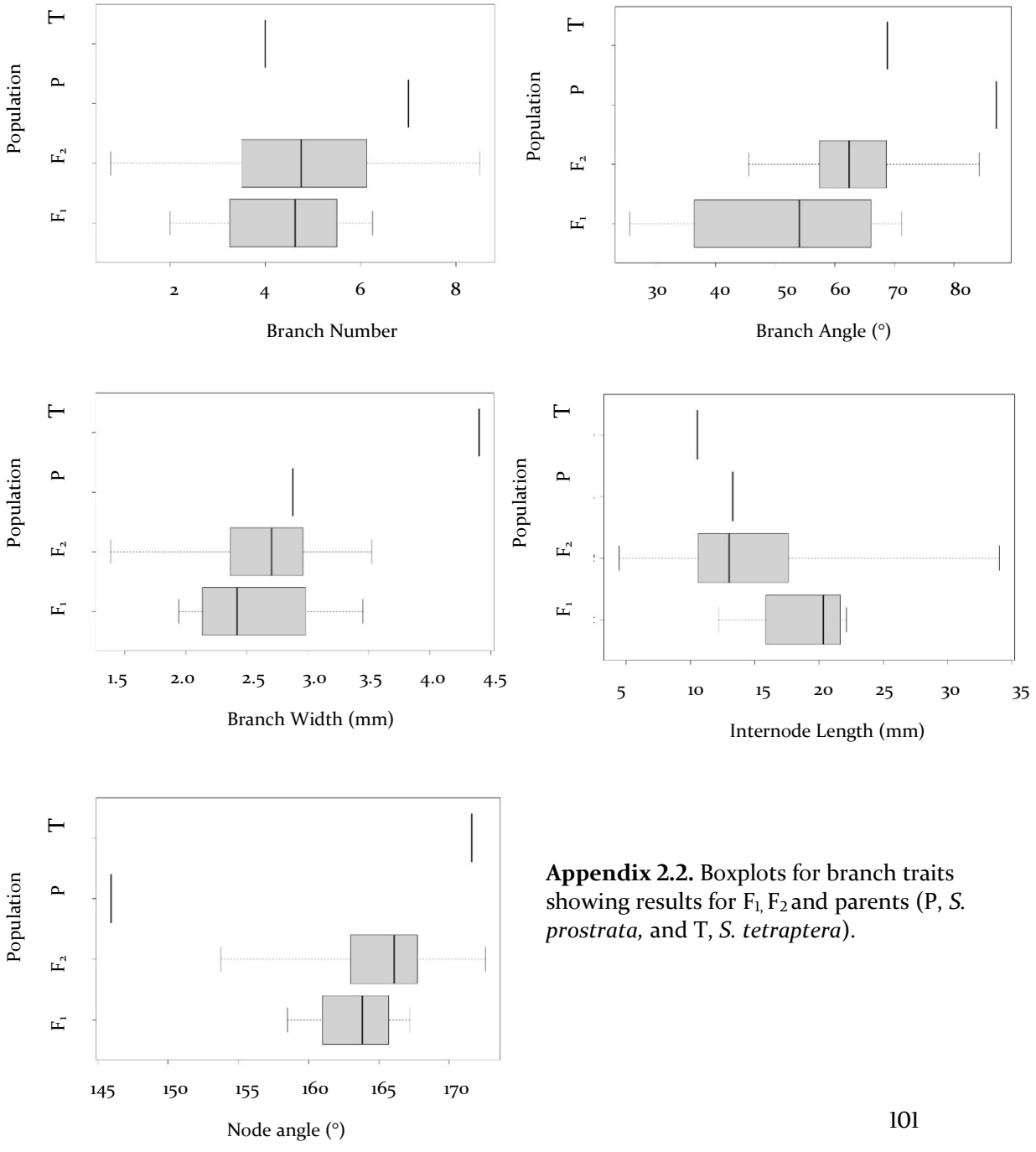
Xu, Y. B., McCouch, S. R., & Shen, Z. T. (1998). Transgressive segregation of tiller angle in rice caused by complementary gene action. *Crop Science, 38*(1), 12-19.

2.8 Appendix

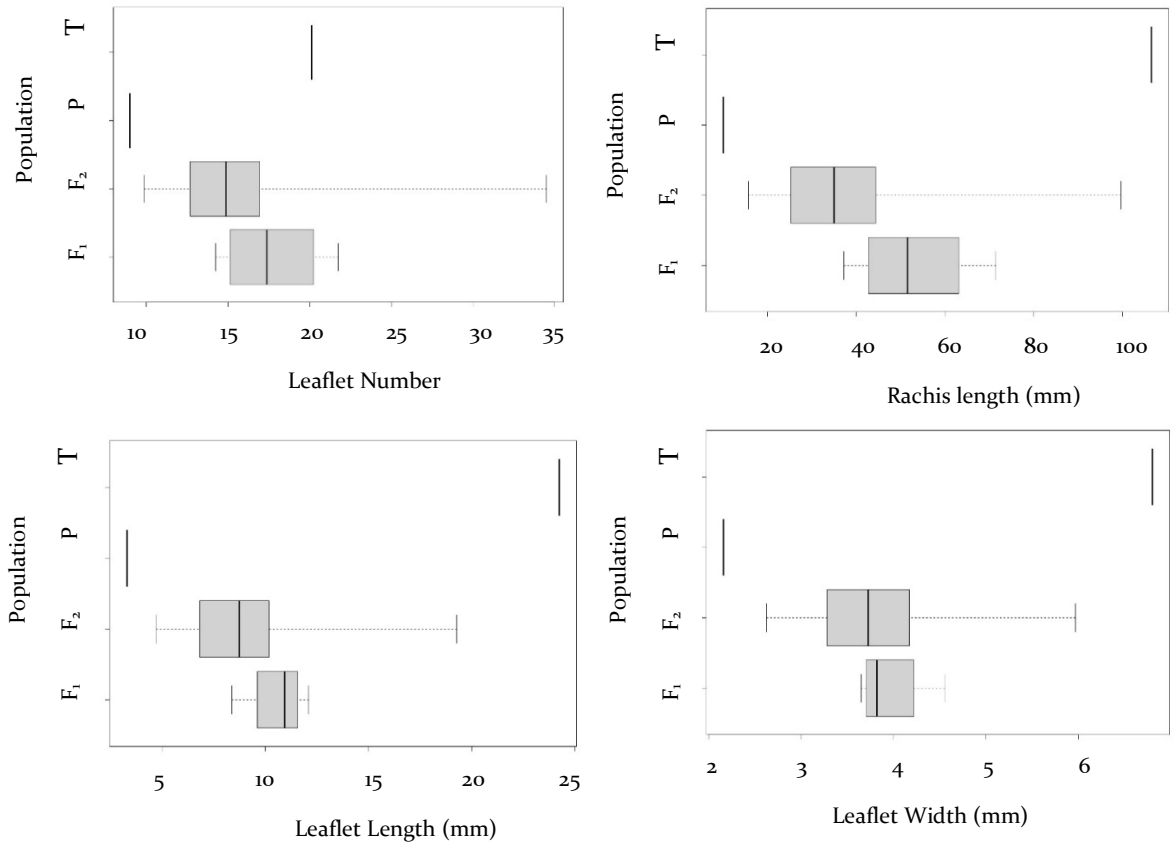
Appendix 2.1. Parental values for the nine divaricate traits from available parental plants growing on site in Lincoln.

	LN	RL	LL	LW	IL	BW	BN	NA	BA
S. tetraptera	20.1	106.5	24.23	6.79	10.28	4.4	4	171.6	68.8
S. prostrata	9	10.1	3.29	2.16	13.27	2.88	7	145.96	87.12

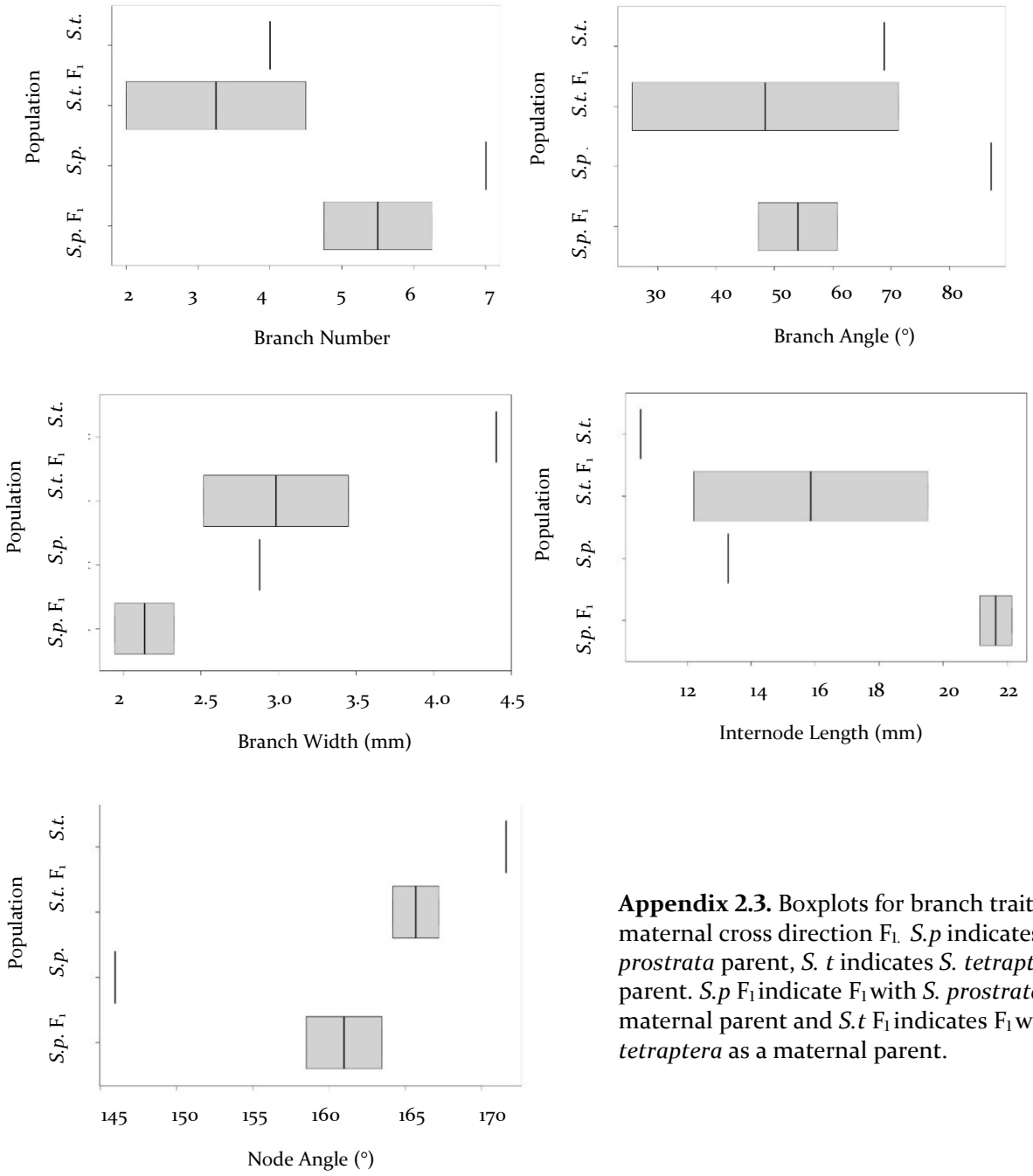
LN – leaflet number, RL – rachis length, LL – leaflet length, LW – leaflet width, IL – internode length, BW – branch width, BN – branch number, NA – node angle, BA – branch angle



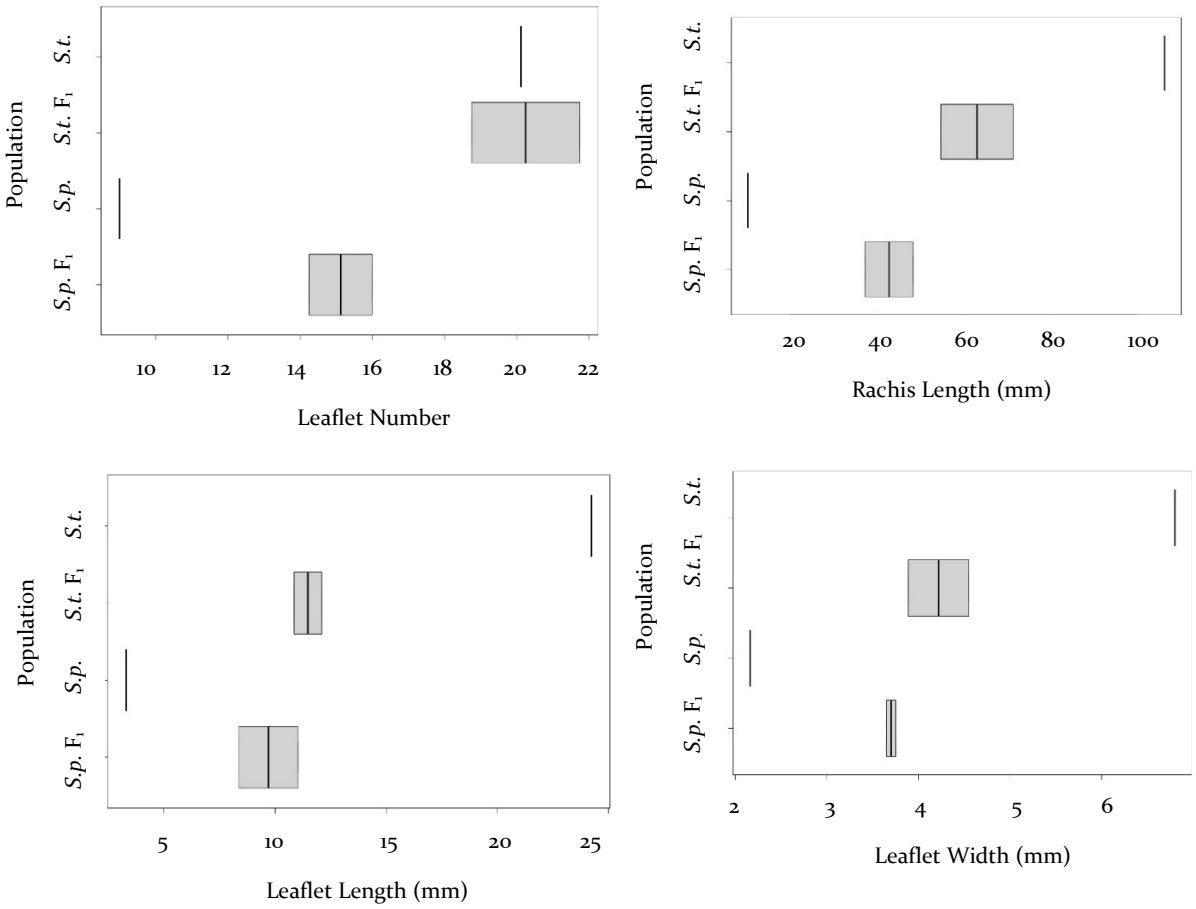
Appendix 2.2. Boxplots for branch traits showing results for F_1, F_2 and parents (P, *S. prostrata,* and T, *S. tetraptera*).



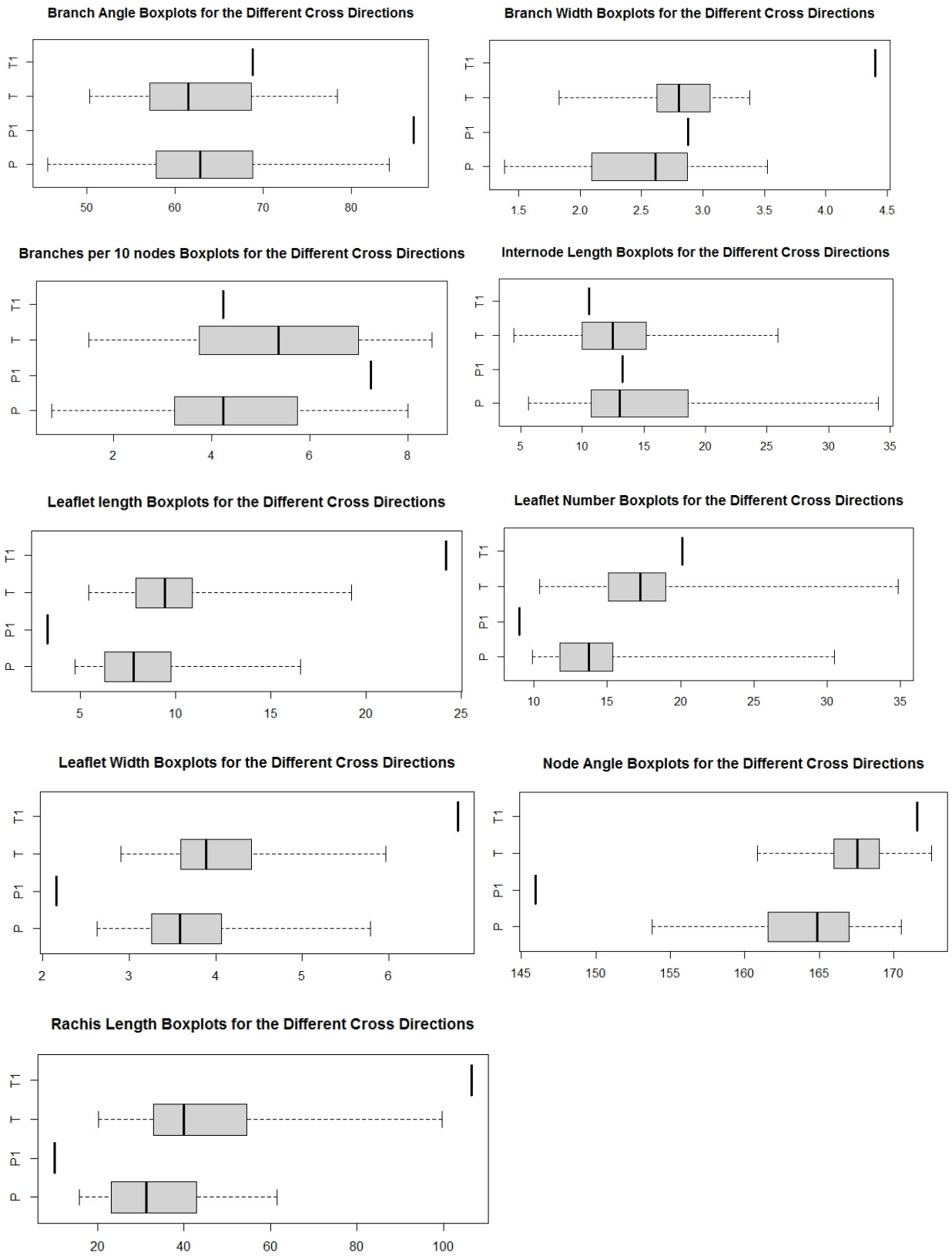
Appendix 2.2. Boxplots for leaf traits for F_1, F_2, and parents (P, *S. prostrata,* and T, *S. tetraptera*).



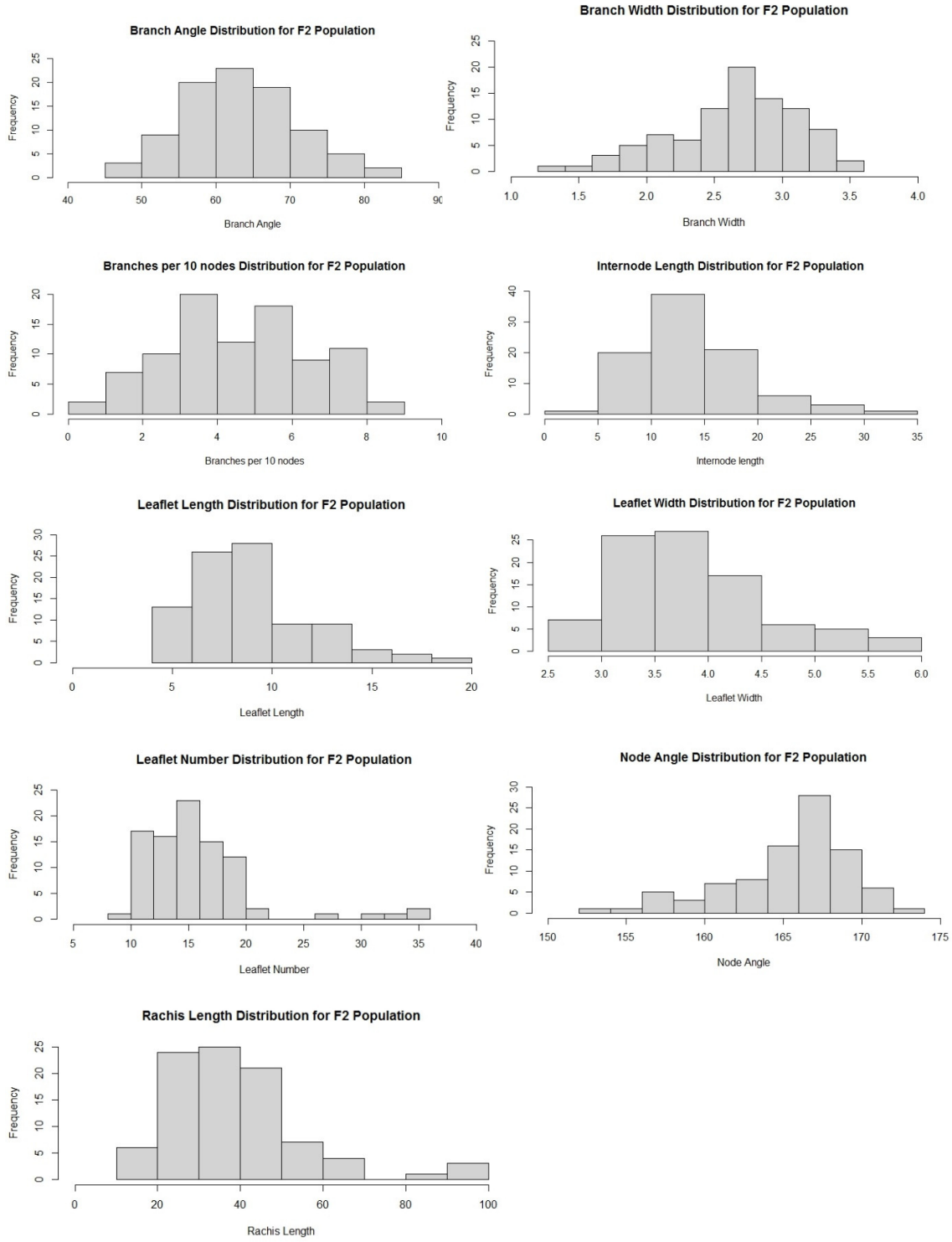
Appendix 2.3. Boxplots for branch traits for the maternal cross direction F_1. *S.p* indicates *S. prostrata* parent, *S. t* indicates *S. tetraptera* parent. *S.p* F_1 indicate F_1 with *S. prostrata* as maternal parent and *S.t* F_1 indicates F_1 with *S tetraptera* as a maternal parent.



Appendix 2.5. Boxplots for leaf traits for the maternal cross direction F_1. *S.p* indicates *S. prostrata* parent, *S. t* indicates *S. tetraptera* parent. *S.p* F_1 indicate F_1 with *S. prostrata* as maternal parent and *S.t* F_1 indicates F_1 with *S tetraptera* as a maternal parent.



Appendix 2.6. Boxplots for the Lincoln F_2 maternal cross direction populations and parents. P indicates F_2 with *S. prostrata* as maternal parent, P1 indicates *S. prostrata,* T indicates F_2 with maternal parent *S. tetraptera* and T1 indicates *S. tetraptera.*



Appendix 2.7. Histogram distributions for the traits measured in the Lincoln F_2. Branch and node angle (°), measurements made in mm units.

3 Chapter Three – Strigolactone and Divarication

3.1 Abstract

Sophora (Fabaceae) is one of many genera in New Zealand that has species with a divaricate growth form: *S. prostrata* is an obligate divaricate, and *S. microphylla* is a heteroblastic species with a divaricating juvenile stage. Divaricating plants are described as small trees or shrubs with small leaves, interlacing branches with wide branch angles yielding a tangled shrubby appearance. Currently nothing appears to be known about the genetic basis behind this plant growth form. Increased branching is one trait consistently used in describing the divaricate growth form and contributes to the dense tangled appearance of these plants. Regulation of axillary shoot branching across land plants is controlled by a family of plant hormones termed strigolactones.

Initial work that described the strigolactone pathway utilised *Pisum sativum* (Fabaceae) mutants with increased axillary branching and identified the *RAMOSUS (RMS)* loci, that are responsible for biosynthesis, signal perception and response of strigolactone in control of shoot branching. These mutants display several phenotypes that are similar to many of the phenotypes seen in the divaricate plant architecture. The similarity of phenotypes between *Pisum rms* mutants and the divaricate form suggest this pathway may contribute to the architecture of divaricate plants. Orthologs of four of the five *RMS* genes were isolated and partially sequenced from *Sophora prostrata* and *S. tetraptera*. Two F_2 populations were then genotyped at each of these loci. Each locus was found to be associated with variation at several of the divarication traits, suggesting the strigolactone pathway may be involved in the evolution of the divaricating form in *Sophora*. In comparisons between *S. prostrata* and *S. tetraptera,* an amino acid replacement in RMS1 was predicted as a deleterious change in the protein for *S. prostrata*. Experiments were conducted applying synthetic strigolactone, GR24, to *S.*

prostrata seedlings and while some changes to leaflet length and width were observed, no significant differences in branch characteristics were observed compared to control plants. The *RMS* markers are also included in linkage maps and QTL analysis (subsequent chapters) where there may be greater power to detect significant associations.

3.2 Introduction

Plant architecture describes the three-dimensional organisation of the plant body and is a composite of branching patterns and size and organization of leaves and reproductive structures (Reinhardt & Kuhlemeier, 2002). It is often used in taxonomy to aid in the classification of taxa and is of importance in agriculture where it affects features such as plant stability in the soil (Quine, 1990; Plourde et al., 2009) and harvest efficiency (Takeda et al., 2013). Plant architecture is determined at the meristems as they give rise to the structures that contribute to the overall plant form, including axillary branches (Chomicki et al., 2017). While plant architecture is regulated by many plant hormones (Shimizu-Sato & Mori, 2001; Eichhorn et al., 2005), the strigolactone hormone plays a pivotal role. Strigolactones are a class of hormones derived from carotenoids; approximately twenty different strigolactone hormones have been identified in plants (Xie et al., 2013). These are known to be involved in seed germination (Cook et al., 1966) by stimulating hyphal growth of mycorrhizal fungi (Akiyama et al., 2005; Xie et al., 2010) and in shoot branch control (Gomez-Roldan et al., 2008; Umehara et al., 2008). Further possible roles in plants have been suggested for strigolactones including in the rhizobium-legume interaction (Soto et al., 2010) and root and root hair growth (Koltai, Dor, et al., 2010).

The involvement of strigolactone in shoot branching was first identified from studies of *Arabidopsis MAX* mutants (Gomez-Roldan et al., 2008), rice *dwarf (d)* mutants (Umehara et al., 2008), and *Pisum RAMOSUS* (Brewer et al., 2009) mutants. These mutants were initially identified by increased branching phenotypes; however, some also displayed other phenotypes including smaller leaves, shorter internodes and thinner stems (Beveridge et al., 1997; Morris et al., 2001; Sorefan et al., 2003; Booker et al., 2004; Auldridge et al., 2006; de Saint Germain et al., 2016). Five mutants were identified and characterised in *P. sativum*, *RMSI-5,* involved in biosynthesis, perception and response of

the strigolactone shoot branch pathway (Stirnberg et al., 2007; de Saint Germain et al., 2016; Flematti et al., 2016; Yao et al., 2016; Ligerot et al., 2017) (Figure 3.1). Five of these mutants have since been cloned. *RMS1* and *RMS5,* known as *MAX4* and *MAX3,* respectively, in *Arabidopsis,* are involved in the biosynthesis of strigolactone in the roots (Flematti et al., 2016). *RMS3* perceives strigolactone and binds covalently to the D ring of strigolactone causing cleavage of the D ring and a conformational change of RMS3 (de Saint Germain et al., 2016; Yao et al., 2016). This enables binding to *RMS4* and assembly

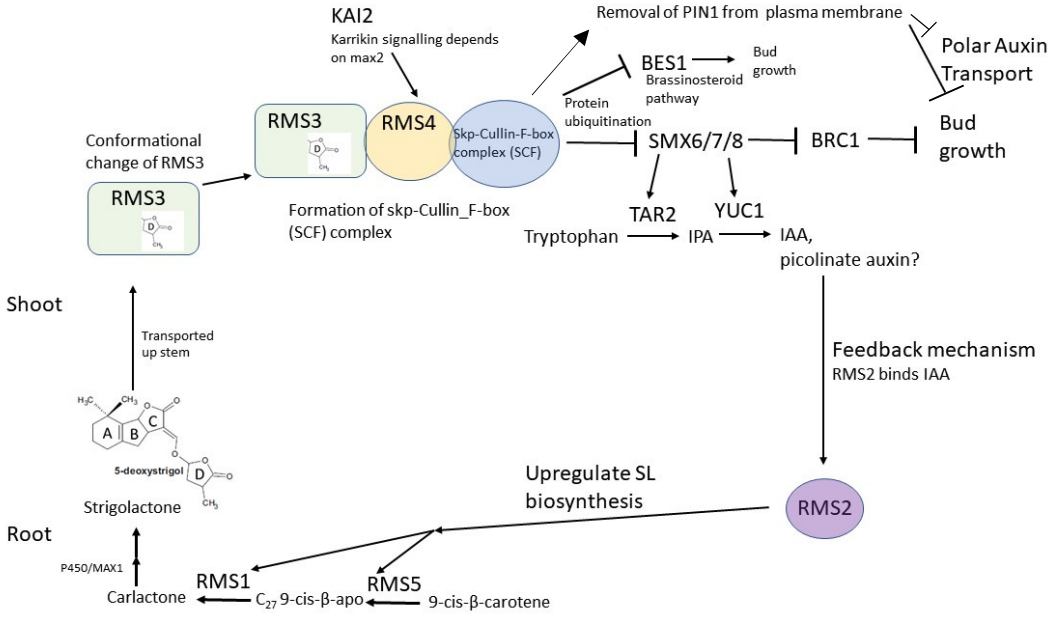


Figure 3.1. The current model of the RMS pathway (modified from Beveridge, 2006; Dun et al., 2009; Braun et al., 2012; Ligerot et al., 2017). *RMS1* and *RMS5* are required together to produce a novel long distance signal, strigolactone, which can only move unidirectional up the shoot. RMS3 hydrolyses the ABC lactone and D ring of strigolactone and binds to the D ring causing a conformational change. This allows the formation of the SCF complex by binding to RMS4. RMS3 and RMS4 are both required together and are involved in signal perception. The SCF complex targets proteins for degradation including BES1, involved in the Brassinosteroid pathway, and SMX6/7/8. RMS4 also integrates signals from KAI2, involved in the karrikin pathway, and removal of PIN1 is associated with strigolactone signalling. The degradation of SMX6/7/8 removes the repression of BRC1, which acts as a repressor to bud growth. RMS4 also acts to repress RMS1, independently of the long distance signals, by suppressing the feedback signal involving RMS2 (Quine, 1990; Dun et al., 2009). When signalling from strigolactone is lacking, SMX6/7/8 upregulate biosynthesis of auxin. RMS2 binds to auxin and acts to upregulate RMS1 and RMS5 in the rootstock and control cytokinin levels in the shoot, which inhibits promotion of axillary bud growth as cytokinin in known to be involved in promotion of branching of axillary buds. RMS3 and RMS4 also act independently in the root to inhibit RMS1 and RMS5 expression.

of the SCF complex (Skp Cullin F box complex) (Woo et al., 2001; Kuroda et al., 2002; Stirnberg et al., 2007), which is involved in protein ubiquitination of downstream targets, such as SMX proteins (Jiang et al., 2013; Bennett & Leyser, 2014; Soundappan et al., 2015). The SMX genes, *SMX6/7/8,* when expressed, repress BRC1 expression, signalling axillary bud growth as BRC1 represses bud growth. In the absence of a strigolactone signal the SMX proteins also are involved in upregulation of the genes *TAR2* and *YUC1* (Ligerot et al., 2017), which are involved in auxin biosynthesis. *RMS2* binds to the auxin produced and acts as a feedback mechanism upregulating *RMS1* and *RMS5* (Peng et al., 1999; Dun et al., 2009; Braun et al., 2012; Ligerot et al., 2017). Orthologs for the *RMS* genes have been identified in a wide range of plants, including willow (Salmon et al., 2014), poplar (Czarnecki et al., 2014) and petunia (Johnson et al., 2006), indicating that the genetic control of branching is largely conserved among diverse species.

In the New Zealand flora, a dramatic shift in plant architecture has occurred multiple times across diverse groups of plants. In at least 18 plant families, that span ~ 15 orders including the Rosales, Apiales and Malpighiales and one gymnosperm order, Pinales, a unique architecture, termed divarication, has arisen. Compared to their arborescent relatives, divaricate species have reduced leaf size with increased branching that occurs at wide angles, yielding a small tree or shrub with interlacing branches (Kelly, 1994). Currently nothing is known about the molecular genetic nature of divaricating forms; however, the strigolactone mutants in model plant species display phenotypes similar to many features of the divaricate form, such as increased branching, smaller leaves and thinner stems. These similarities and the conserved nature of the strigolactone pathway across diverse plant groups suggest that strigolactone biosynthesis and perception pathways may be a good candidates for the regulation of the divaricate growth form.

This chapter reports on a diverse set of investigations largely focused on the involvement of strigolactones on the divaricate growth form in *Sophora prostrata* , relative to its non-divaricating congener, *S. tetraptera* (Figure 3.2). These investigations include (1) the isolation and sequencing of partial *RMS* orthologs from *S. prostrata* and *S. tetraptera*, (2) marker development from these gene sequences and another unrelated candidate gene, *ZIG,* (3) a new F_2 population developed from seed collected from existing F_1 plants (these were grown in a greenhouse in more controlled conditions than the original F_2 population), (4) phenotyping of the new F_2 population for the same traits as those described for the Lincoln population in the previous chapter (Chapter 2), and (5) the involvement of strigolactone in the divaricate form of *S. prostrata* investigated by application of synthetic strigolactone, GR24, to *S. prostrata* seedlings.

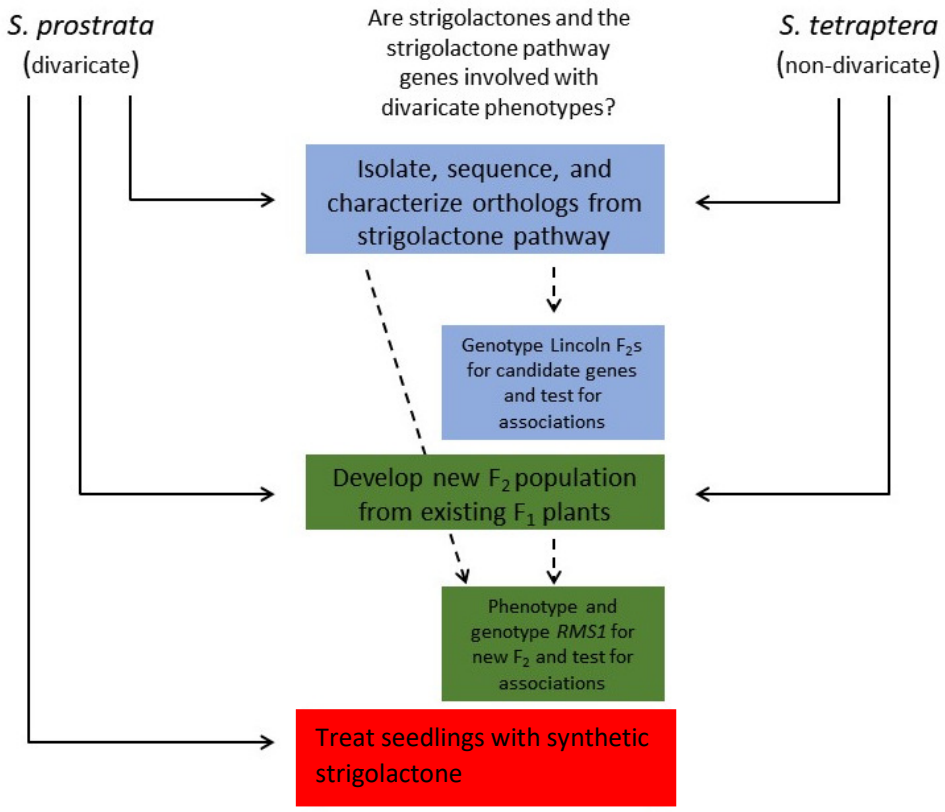


Figure 3.2. Diagram of the layout of this chapter. The colours represent different major sections that are presented to investigate the involvement of strigolactone on the divaricate growth form.

3.3 Materials and Methods

3.3.1 Isolation and sequencing of *RMS* genes and *ZIG* in *Sophora prostrata* and *S. tetraptera*

Plant material and DNA extractions

Plant tissue was collected from the Lincoln F_2, and parents at the 'East Block' site in Lincoln as described in Chapter 2. Tissue collection for the F_1 plants was performed at the Manaaki Whenua Landcare Research campus in Lincoln as described in Chapter 2. DNA extractions for the Lincoln F_2, F_1, and parents were described in Chapter 2. The DNA from parent species, *S. tetraptera* and *S. prostrata,* was used in the isolation and sequencing performed in this chapter.

RMS1/MAX4

Pisum sativum RMS1 gDNA sequence was downloaded from Genbank identified by the accession number AY557341 (Foo et al., 2005). BLAST searches using the *PsRMS1* sequence were used to retrieve sequences for *RMS1* orthologs from the Papilionoideae (Fabaceae). Sequences from *Arachis* (XM_016090204 and XM_016326324), *Medicago* (XM_013593530), *Cicer* (XM_004501100), *Pisum* (AY557341 and AY557342), *Glycine* (XM_003522665), *Phaseolus* (XM_007137159) and *Vigna* (XM_014644999) were downloaded from Genbank. These sequences were aligned in Geneious 9.1.8 (https://www.geneious.com) and multiple forward and reverse primers were designed (Table 3.1; primers F1, R1, F3 and R3) from the most conserved regions with degenerate nucleotides (IUPAC ambiguity codes) to allow for variation among taxa.

The initial primer combination (RMS1_F1 and RMS1_R1) gave multiple bands (Figure 3.3) so, of the multiple primers designed, those that were internal to the initial primers (RMS1_F3 and RMS1_R3) were used on the product of the first PCR (nested PCR)

to give a single band that could be directly sequenced via Sanger sequencing. As the length of the products was ~ 3 kb, a full length sequence could not be obtained from one sequencing reaction. New primers, specific to *Sophora,* were designed from the *Sophora* sequence, allowing some overlap between sequence fragments in order to align these into one large gene fragment (Table 3.1 and Figure 3.4). This process, primer walking (designing new primers from previous sequence) was repeated two more times to complete sequencing of the initial 3kb product. Sequencing of the entire 3 kb product required four different PCR (referred to as PCR reaction, Table 3.1) and sequencing reactions for *S. prostrata* and five for *S. tetraptera,* as the product of the latter was a larger size (Figure 3.4).

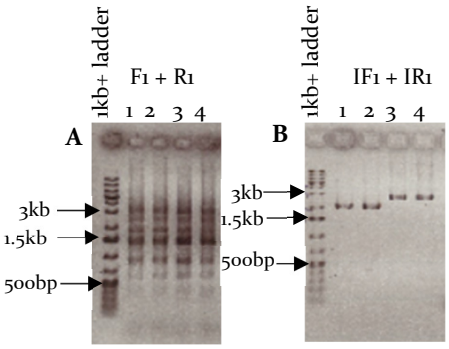


Figure 3.3. Example of PCR to obtain *RMSI* sequence. The initial PCR step displaying multiple bands (A) and a single band after nested PCR (B). Samples 1 and 2 represent *S. prostrata*. *S. tetraptera* samples are labelled 3 and 4. Black arrows indicate band sizes on the size standard Generuler 1kb Plus DNA ladder from Thermo Scientific (Thermo Fisher Scientific).

The initial PCR amplification was performed in a 10 µL volume with 1x HF Phusion buffer (Thermo Fisher Scientific), 50 µM of each dNTP, 0.5 µM of forward primer, 0.5 µM of reverse primer, 3% DMSO, 1.0 units/50 µL Phusion Taq (Thermo Fisher Scientific) and 1 µl gDNA. Amplification by PCR for the first and second PCR reactions was obtained by: 95°C for 3 minutes, then 35 cycles of 95°C for 30 seconds, 57°C for 30 seconds, 72°C for 3 minutes, followed by 72°C for 10 minutes. The PCR protocol for PCR 3 (Table 3.1), with the first specific *Sophora* primers, was the same except that the annealing temperature was 53°C. Amplification for PCR 4 was 95°C for 3 minutes, then 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 2 minutes, followed by 72°C for 10 minutes. As the *S. tetraptera* fragment is around 1kb larger than *S. prostrata* new primers specific to *S. tetraptera* were designed from the previous

sequence to amplify and sequence the rest of *RMSI* from *S. tetraptera*. Amplification for the fifth PCR of *S. tetraptera* was performed by 95°C for 3 minutes, then 35 cycles of 95°C

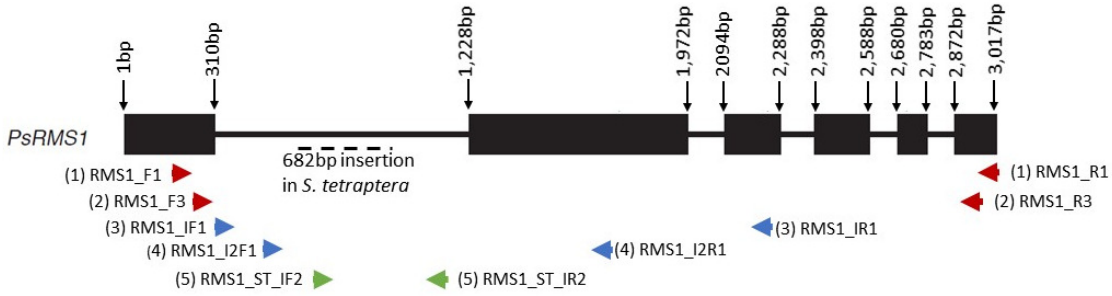


Figure 3.4. *RMSI* gene schematic indicating positions of primers used to amplify and sequence RMSI in *Sophora*. Positions of primers and exons are based on *Pisum sativum RMSI* gene annotations (AY557341). Introns indicated by the solid black line, exons indicated with solid black rectangles. The insertion in the intron of *S. tetraptera* indicated with a dashed line. Initial primers (with degenerate nucleotides) designed from conserved regions in the Papilionoideae are indicated with red arrows. The numbers in brackets next to primer names indicate the order the primer pairs were used with (1) indicating the initial PCR that produced multiple bands and (2) indicating the following nested PCR to give a single band. The blue arrows indicate primers designed specific to *Sophora* from the previous sequencing used to sequence the internal portions of the *RMSI* fragment. Green arrows represent the primers specific to *S. tetraptera* that enabled sequencing of the large insertion in *S. tetraptera*.

for 30 seconds, 45°C for 30 seconds, 72°C for 30 seconds, followed by 72°C for 5 minutes.

RMSI sequencing

To prepare for sequencing 8 µL of PCR product was purified in a 16.5 µL reaction using 0.25 µL of Exonuclease I (20 U/µL) (Thermo Fisher Scientific), 0.5 µL of FastAP Thermosensitive Alkaline Phosphatase (1 U/µL) (Thermo Fisher Scientific) and 7.75 µL of sterile water. This reaction was heated to 37°C for 30 minutes and then 80°C for 15 minutes. The purified product was run on a 1% agarose gel, stained with ethidium bromide and visualised under UV light to infer the concentration needed for sequencing reactions. The product was sent to the Massey Genome Service for sequencing on the ABI3730 DNA analyzer. Sequencing was completed from both 5` and 3` ends of the PCR templates.

Table 3.1. Primers used in amplifying and sequencing the *RMS* and *ZIG* candidate genes fragments.

Primer	Sequence	PCR	Tm (°C)	%GC	Resulting product
RMSI_F1	AyrkyGCATGGACTAGTrTC	1	45-55	35	Multiple bands
RMSI_R1	ACCCAGCATCCATGyAACCC	1	53-56	55	Multiple bands
RMSI_F3	yvbCAAGArAGvTGGGAAGG	2*	47-58	40	Single band
RMSI_R3	AywAykGAGATTACCACrCC	2*	45-54	35	Single band
RMSI_IF1	CTTGTGTCTCTTATGACAG	3	46.8	42.1	Single band
RMSI_IR1	GCTTATCAAGAATGGCGG	3	48	50	Single band
RMSI_I2F1[1]	TGTCAGTTGAGTTATAC	4	45.5	35.3	Single band
RMSI_I2R2[1]	GTGTGCAGAATGAATGAG	4	51.4	44.4	Single band
RMSI_St_I_F2	CCATAGGAATTAGCAAC	5	42.2	41.2	Single band
RMSI_St I_R2	CAATGTTATGTAGTGAC	5	39.8	35.3	Single band
RMS2_F1	TCATTTGGAGTCTCTTG	1	47.6	41.2	Single band
RMS2_R1	CTACTGCAkkATGGTAAC	1	49.7	44.4	Single band
RMS4_F3	yTnCCGsmGGAGATyyTGwC	1	49-60	40	Multiple bands
RMS4_R3	TCyGTrCTCATrTCAyTCTC	1	45-54	35	Multiple bands
RMS4_F1	TGCGrGGnAACGCGCGTGAC	2*,4	57-62	65	Two bands
RMS4_R1	CTCTGGTTsACATCyTCATC	2*,3	49-52	40	Two bands
RMS4_IF1	TTACGCTTCTTCACCTCG	3	48	50	Single band
RMS4_IR1	CAATCCACATGAAGCCGC	4	50.3	55.6	Single band
RMS5_F2	TwGGGAAyACkAAGGThATG	1	45-52	35	Multiple bands
RMS5_R2	rTkvmTyGGAATTTCrAkyC	1	39-56	20	Multiple bands
RMS5_F1	yTrTGyATGTGGGAAGGTGG	2*	49-56	45	Single band *S. prostrata*
RMS5_R3	AGAAkrTArCCATCrTCTTC	2*	43-52	30	Single band *S. prostrata*
RMS5_SP_EF1	GGTGGTGATGTTTGGGAGG	3,5	53.2	57.9	*S. prostrata* band
RMS5_SP_IR2	TCCACATACCGTTGGATC	3,6,7	48	50	*S. prostrata* band
RMS5_Ex2_F	AAGATGCCmCCAArGAG	4	44-50	47.1	*S. prostrata* band
RMS5_Ex4R	CCAyTGrTAAGArCAAGC	4	43-50	38.9	*S. prostrata* band
RMS5_SP_Ex2AR[S]	TACTACGTGGAATCAAC	5	42.2	41.2	*S. prostrata* band
RMS5_SP_Ex4AF[S]	GTGGCTACTTCATGTTGG	6	48	50	*S. prostrata* band
RMS5_SP_Ex3AF[S]	CCTGCAGAACATGACTC	7	47.1	52.9	*S. prostrata* band
RMS5_SP_Ex3AF[S]	CCTGCAGAACATGACTC	8	47.1	52.9	*S. tetraptera* band
RMS5_SP_Ex4R[S]	CCAyTGrTAAGArCAAGC	8	43-50	38.9	*S. tetraptera* band

Primer	Sequence	PCR	Tm (°C)	%GC	Resulting product
iF_ZIG	AATGGACYTYGAGGCAAG	1	45-50	44.4	Single band
iR_ZIG	CCAYTTGTTCTTGTTCATC	1	44-47	36.8	Single band

* used previous product as template, [1] Primers used to genotype *RMS1*, PCR indicates order primers were used to sequence the partial fragment of each gene, [S] indicates *RMS5* primers designed that are specific to *S. prostrata*.

RMS2

PsRMS2 sequence was obtained from Genbank with the accession number MG495397 (Ligerot et al., 2017). BLAST searches restricted to the Papilionoideae using *PsRMS2* sequence yielded orthologs from *Pisum* (MG495397), *Medicago* (CU475569), *Lotus* (AP004968) and *Cicer* (XM_004490351). Primers were designed from conserved regions in alignments of these sequences (Table 3.1). These primers were then tested in *S. prostrata* and *S. tetraptera* to amplify a region (of ~ 1,200 bp based from the *Pisum* sequence length between designed primers) of *RMS2*. Initial PCR amplification was performed in a 10 µL volume as described for *RMS1*. Amplification by PCR was obtained by: 95°C for 3 minutes, then 35 cycles of 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute, followed by 72°C for 5 minutes. This produced a single product from each *Sophora* species, which was used as template for sequencing, as described for *RMS1*, using the same forward and reverse *RMS2* primers.

RMS4/MAX2

Pisum sativum RMS4 sequence was obtained from Genbank identified by the accession number DQ403159 (Johnson et al., 2006). BLAST searches restricted to the Papilionoideae using *PsRMS4* sequence yielded orthologs from *Phaseolus* (XM_007132787), *Arachis* (XM_016080930), *Vigna* (XM_014638348), *Medicago* (XM_003607544), *Cicer* (XM_004505434) and *Pisum* (DQ_403159). Primers were designed from conserved regions in alignments of *RMS4* sequences (Table 3.1). Initial

PCR amplification was performed in a 10 µL volume as described for *RMS1.* Amplification by PCR was obtained by: 95°C for 3 minutes, then 35 cycles of 95°C for 30 seconds, 58°C for 30 seconds, 72°C for 2 minutes, followed by 72°C for 10 minutes. Following initial and internal PCRs, two separate bands (~ 1,500 and 1,750 bp estimates) were still produced. To isolate the single product of the correct size (~ 1,750 bp predicted from the *Pisum* sequence based on where primers situated) PCR products were separated by gel electrophoresis on a 0.8% agarose gel for 2 hours, stained with ethidium bromide and visualised using an Invitrogen Safe Imager Blue Light Transilluminator (Thermo Fisher Scientific). Individual bands were excised using a razor blade and the excised fragments were purified using the Zymo Gel Extraction Kit (Zymo Research Corporation). These bands were sequenced as described above for *RMS1* using the primers RMS4_F1 and RMS4_R1. Sequences from the gel excised fragments, (~ 1 kb), were BLASTed and based on sequence matches, identified to be the correct *RMS4* orthologs. Primers were designed from the *Sophora* sequences (Table 3.1) and the gel excised fragment was used as a template for a further PCR to sequence the remaining *RMS4* fragment using the same forward and reverse primers. Amplification by PCR was obtained by: 95°C for 3 minutes, then 35 cycles of 95°C for 30 seconds, 58°C for 30 seconds, 72°C for 90 seconds, followed by 72°C for 5 minutes for PCR step 3 with primers RMS4_IF1 and RMS4_R1 and for step 4, with primers RMS4_F1 and RMS4_IR1, the annealing temperature was 65°C.

RMS5/MAX3

Pisum sativum RMS5 sequence was downloaded from Genbank identified from the accession number DQ403160 (Johnson et al., 2006). BLAST searches using *PsRMS5* sequence, identified from were restricted to the Papilonoideae and yielded orthologs from *Arachis* (XM_016116193), *Glycine* (XM_003516754), *Vigna* (XM_014642045 and XM_017562054), *Lotus* (GU441766), *Pisum* (DQ403160), *Cicer* (XM_004513878),

Trifolium (KJ127514) and *Medicago* (XM_003622507). Primers were designed from conserved regions in sequence alignments of these *RMS5* sequences (Table 3.1). PCR amplification was performed in a 10 µL volume as described for *RMS1*. Amplification by PCR was obtained by: 95°C for 3 minutes, then 35 cycles of 95°C for 30 seconds, 57°C for 30 seconds, 72°C for 3 minutes, followed by 72°C for 10 minutes. After the second PCR reaction only *S. prostrata* produced a band that could be sequenced. Primers, specific to *S. prostrata,* were developed from this sequence, however, the large size of the first intron, which contained a high frequency of single nucleotide repeats, made sequencing difficult so primers specific to each exon were also designed from the original alignment of Papilonoideae and amplified in *Sophora*. Amplification in *S. tetraptera* with the consensus primers was not successful, therefore, primers specific to the *S. prostrata* sequence were designed and used to amplify *RMS5* in *S. tetraptera* as previous *RMS* gene sequences had shown these species to only differ by a few SNPs. Gel extraction using the Zymo gel extraction kit was performed, as described for *RMS4,* for the band of expected size (based on the size of *Pisum RMS5* according to the position of the primers) for *S. tetraptera* and sequenced, as described for *RMS1,* with the *S. prostrata* specific *RMS5* primers RMS5_3AF and RMS5_4R. The product of each PCR with single bands was used as a template for sequencing, as described for *RMS1,* using the forward and reverse primers from that PCR reaction.

ZIG

Primers for *ZIG* (Table 3.1) were previously designed (Symonds, unpublished) for *Sophora* from consensus sequence including *Arabidopsis* (NM_123313), *Capsella* (XM_006291751), *Eutrema* (XM_006395245), *Nicotiana* (XM_009593661), *Camelina* (XM_010452500), *Brassica* (XM_013734377 and XM_022703989), *Arachis* (XM_016090671, XM_025826670 and XM_016325864), and *Raphanus* (XM_018602866)

. PCR amplification was performed in a 25 µL volume with 1x Thermopol buffer (New England BioLabs Inc), 50 µM of each DNTP, 0.2 µM of forward primer, 0.2 µM of reverse primer and 1.25 units/50 µL *Taq* DNA polymerase (New England Biolabs Inc). Amplification by PCR was obtained by: 95°C for 5 minutes, then 35 cycles of 95°C for 30 seconds, 55°C for 40 seconds, 72°C for 90 seconds, followed by 72°C for 5 minutes. PCR products were used as templates for sequencing, as described for *RMSI,* with the same forward and reverse primers.

3.3.2 Sequence analysis of candidate genes

For each gene of interest, the individual sequences for each species were processed independently in Geneious 9.1.8 (https://www.geneious.com) (Kearse et al., 2012). All sequences from a given gene were aligned, trimmed, and, where necessary, edited to form a single contiguous sequence based on the consensus. The final gDNA sequences from *S. prostrata* and *S. tetraptera* were then aligned with one another to identify SNPs and indels and aligned with the relevant annotated *Pisum sativum* ortholog to infer intron/exon boundaries. To examine the sequences for potential amino acid replacements, inferred introns were spliced out of the alignments and the translations examined. Once translated the RMS protein sequences were input in *fasta* format into PROVEAN web server (Choi & Chan, 2015). The default delta score cut-off for significance of -2.5 was used. Amino acid polymorphisms between the *Sophora* species were specified to the PROVEAN tool for analysis of their predicted effect. To further examine an RMS1 site predicted to have a deleterious replacement in the *S. prostrata* sequence, translated sequences were downloaded from Genbank for *Arabidopsis thaliana* (Brassicaceae) (NP_001329787), *Cucumis melo* (Cucurbitaceae) (XP_008445014), *Daucus carota subsp sativus* (Apiaceae) (XP_017243247), *Glycine max* (Fabaceae) (AQY54424), *Jatropha cucrus* (Euphorbiaceae) (NP_001306853), *Nicotiana*

tabacum (Solanaceae) (NP_001312826), *Oryza sativa* (Poaceae) (XP_015642760), *Physcomitrella patens* (Funariaceae) (ADK36681), *Pisum fulvum* (Fabaceae) (MH068753), *Pisum sativum* (Fabaceae) (AAS66906) , *Selaginella moellendorffi* (Selaginellaceae) (XP_002988101), *Sesbania cannibana* (Fabaceae) (AWK40013) and *Zea mays* (Poaceae) (PWZ31381) and were aligned in Geneious 9.1.8 with the *Sophora* sequences.

No *ZIG* ortholog could be identified from *Pisum sativum* so the *Arabidopsis thaliana ZIG (AtVTI1a)* sequences were used in the alignment with *Sophora ZIG* sequences.

3.3.3 Marker development for *RMS* genes in *S. prostrata* and *S. tetraptera* *RMS1*

Previously developed primers (Tables 3.2) that yielded PCR product size differences between *S. tetraptera* and *S. prostrata* were used to genotype F_2 samples. Primers were tested on the four parent samples, four F_1 samples, and a subset of seven F_2 samples. All samples were amplified in a 10 µL volume with 1x HF Phusion buffer (Thermo Fisher Scientific), 50 µM of each dNTP, 0.5 µM of forward primer, 0.5 µM of reverse primer, 3% DMSO and 1.0 units/50 µL Phusion Taq (Thermo Fisher Scientific). Amplification by PCR was obtained by: 95°C for 3 minutes, then 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 3 minutes, followed by 72°C for 5 minutes. PCR products were run on a 1% agarose gel, stained with ethidium bromide, and photographed under UV. Bands were scored from the gels based on PCR product size; the *S. tetraptera* allele is larger than the *S. prostrata* allele (Figure 3.3).

Table 3.2. Primers used to genotype *RMS* and *ZIG* genes for *Sophora*

Primer name	Allele	Sequence	Tm	%GC
RMS1_I2F1	-	TGTCAGTTGAGTTATAC	45.5	35.3
RMS1_I2R2	-	GTGTGCAGAATGAATGAG	51.4	44.4
RMS2_1257_St_A1F	ST	GGAACAGGATCTTTT**G**GG	50.4	50
RMS2_1261_St_A5R	ST	GACCTACCTCAGAAA**T**CA	47.8	44.4
RMS2_1258_SP_A2F	SP	GGAACAGGATCCTTT**G**GT	51.7	50
RMS2_1262_SP_A6R	SP	GACCTACCTCAGAAA**T**CG	49.2	50
RMS4_1265_ST_A9_F	ST	CGCCGGTACTATCC**T**G	51.8	62.5
RMS4_1271_ST_A13_R	ST	GAGAATAACAAGCC**G**GG	47.1	47.1
RMS4_1266_SP_A10_F	SP	CGCCGGTACTATCC**T**C	51.5	62.5
RMS4_1272_SP_A14_R	SP	GAGAATAACAAGCC**G**GA	45.6	41.2
RMS5_1283_ST_A29F	ST	GCGAGAAAGTAATGATATATC**C**AT	51	33
RMS5_1285_ST_A27R	ST	GTGCTTCAATTGGC**G**CA	54.3	52.9
RMS5_1284_SP_A30	SP	GCGAGAAAGTAATGATATATC**C**AC	52	38
RMS5_1286_SP_A28R	SP	GTGCTTCAATTGGC**G**CT	54	52.9
ZIG_1289_ST_A33_F	ST	CATCCTTTACTGATGT**C**CA	48.9	42.1
ZIG_1293_ST_A37_R	ST	CCATTTAATAGATCACT**C**CT	46.3	35.0
ZIG_1290_SP_A34_F	SP	CATCCTTTACTGATGT**C**CC	49.9	47.4
ZIG_1294_SP_A38_R	SP	CCATTTAATAGATCACT**C**CC	47.5	40.0

Allele refers to the species the primers were designed to amplify, ST represents *S. tetraptera* and SP represents *S. prostrata* specific primers. Intentional mismatch designed in primers represented in bold.

RMS2

For *RMS2,* a marker system was designed around SNP polymorphisms that distinguish *S. tetraptera* and *S. prostrata* alleles. Species-specific forward and reverse

primers were designed (Table 3.2) from *Sophora* sequences with the 3' end of each primer ending at the SNP. The primers included a second intentional mismatch at the 2nd or 3rd site (Table 3.2), from the SNP, based on the recommendations in Liu *et al.,* 2012. Both parent species, F_1 and a subset of seven F_2 samples were amplified for both species (allele)-specific primer pairs in separate reactions. PCR reactions were performed in a 10 µL volume with 1x Thermopol buffer (New England BioLabs Inc), 50 µM of each DNTP, 0.2 µM of forward primer, 0.2 µM of reverse primer and 1.25 units/50 µL *Taq* DNA polymerase (New England Biolabs Inc). Amplification by PCR was obtained by: 95°C for 3 minutes, then 35 cycles of 95°C for 30 seconds, 55°C for 30 seconds, 68°C for 40 seconds, followed by 68°C for 5 minutes. Products were run on a 1% agarose gel and visualised with ethidium bromide under UV light. Genotypes were scored as the presence/absence of bands from both reactions with homozygotes having only one successful PCR product and heterozygotes having successful PCR products in both reactions.

RMS4

SNP-based allele specific primers (Table 3.2) were designed based on *RMS4* sequences as they were for *RMS2* above. F_1, F_2, and parent samples were amplified with the *RMS4* primer pairs in a 10 µL as described above for *RMS2*. Amplification by PCR and gel visualisation were performed as described above for *RMS2* but with an extension temperature at 68° for 1 minute and 20 seconds.

RMS5

Allele specific SNP-based primers (Table 3.2) were designed from partial *RMS5* sequences from *S. tetraptera* and *S. prostrata* as described above for *RMS2*. All samples were amplified in a 10 µL volume as described for *RMS2* above. Amplification by PCR

was as described above for *RMS2* except with an annealing temperature of 54°C for 30 seconds. PCR products were visualised and scored as described for *RMS2* above.

ZIG

SNP-based allele-specific primers (Table 3.2) were designed from partial *ZIG* sequences from *S. tetraptera* and *S. prostrata* as described for *RMS2* above. All samples were amplified with the *ZIG* primer pairs in a 10 µL PCR reaction as described for *RMS2*. Amplification by PCR was obtained by: 95°C for 3 minutes, then 35 cycles of 95°C for 30 seconds, 50°C for 30 seconds, 68°C for 30 seconds, followed by 68°C for 5 minutes. PCR products were visualised and scored as described for *RMS2*.

3.3.4 Genotyping Lincoln F_2 *S. prostrata x S. tetraptera* population

Tissue collection, DNA extraction and phenotype data collection for the F_2 population located in Lincoln were described in Chapter 2. All Lincoln F_2 were genotyped for the *RMS* and *ZIG* candidate genes using the methods described above. ANOVA tests were performed for each of the candidate genes using the basic stats package in R v.3.4.2 (R Development Core Team, 2016). A Bonferroni correction (Dunn, 1961) was performed on the ANOVA results to account for multiple testing performed in Microsoft Excel (2016). Levene's test (Brown & Forsythe, 1974) was performed using the *leveneTest* function in the car v. 3.0-2 package to test for homogeneity of variance between tested groups. As all the data met the assumption of homogeneity of variance, the post hoc test Tukeys HSD (Tukey, 1949) was performed using the *TukeyHSD* function in the stats v. 3.4.2 package. ANOVA was also performed for the maternal direction populations as described above.

3.3.5 Generation of a new *S. prostrata x S. tetraptera* F_2 population: the Palmerston North (PN) F_2

A second F_2 population of 69 individuals was developed from seed collected from the 10982-9 F_1 individual (see Chapter 2 for details on this individual) to generate a population of young plants grown in a more controlled environment than the Lincoln F_2 population.

As seed germination is very slow for untreated *Sophora* seed, small scale experiments were conducted to explore the effects of H_2SO_4 treatments as this has been reported to improve germination of the congener *Sophora secundiflora* (Fabaceae) (Ruter & Ingram, 1991). Based on those initial tests (data not shown), the best gemination rates were determined to take place using concentrated H_2SO_4 for 45 minutes, followed by a 2 hour wash with water. Following treatment, seed were placed in petri dishes that contained filter paper and MilliQ water and put under growth lights to stimulate germination.

In December 2016, several hundred seed were collected from the F_1 individual 10982-9 in Lincoln (see Chapter 2 for details on this plant's lineage). One-hundred seed were treated as described above. Following germination (radicle emergence), the seed coat was manually removed and the embryos then transferred into individual pots in the greenhouse. The plants were maintained in greenhouse conditions (Massey University Plant Growth Unit in Palmerston North) and watered every second day.

3.3.6 Microsatellite screening of PN F_2 population to confirm parentage

To secure F_2 seed that were the products of F_1 self-fertilization, dozens of flowers on multiple F_1 plants were bagged prior to anthesis; however, seed set from these bagged flowers was very low and did not provide a sufficient number of seed. Therefore, mature seed from unbagged flowers were collected (as in above section). While the maternal

parentage of these seed was unambiguous, the paternal parentage was uncertain, therefore, the F_2 plants started from those seed were screened with a small number of microsatellite markers to test for deviations from the maternal parent's combination of alleles. Four of the microsatellites (Van Etten et al., 2014) used to identify the Lincoln F_2 population (see Chapter 2) were genotyped (Table 3.3) in the PN F_2 population as these four identified alternative alleles from the plants identified as recruiters in the Lincoln F_2 population. Microsatellite genotyping was performed as per the Lincoln F_2 (see Chapter 2).

Table 3.3. Microsatellite primers used to genotype F_2 (primer details from Van Etten et al. 2014).

Locus		Primer sequence (5'-3')	Repeat motif	Size range (bp)	*Ta* (°C)
Sop-42	F	CCATACCTGACACTTGCGG	$(AG)_9$	179 - 187	53
Sop-42	R	TTGAGTCCAACATGAATGGC			53
Sop-248	F	TCCCGGAAATCTCATTCAAAGG	$(GTT)_{13}$	270-290	53
Sop - 248	R	ACTCAAGGAGTTTAGGTAGCG			53
Sop-802	F	ACAAAGCCTCATACAGAGC	$(GTT)_{10}$	303-309	53
Sop-802	R	GAATGACCAAGGTATCGCC*			53
Sop-807	F	AGTGTACCTTGACGATTGTG	$(AT)_9$	319-325	53
Sop-807	R	TCAGTTGGTGAACTATCAAC*			53

Note : Ta = annealing temperature used in PCR, * PIG tail (GTTTCTT) added to the 5' end of each reverse primer

3.3.7 Phenotyping of PN F_2 population

The new F_2 population was phenotyped for leaf and branch number traits in the first week of February 2018, after a little over a year of growth. To allow plants to develop further and determine the best methods, other branch traits were phenotyped in January 2019 (Table 3.4). Phenotyping was performed using digital calipers as described in Chapter 2 and by processing photographs of clay impressions of branches for node and branch angles in Fiji (ImageJ) v.1.5.1 (Schindelin et al., 2012). The phenotypes measured included leaflet number, rachis length, leaflet length and width, branch width, node angle, branch angle, internode length, number of primary (X1Br), secondary (X2Br),

tertiary (X3Br) and total number of branches (TotalBr). All leaf and branch size measurements were made in mm units.

Four leaves and three leaflets per leaf were measured per plant; measurements of each trait for each individual were averaged for analysis. Leaf traits, internode length and branch width were performed as described for the Lincoln F_2 in Chapter 2. Leaflet length and width were taken from the longest and widest parts of the leaf, respectively. Rachis length was measured from the base of the petiole to the tip of the rachis. Branches were counted including primary branching (branches developed from the main stem), secondary branching, which includes branches growing from the primary branches and tertiary branching, which includes branching from the secondary branches. These were also summed to make a total branch number for each plant. Internode length was measured from the top of one node to the top of the next node. Branch width was measured at the fifth internode from the base for each plant in the centre of the internode. For node angle and branch angle impressions of the branches were made at the nodes using a modelling compound and these impressions were photographed. Nodes were chosen starting from the fifth node from the base, with each consecutive node measured where possible, and included branches when present. The angles of the nodes and branches were calculated from the impression photos using Fiji v. 1.51 (Schindelin et al., 2012). Five measurements for internode length, node angle and branch angle, where possible, were made and averaged for analysis.

Trait analysis of the PN F_2 population was performed using R v. 3.4.2 (R Development Core Team, 2016). Histograms and boxplots were generated using the graphics v. 3.4.2 package. Correlation coefficients were performed using the stats v. 3.4.2 package. Descriptive statistics were calculated using the R package psyche v. 1.8.3.3 (Revelle, 2016). The package corrplot v. 0.84 (Wei & Simko, 2016) was used to generate

the heat map and the PCA visualised with ggbiplot v. 0.55 (Vincent, 2011). Secondary and tertiary branch traits were excluded from the PCA calculations due to missing data as some plants did not have branching beyond the primary branches.

Table 3.4. Traits phenotyped in the PN F_2 indicating dates measuring was performed

Trait	Date of phenotyping
Leaflet Number	1-7 February 2018
Rachis Length	1-7 February 2018
Leaflet Length	1-7 February 2018
Leaflet Width	1-7 February 2018
Internode Length	14-18 January 2019
Branch Width	14-18 January 2019
Number of Primary Branches (X1Br)	1-7 February 2018
Number of Secondary Branches (X2Br)	1-7 February 2018
Number of Tertiary Branches (X3Br)	1-7 February 2018
Total Branch Number	1-7 February 2018
Node Angle	14-18 January 2019
Branch Angle	14-18 January 2019

3.3.8 Genotyping PN F_2 for *RMSI*

To follow up on the associations found between *RMSI* and divaricate phenotypes in the Lincoln F_2 population, the PN F_2 population was genotyped at *RMSI*. Leaf tissue was collected from each of the PN F_2 individuals and dried in silica gel for later DNA extraction. DNA extractions were performed with a modified cetyltrimethylammonium bromine (CTAB) protocol (Doyle & Doyle, 1987). All of the 69 PN F_2 plants were genotyped for the candidate gene marker, *RMSI,* as described above. Analyses for *RMSI* were performed in R v. 3.4.2 (R Development Core Team, 2016) using the methods

described above for the Lincoln F_2. Leaflet number and rachis length were indicated to violate the assumption of homogeneity of variance by Levenes test and so may not be suitable for the Tukey HSD post hoc test. Therefore the Games-Howell (Games & Howell, 1976) post hoc test was also performed for these traits using the function *oneway* with posthoc = 'games-howell' specified in the userfriendlyscience v. 0.7.2 package.

3.3.9 Synthetic strigolactone experiments on *S. prostrata* and *S. tetraptera* seedlings

To test the potential involvement of the strigolactone hormone in the divaricate growth form, synthetic strigolactone, GR24 (Chiralix B.V., 2018), was applied to *S. prostrata* and *S. tetraptera* seedlings grown under uniform conditions in a growth chamber at Massey University.

Fifteen seeds of *S. prostrata* were put through the germination scheme described above; however, germination rates were quite low. As a result, only six *S. prostrata* seedlings were initially grown in a seed-raising mix (Oderings) in four inch pots under 16 hours of light at ~ 21°C as a pilot study. For the pilot, 50 mL of 100 nM GR24 (Chiralix B.V., 2018) in water was applied to the soil or as 1 µl directly on each node axis of the seedling starting approximately 30 days after planting in soil. As germination rates were very low, only four seedlings were treated with GR24, two with water application and two with node treatment, and two seedlings were treated with water as control plants (Table 3.5).

Treatments were completed every second day for two months before plants were phenotyped. Measurements taken for each plant were: number of branches, branch length, stem width, leaf number, leaflet number, rachis length, leaflet length and leaflet width. From such a small sample size, results on the effect of GR24 could not be

conclusively determined but the application of GR24 by watering in was determined to be the most efficient application as direct node application was too time intensive.

Table 3.5. Methods of GR24 treatment trialled in the pilot with *S. prostrata* seedlings.

Treatment	Method of Treatment	Number of seedlings
Control	50 ml water added to soil	2
Node treatment	1 µl added to each node axis	2
SL water treatment	50 ml of 100 nM GR24 in water added to soil	2

To scale up, fifty seed of each species were put through the germination scheme. However, germination rates were (again) low for *S. prostrata* and the seedlings grew slower than most *S. tetraptera* seedlings. The low germination of *S. prostrata* seeds occurred due to fungal contamination (despite the severe germination treatment). As a result, eight *S. prostrata* and eight *S. tetraptera* seedlings, grown as described above, were used in another small-scale experiment. *S. tetraptera* seedlings were also treated to determine possible treatment effects that GR24 may have on plants that should not be deficient in strigolactone. Only eight plants of *S. tetraptera* were included as these were the only seedlings near a similar age to the *S. prostrata* seedlings. In this experiment, four seedlings of each species were treated with GR24 and four were treated with water as a control; all treatments were applied via 40 mL every two days. Treatments were started when buds were beginning to develop and results were measured after two months of treatment. Phenotype measurements were the same as those for the initial trial.

Following the small-scale experiment, a new larger experiment was initiated on 9 February 2018 with newly collected (December 2017) *S. prostrata* seeds as the seeds previously used were collected in 2016 and the age of seeds may have contributed to the

low germination rates. The previous germination scheme was developed for the F_2 seedlings, rather than *S. prostrata* so the time of treatment of H_2SO_4 was reduced for *S. prostrata* seeds. Sixty seeds from *S. prostrata* were treated in concentrated H_2SO_4 for 20 minutes followed by a 2 hour wash in water and grown as described above for the pilot. Only nineteen plants of a similar age germinated as healthy plants; ten were treated with GR24 and nine were used as water controls. Treatment was started on the 1st of June 2018 when plants were about one month old and phenotyping was performed on the 10th August 2018 about two months after treatment began. GR24 was applied by adding 40ml of 100nM solution in water to each pot and was applied every second day until plants were measured. The phenotypes measured were internode length, stem width, branch number, branch length, rachis length, leaflet number, leaflet length, and leaflet width as described for these traits in Chapter 2. Buds were considered branching when the first leaves could be seen forming. For all trials, boxplots were performed in the graphics v.3.4.2 package and *t*-tests were performed using the function *t.test* in the basic stats package in R v. 3.4.2 (R Development Core Team, 2016).

3.4 Results

3.4.1 Isolation, sequencing, and characterisation of *RMS* genes and *ZIG* from *S. prostrata* and *S. tetraptera*

RMSI characterisation

The partial sequences obtained from *RMSI* showed a clear size difference (Figure 3.5) between the *S. prostrata* and *S. tetraptera* alleles. Alignments between the *Sophora* sequences and the *Pisum RMSI* ortholog revealed that the size difference is mainly due to a larger indel (686 bp) in the first intron. The total length of the partially sequenced *RMSI* from *S. prostrata* is 2,813 bp whereas for *S. tetraptera* the size of the *RMSI* partial sequence is 3, 556 bp where the size difference includes the intron 1 indel and other smaller indels spread throughout the sequence (Figure 3.6). BLAST searches using the new *Sophora RMSI* sequences returned results from other Fabaceae taxa including *Glycine max, Pisum sativa* and *Sesbania cannabina CCD8 (RMSI)* sequences. Based on alignments with other legume *RMSI* orthologs, the inferred amino acid sequences from *S. prostrata* and *S. tetraptera* differ by four replacements (Table 3.6). For three of those positions, *S. tetraptera* shares the same amino acid possessed by the *Pisum RMSI*

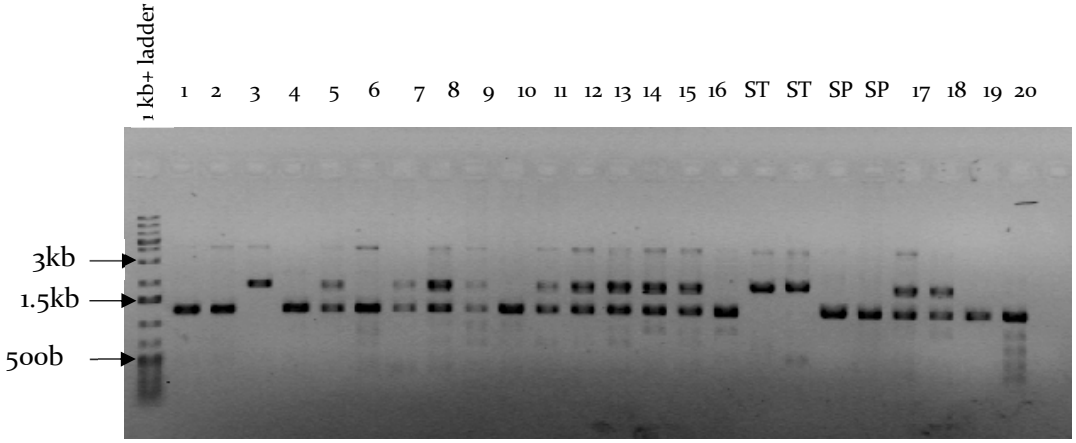


Figure 3.5. Example of *RMSI* marker genotyping. Numbers 1 to 20 indicate samples from the Lincoln F_2 population. Two *S. tetraptera* samples (ST) and two *S. prostrata* samples (SP) display the size difference of *RMSI* between the two species. Heterozygotes were scored when two bands within an individual were present. Homozygotes were scored when only one band was present, with band size representing the allele. The black arrows indicate the 1.5kb and 3kb bands on the Generuler 1kb Plus DNA ladder from Thermo Scientific (Thermo Fisher Scientific).

sequence and at the fourth position, *P. sativum* differs from both *Sophora* sequences (Table 3.6).

Table 3.6. Amino acid changes in *RMS1* sequence for *S. prostrata* and *S. tetraptera*.

	Pisum sativum	*S.tetraptera*	*S. prostrata*	*PROVEAN score*
Size of nucleotide sequence	3,123bp	3,556bp	2,813bp	
Size of amino acid sequence	561	410	410	
Amino acid 273	Glu (E)	Glu (E)	Gln (Q)	-1.172
Amino acid 400	Leu (L)	Leu (L)	Met (M)	-1.683
Amino acid 416	Lys (K)	Arg (R)	Ser (S)	-3.205
Amino acid 427	Ser (S)	Ser (S)	Gly (G)	-1.75

Amino acid positions are in reference to the *Pisum sativum RMS1* sequence. PROVEAN score refers to the amino acid replacement tests for deleterious mutations.

PROVEAN analysis of the inferred *Sophora RMS1* translated sequences was carried out to predict if replacements in *S. prostrata* could have a deleterious effect on the protein. In the analysis, 154 sequences, and the top 30 clusters, were identified and used by PROVEAN. The software revealed one of the amino acid replacements, at position 416, from an arginine to a serine (Figure 3.6) was predicted to have a deleterious effect (Table 3.6) on the protein with a delta score of -3.205. All the other sites were predicted to be neutral. To follow up on this, CCD8 protein sequences for species from diverse plant families including representatives in the Brassicaceae, Fabaceae, Cucurbitaceae, Apiaceae, Euphorbiaceae, Solanaceae, Poaceae, Funariaceae, and Selaginellaceae were downloaded from Genbank (see section 3.3.2) and alignments made in Geneious v. 9.1.8. At the position of the predicted deleterious polymorphism, all sequences, except *Pisum sativum* and *S. prostrata,* had an arginine. *P. sativum* had lysine while *S. prostrata* had serine at that position.

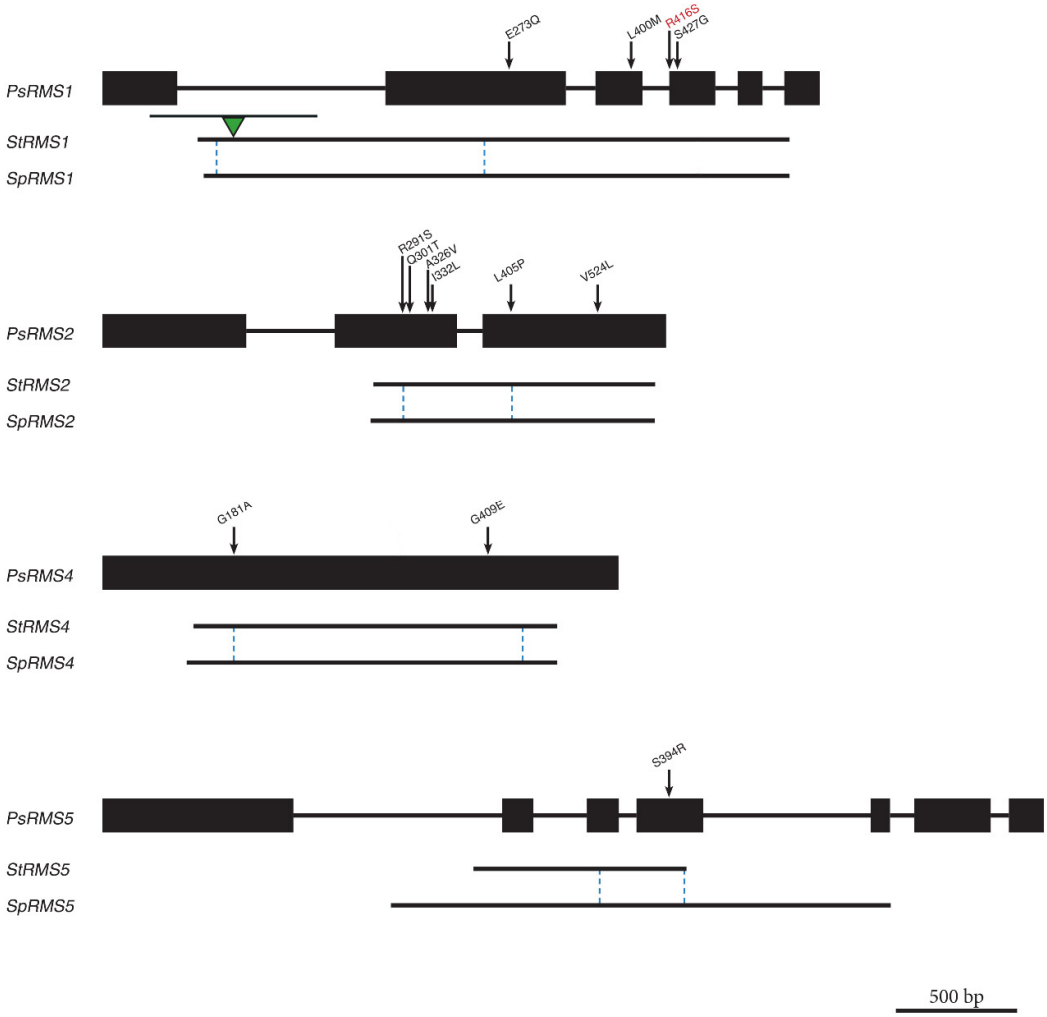


Figure 3.6. Diagrams displaying the structure of the RMS candidate genes. *Pisum sativum* (Ps) is used as a reference with exons indicated by solid rectangles and introns indicated by the solid lines. *S. prostrata* (Sp) and *S. tetraptera* (St) partial fragments obtained for each gene are indicated below *P. sativum*. Amino acid differences between the *Sophora* sequences are indicated with arrows labelled with the amino acid replacement and the position relative to the *Pisum* sequence. The amino acid replacement in red in *RMS1* is predicted as potentially deleterious. The indel in *RMS1* for *S. tetraptera* is indicated with the green triangle and the line above it indicates the size of the insertion. The blue dashed lines in *RMS2, RMS4* and *RMS5* indicate the positions of the species-specific primers, located at a SNP, used to genotype these genes. The dashed lines in *RMS1* indicate the positions of the primer pair used to genotype this gene by product size. A scale bar is indicated in the bottom right of the figure.

RMS2 characterisation

The partial *RMS2* sequences obtained were 1,193 bp for *S. tetraptera* and 1,179 bp for *S. prostrata.* BLAST results for the *Sophora* sequences returned *Pisum RMS2* and a *Prunus* F box protein as well as several predicted RMS2 proteins. The *Sophora* sequences

were aligned with *Pisum RMS2*. The alignment identified six sites with amino acid differences between the two *Sophora* species (Figure 3.6 & Table 3.7). PROVEAN analysis of *RMS2* used 161 sequences and predicted the effects of all amino acid replacements as neutral.

Table 3.7. Amino acid differences between *S. prostrata* and *S. tetraptera* for *RMS2*.

Fragment position	*Pisum sativum*	*S. tetraptera*	*S. prostrata*	*PROVEAN score*
Size of nucleotide sequence (bp)	2,558	1,193	1,179	
Size of amino acid fragment	617	367	364	
Amino acid 291	Ser (S)	Arg (R)	Ser (S)	1.017
Amino acid 301	Gln (Q)	Gln (Q)	Thr (T)	-1.363
Amino acid 326	Pro (P)	Ala (A)	Val (V)	-1.232
Amino acid 332	Ile (I)	Ile (I)	Leu (L)	-0.724
Amino acid 405	Pro (P)	Leu (L)	Pro (P)	1.403
Amino acid 524	Leu (L)	Val (V)	Leu (L)	2.729

Amino acid positions are in reference to the *Pisum sativum RMS2* sequence. PROVEAN score refers to the amino acid replacement tests for deleterious mutations.

RMS4 characterisation

The partial *RMS4* sequences obtained from *Sophora* were 1,502 bp for *S. tetraptera* and 1,524 bp in *S. prostrata*. BLAST results showed hits from two legumes: *Glycine max MAX2 (RMS4)* and *Sesbania cannabina MAX*. Two amino acid differences were observed between *S. tetraptera* and *S. prostrata* for the *RMS4* fragment (Figure 3.6 & Table 3.8). *PROVEAN* software identified 139 sequences for analysis. The two amino acid replacements were predicted to have neutral effects.

RMS5 characterisation

The partial *RMS5* sequence from *S. tetraptera* and *S. prostrata* were quite different in size, at 878 bp and 2,060 bp, respectively, as *S. tetraptera* sequence could only be obtained from one primer pair, Ex3AF and Ex4R (see Table 3.1). The sequences were aligned in Geneious with the *Pisum sativum RMS5* sequence and BLAST searches returned results from *Glycine max CCD7(RMS5)*. More dissimilar results included *Lotus CCD7-like,* and *Trifolium CCD7.* Only one amino acid difference was observed between the sequences of the two *Sophora* species (Figure 3.6 & Table 3.9); however only a small portion of the gene was sequenced, especially for *S. tetraptera.* This was predicted as a neutral replacement by PROVEAN analysis using 41 supporting sequences.

Table 3.8. Amino acid differences between *S. prostrata* and *S. tetraptera* for *RMS4*.

Fragment position	*Pisum sativum*	*S. tetraptera*	*S. prostrata*	*PROVEAN score*
Size of nucleotide fragment (bp)	2,127	1,502	1,524	
Size of amino acid fragment	709	460	460	
Amino acid 181	Ala (A)	Gly (G)	Ala (A)	-1.966
Amino acid 572	Glu (E)	Gly (G)	Glu (E)	-0.621

Amino acid positions are in reference to the *Pisum sativum RMS4* sequence. PROVEAN score refers to the amino acid replacement tests for deleterious mutations.

Table 3.9. Amino acid differences between *Sophora* species for *RMS5*.

Fragment position	*Pisum*	*S. tetraptera*	*S. prostrata*	*PROVEAN score*
Size of nucleotide fragment (bp)	3,880	878	2060	
Size of amino acid fragment	620	88	186	
Amino acid 394	Lys (K)	Ser (S)	Arg (R)	-2.341

Amino acid positions are in reference to the *Pisum sativum RMS5* sequence. PROVEAN score refers to the amino acid replacement tests for deleterious mutations.

***ZIG* characterisation**

The *ZIG* sequences obtained were 1,483 bp in *S. prostrata* and 1,324 bp in *S. tetraptera*. A BLAST search returned results from *Medicago* and *Glycine* v-SNARE 13. *ZIG* is a v SNARE VTI11 protein involved in a complex of four v SNARE proteins (Hashiguchi et al., 2010). Determining the intron/exon boundaries of *ZIG* from the *Sophora* sequences was not successful and consequently the PROVEAN analysis could not be run for the *ZIG* SNPs.

3.4.2 Genotyping the *RMS* genes in the Lincoln F_2 population

All candidate gene markers were successfully genotyped in the Lincoln F_2 population resulting in no missing data. The genotypes in the F_2 for most markers displayed a segregation ratio close to the expected ratio of 1:2:1 (Table 3.10). The exception to this was *ZIG* which had a segregation ratio of 0.3:1.2:1. Statistically significant associations were observed between phenotypes and genotypes for some traits in all candidate genes (Table 3.11 & Appendix 3.1). The most frequent associations were observed for leaf traits. *RMS1* displayed statistically significant associations for six of nine traits (Table 3.11). Following a Bonferroni correction for multiple tests, *RMS1, RMS2* and *ZIG* were the only candidate genes with statistically significant associations (Appendix 3.2). *RMS1,* after correction, displayed statistically significant associations for leaflet length, width and rachis length, *RMS2* indicated leaflet number and leaflet width still significant and *ZIG* displayed a statistically significant association for leaflet width after correction. Levene's test for homogeneity of variance among groups did not indicate statistically significant differences in variance between genotype and traits (Appendix 3.3) indicating TukeysHSD post hoc test could be performed on the data. TukeysHSD indicated for *RMS1* that leaflet length, leaflet width, rachis length, branch width and node angle were statistically significant between the parental homozygote

alleles (Appendix 3.4). Leaflet number and leaflet width for *RMS2* indicated statistically significant differences between parental homozygotes and leaflet width was also indicated statistically significant for *ZIG*.

Table 3.10. The results from genotyping the *RMS* and *ZIG* genes in the Lincoln F_2 and associations with phenotype traits.

Gene	Genotype ratio	Significant phenotypic traits
RMS1	1: 2.2 : 0.6	LL, LW, RL, IL, BW, BN, NA
RMS2	0.5: 1.8 : 1	LN, RL, LW
RMS4	0.5: 1.8 : 1	LL, NA
RMS5	0.9: 1.5: 1	LL, LW
ZIG	0.3:1.2:1	LW, BW

Genotype ratio structured as: *S. prostrata* homozygote: Heterozygote: *S. tetraptera* homozygote. Trait abbreviations are: LL (leaflet length), LW (leaflet width), RL (rachis length), IL (internode length), BW (branch width), BN (branch number), NA (node angle).

Plants with the homozygote *S. prostrata RMS1* allele had smaller branches, longer internodes, smaller leaves and smaller node angles consistent with the phenotype of the *S. prostrata* parent and vice versa for F_2 with the *S. tetraptera* homozygote allele (Appendix 3.5). When the analysis was split to consider cross direction, the AB population (with maternal parent *S. tetraptera*) only had statistically significant differences for leaflet length, width and rachis length with *RMS1* and only leaflet length statistically significant after Bonferroni correction (Appendix 3.6). TukeysHSD test indicated statistically significant difference between the parental homozygotes for leaflet length (Appendix 3.7). The CD population (maternal parent *S. prostrata*) showed statistically significant differences for *RMS1* in leaflet length, rachis length, branch width, and internode length (Appendix 3.8) with rachis length statistically significant after Bonferroni correction. Levene's test indicated violation of the assumption for homogeneity of variance in the CD *RMS1* for rachis length (Appendix 3.9) indicating the

Games-Howell post hoc test may be more appropriate. The post hoc tests indicated a statistically significant difference between parental homozygotes for leaflet length, rachis length, and leaflet number with *RMS1* in the CD population (Appendix 3.10).

Table 3.11. Table of P-values from ANOVA tests in the *Sophora* F_2 for the candidate gene markers.

	RMS1 *Lincoln*	*RMS1* *PN F_2*	*RMS2* *Lincoln*	*RMS4* *Lincoln*	*RMS5* *Lincoln*	*ZIG* *Lincoln*
Leaflet Number	0.401	0.00955**	0.0015**[B]	0.121	0.725	0.114
Leaflet Length	0.000307**[B]	$1.58e^{-8}$**[B]	0.0184*	0.00583**	0.0826	0.0115*
Leaflet width	0.000841**[B]	$5.51e^{-8}$**[B]	0.00478**[B]	0.212	0.0144**	0.00548*[B]
Rachis Length	0.000158**[B]	$7.88e^{-8}$**[B]	0.00959*	0.0238*	0.582	0.0168*
Internode Length	0.0272*	0.0344*	0.77	0.847	0.389	0.321
Branch Width	0.0145**	0.123	0.986	0.605	0.6	0.156
Branch Number	0.0828	-	0.214	0.451	0.988	0.232
Node Angle	0.0121**	0.00966**	0.035*	0.0646	0.283	0.395
Branch Angle	0.303	0.485	0.393	0.715	0.0814	0.987
Primary Branches	-	0.0654	-	-	-	-
Secondary Branches	-	0.00818**	-	-	-	-
Tertiary Branches	-	0.0853	-	-	-	-
Total Branches	-	0.00348**[B]	-	-	-	-

[*] indicates significance p>0.05, - indicates measurements that were not performed in that population, ** indicates significant difference between parent homozygote alleles after post-hoc tests. [B] indicates significance after Bonferroni correction.

3.4.3 New PN F_2 phenotyping

Nine traits were measured for the 69 individuals in the PN F_2 population with branch number also divided into primary, secondary, tertiary and total branch number. Descriptive statistics for each trait are provided in Table 3.12. Histogram distributions (Figure 3.7 & Appendix 3.16) and the skewness values (Table 3.12) indicate that most

traits display a right (positive) skew as values above 0.5 are considered to display moderate skew (Bulmer, 1979). Rachis length, leaflet number, leaflet length, secondary branching, tertiary branching and branch width all display significant skew indicated by the skew value above 1. Primary branching and node angle are the only traits to show an approximately normal distribution as indicated by the value close to 0. Kurtosis measures the tails of the distributions where values above 1 indicate most traits display a tail in their distributions.

In this population, there were several between trait correlations (Figure 3.8). The leaf traits were all positively correlated and also positively correlated with node angle and branch width. The leaf traits negatively correlated with the branch number categories, except tertiary branching, indicating that leaf size, node angle and branch width increase with decreasing branch number and vice versa. Branch angle was only correlated with node angle. These results are generally consistent with the results in the Lincoln F_2 population (described in Chapter 2); however, some differences were observed between populations which is not unexpected when different methods are used for phenotyping traits and with the differences in plant age and environment.

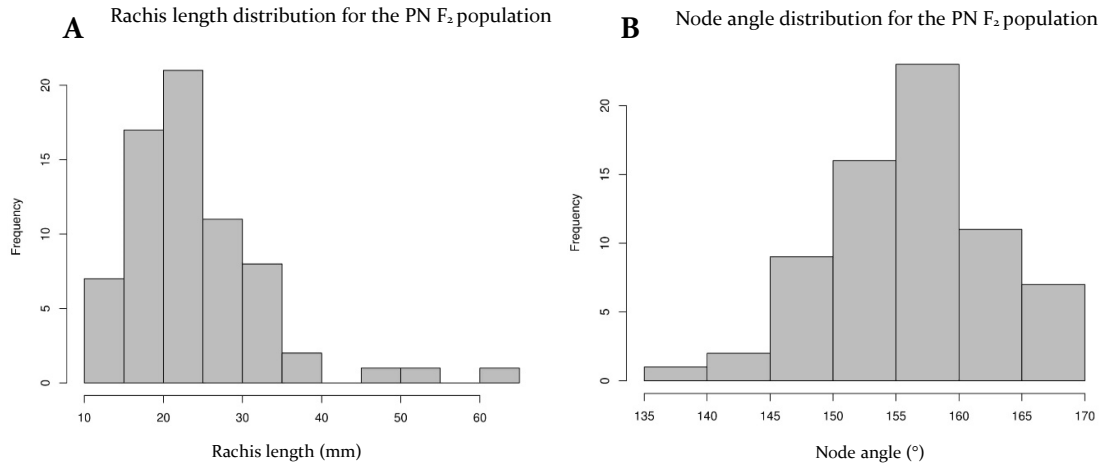


Figure 3.7. Examples of histogram distributions. Rachis length (A) displaying a positive skew and node angle (B) displaying a fairly normal distribution.

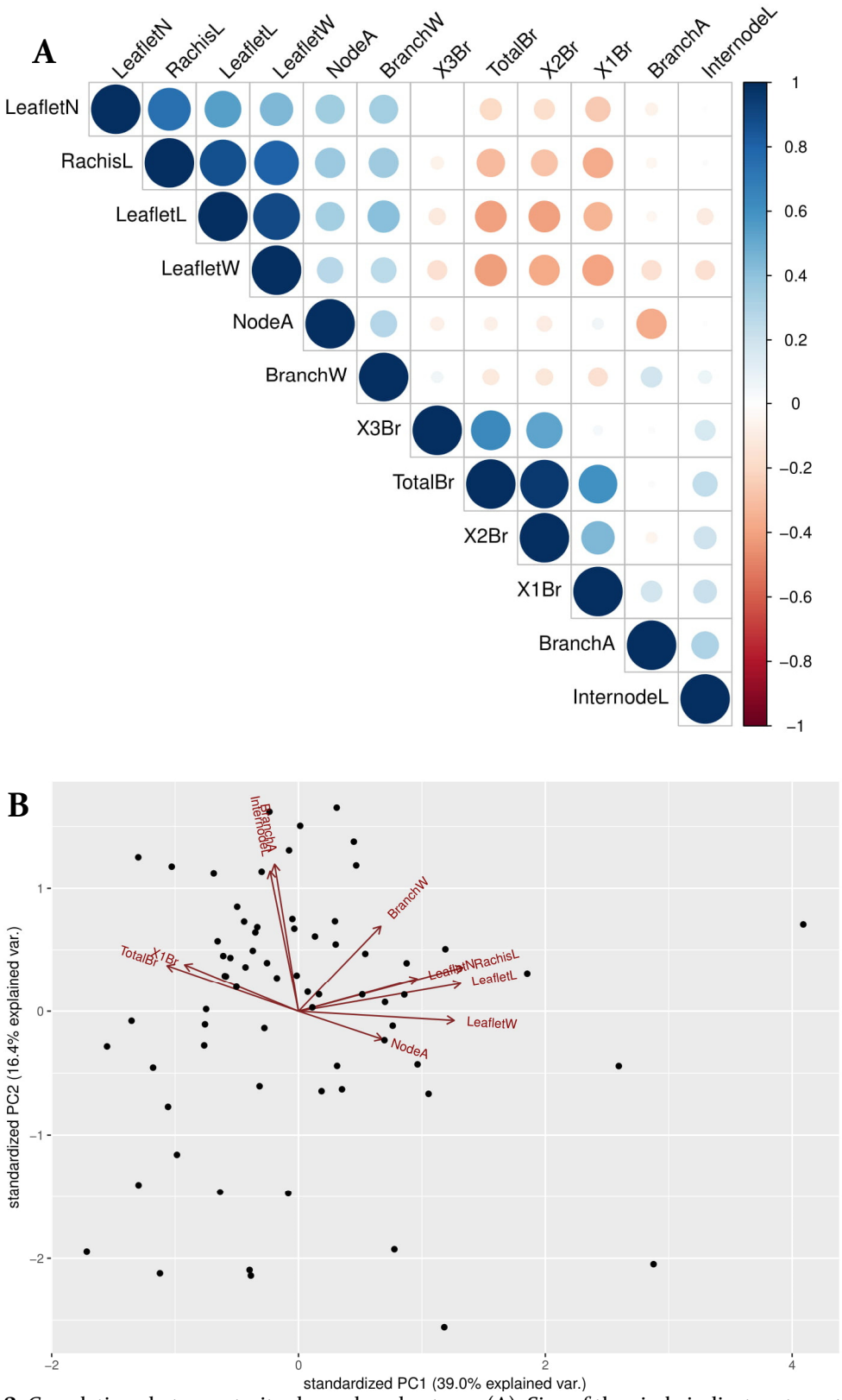


Figure 3.8. Correlations between traits shown by a heatmap (A). Size of the circle indicates strength of correlation and colour indicates the direction of correlation where blue is a positive correlation and red is negative correlation. X1Br-primary branch number, X2BR – secondary branch number. PCA (B) showing the two principal components explaining the largest variation in the data.

Table 3.12. Descriptive statistics for the PN F_2 population.

Trait	n	mean	sd	median	min	max	range	skew	kurtosis	se
LeafletN	69	9.55	2.18	9.25	5.25	18.25	13	1.1	2.46	0.26
RachisL	69	24.19	8.89	22.35	12.03	62.51	50.48	1.88	4.81	1.07
LeafletL	69	6.28	1.7	6.07	3.38	14.3	10.92	1.7	5.71	0.21
LeafletW	69	3.62	0.68	3.49	2.38	6.04	3.66	0.95	1.86	0.08
TotalBr	69	24.9	15.31	21	1	71	70	0.95	0.4	1.84
PrimaryBr	69	11.2	4.34	11	1	26	25	0.16	1.27	0.52
SecondaryBr	69	11.7	10.62	9	0	52	52	1.46	2.37	1.28
TertiaryBr	69	2	3.97	0	0	17	17	2.28	4.66	0.48
NodeA	69	156.09	6.71	156.44	139.11	169.55	30.44	-0.05	-0.36	0.81
BranchA	69	85.91	13.28	85.09	60.56	123.88	63.32	0.32	-0.05	1.6
BranchW	69	3.74	1.46	3.35	1.74	9.9	8.16	1.74	4.04	0.18
InternodeL	69	17.05	6.1	16.19	6.63	38.12	31.49	0.63	0.64	0.73

3.4.4 Genotyping *RMSI* in the PN F_2 population

The PN F_2 population was also genotyped for *RMSI* to check if the associations detected in the Lincoln F_2 population were consistent in a population developed from a single F_1 and when grown in a more controlled environment. This population consisted of 69 F_2, grown from seed collected from the 10892-9 individual, and displayed a ratio of 1: 2.4: 0.7 (*S. prostrata* homozygote allele: Heterozygote: *S. tetraptera* homozygote allele). These plants showed statistically significant differences in genotype for leaflet number, rachis length, leaflet length, leaflet width, node angle, internode length, number of secondary branches and total branch number (Tables 3.11, Appendix 3.2 and Appendix 3.11). Levene's test for homogeneity of variances indicated leaflet number and rachis length violate the assumption (Appendix 3.12) indicating TukeysHSD post hoc test may not be suitable and therefore the Games-Howell post hoc test was also performed

for these traits. Post hoc testing indicated leaflet length, leaflet width, rachis length, node angle and secondary branching are statistically significant between parental homozygote alleles (Appendix 3.13 & 3.14).

3.4.5 Synthetic strigolactone application to seedlings

To test the associations seen between *RMSI* and the traits that displayed significant differences between parental alleles, synthetic strigolactone, GR24, was supplied to *S. prostrata* seedlings (Figure 3.9). *RMSI* is part of the pathway involved in the biosynthesis of strigolactones where *rms1* mutants produce less or no strigolactone (Beveridge et al., 1997; Sorefan et al., 2003; Booker et al., 2004; Koltai, LekKala, et al., 2010). The association of the *RMSI* marker with divaricate traits in this work may indicate that a lack of strigolactone production or perception contributes to the divaricate growth form in *S. prostrata*. Application of synthetic strigolactone to *S. prostrata* seedlings may result in changes to these phenotypes if a lack of strigolactone is responsible and could indicate whether or not strigolactone biosynthesis genes are involved. A trial was conducted with nineteen plants, as this was the number of healthy *S. prostrata* seedlings of similar age that survived the germination treatment. The seedlings were difficult to grow from seed, often affected by fungus despite the treatment in H_2SO_4. Preliminary trials revealed the most suitable and practical way to deliver the synthetic strigolactone was through a GR24 solution watered into the soil. None of the branch traits measured were significantly different between the control and treated plants (Figure 3.10, Table 3.13). Of the leaf phenotypes, leaflet length and width did show significant differences between the control and strigolactone-treated plants (Figure 3.10, Table 3.13), even with the very small number of samples. A smaller trial involving eight *S. tetraptera* seedlings was also performed and only leaflet width was indicated as significantly different between treatments for *S. tetraptera* (Appendix 3.15) however, the

sample size is very small. *S. prostrata* did not show any significant differences in this eight plant study.

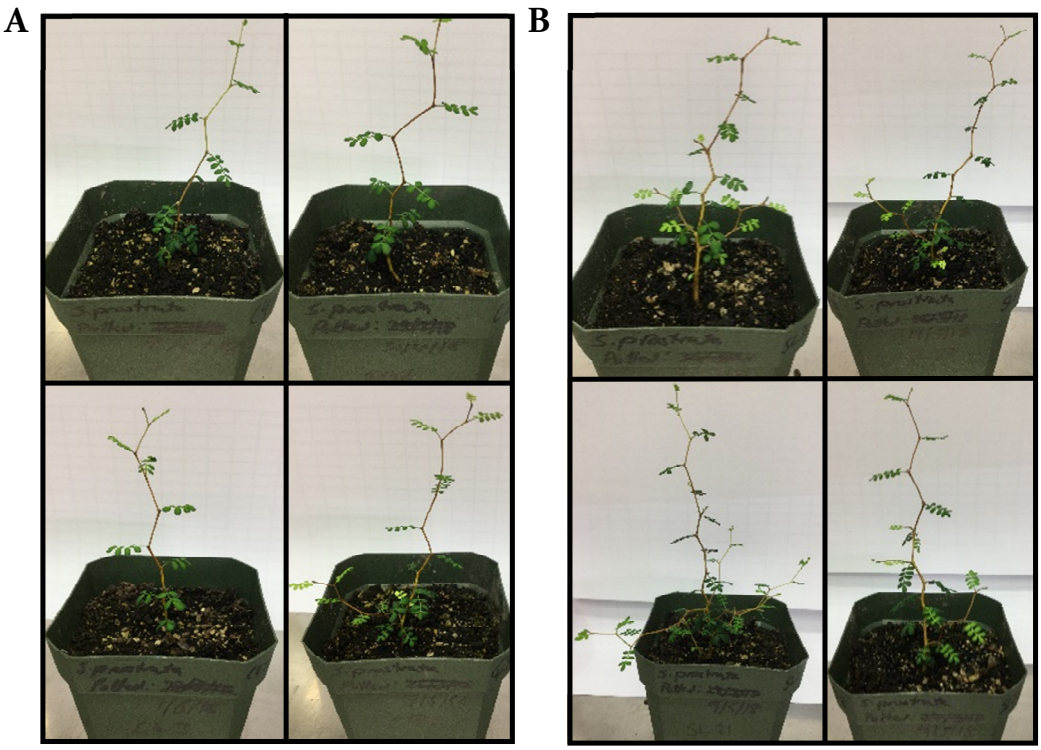


Figure 3.9. Example *S. prostrata* seedlings treated with water control (A) and the synthetic strigolactone, GR24 (B).

Table 3.13. *t*-test results for application of GR24.

Trait	Control mean	GR24 mean	t	df	p-value
Leaflet number	7.7	6.8	1.9698	16.867	0.06551
Rachis length	14.4	13.7	0.81479	16.869	0.4265
Leaflet length	4.4	5.0	-4.5506	16.2	0.0003178*
Leaflet width	2.9	3.3	-4.5162	16.239	0.0003395*
Internode length	6.8	7.0	-0.24918	16.494	0.8063
Stem width	1.1	1.1	0.30859	12.116	0.7629
Branch number	2.1	2.6	-0.64622	16.752	0.5269
Branch length	38.0	27.6	1.208	11.365	0.2516

* indicates significant difference

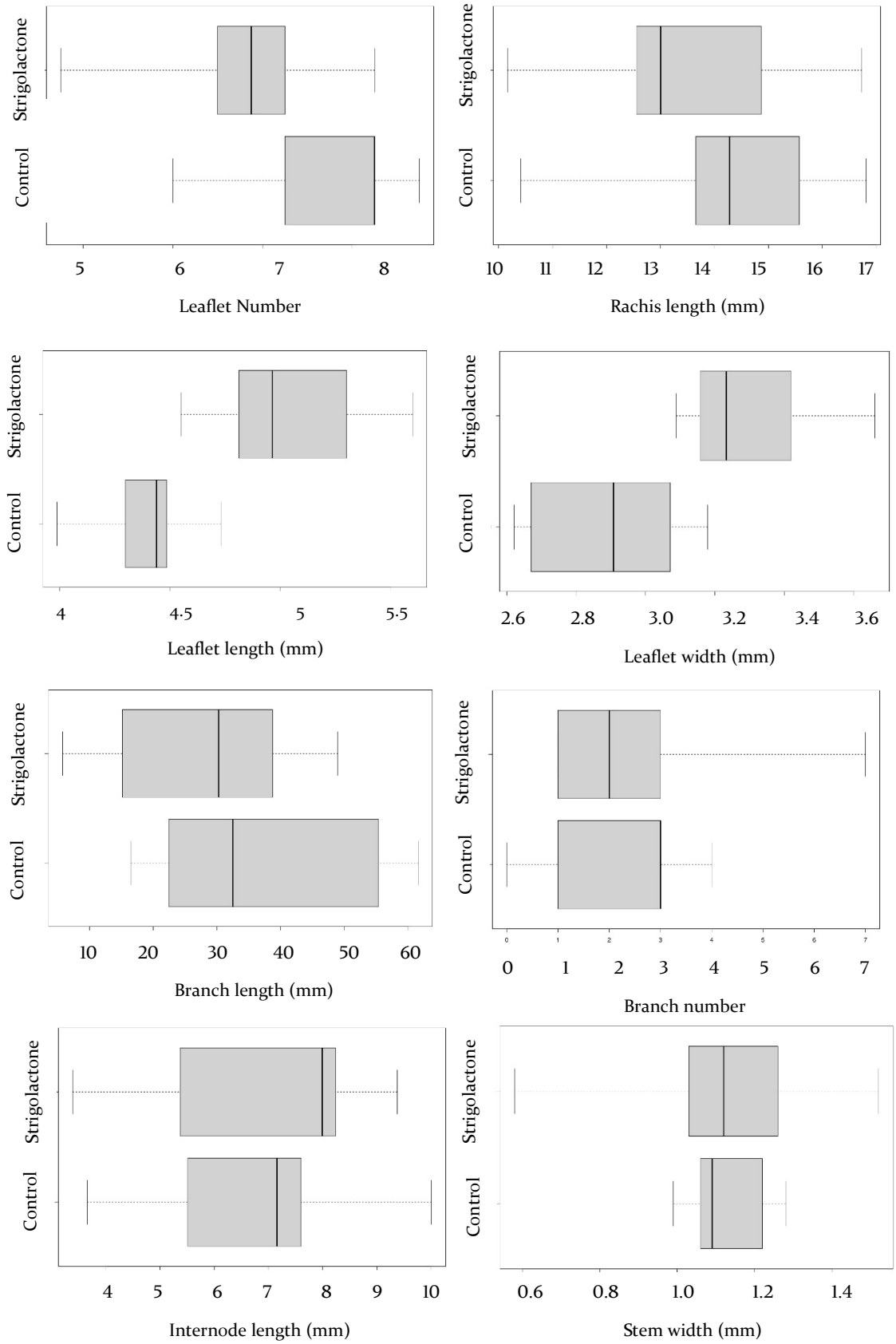


Figure 3.10. Example boxplots displaying comparisons between the GR24 treated *S. prostrata* seedlings (SL) and the control seedlings (C) for the traits. Control plants represented in the top box and whisker chart and treated plants by the bottom chart.

3.5 Discussion

The strigolactone hormones are an important component in the control of axillary shoot branching in seed plants. The shoot branching pathway is highly conserved across seed plants with gene orthologs found in diverse plant groups including *Arabidopsis, Pisum, Populus, Petunia, Maize* (Johnson et al., 2006; Czarnecki et al., 2014), *Pinus* and *Picea* (Delaux et al., 2012). Strigolactone is produced in the Charales, the family thought to be the sister to land plants, but not other chrysophyte algae suggesting an origin of around 750-1200mya from which the control of shoot branching then developed (Waters et al., 2017). The genes in this pathway were initially identified and characterised from mutants in *Pisum sativum* and *Arabidopsis thaliana*. These mutants displayed an increased branching phenotype but often other phenotypes were present including smaller leaves, thinner stems and shorter plant height (Beveridge et al., 1997; Morris et al., 2001; Sorefan et al., 2003; Booker et al., 2004; Johnson et al., 2006; de Saint Germain et al., 2016). These phenotypes are similar to many of the phenotypes commonly described for the divaricate growth form, such as increased branching with smaller leaves. As this pathway is conserved in land plants and mutant phenotypes are similar to those of the divaricate growth form, these genes constitute good candidates for the genetic transition to the divaricating growth form. Orthologs from four of the genes from the strigolactone pathway (Figure 3.1), *RMS1, RMS2, RMS4* and *RMS5,* were developed into species-specific molecular markers for *Sophora.* Two of these genes, *RMS1* and *RMS5,* are involved in the biosynthesis of the strigolactone hormone where knock-out mutants of these genes in *Arabidopsis thaliana* and *Pisum sativum* lack strigolactone (Foo et al., 2001; Foo et al., 2005; Flematti et al., 2016). *RMS4* is involved in perception and signalling of the strigolactone signal where RMS4 binds to RMS3, following perception and binding the strigolactone D ring by RMS3. This forms the SCF complex that degrades downstream targets, such as the SMX gene products

(Woo et al., 2001; Stirnberg et al., 2007; Jiang et al., 2013; Soundappan et al., 2015). *RMS2* is involved in a feedback loop associated with auxin biosynthesis where RMS2 binds to auxin and upregulates biosynthesis of strigolactone in the roots (Ligerot et al., 2017). Here, I investigated polymorphisms in several of the *RMS* genes and another candidate gene from outside this pathway, *ZIG*.

3.5.1 Molecular variation in candidate genes

The partial sequences obtained from the candidate genes for *S. prostrata* and *S. tetraptera* generally revealed few polymorphisms between *Sophora* species with genes displaying only one to six amino acid replacements (Tables 3.14 – 3.17); *RMS2* displayed the most amino acid replacements with six in total. In contrast, the number of amino acid replacements between *Pisum sativum* and the *Sophora* partial sequences was much higher ~ 30 – 50 replacements (data not shown). A study of *CCD8* (*RMS1)* among seven monocots and eight dicots indicated variation in gene length, mainly occurring from the number and size of introns/exons ranging from 4-7 exons. The sequence similarity compared to *Zea mays* ranged from 52% - 76% with monocots the most similar indicating *CCD8* displays considerable natural variation among plant groups (Batra et al., 2019).

Among the key aims of the work in this chapter was to identify interspecific variation in candidate genes and develop allele-specific markers for each gene of interest. Between indels and SNPs, a PCR-based genotyping scheme was developed for each candidate gene. While they were useful here for single locus association tests, their application in the Lincoln F_2 population means that those genes can potentially be placed on linkage groups (Chapter 4) and used in QTL mapping (Chapter 5).

The results from using these markers in the Lincoln F_2 population show that many of the *RMS* genes are associated with divaricate-related traits. *RMS1* showed

significant differences for several traits including node angle, leaflet width, leaflet length, rachis length, branch width and internode length suggesting *RMSI* could be contributing to all of these traits in the divaricate growth form. As the Lincoln East block F_2 are older and have been exposed to environmental conditions that may have influenced growth of these traits, a second F_2 population was developed in a more controlled environment.

3.5.2 The new PN F_2 population

The new PN F_2 population consisted of 69 individuals grown in a glasshouse. Divaricate traits such as leaf and branching phenotypes were measured for this population similar to the Lincoln F_2 population; however, the difference in age meant some traits had to be measured in different ways. These phenotypes showed approximately normal distributions for some phenotypes, such as leaflet number and rachis length, with significant positive skew. As these plants are much younger than the Lincoln population, size data cannot be compared; however, these traits often had similar distribution patterns to those of the Lincoln population (see Chapter 2) with most traits showing approximately normal distributions with moderate to significant positive skew. Comparisons between traits for the PN F_2 show that most traits are correlated (Figure 3.6). Leaf traits are all positively correlated with each other and also with node angle and branch width; all of these traits also displayed negative correlations with branch number indicating that as leaf size, branch width and node angle increase, branch number decreases and vice versa. This is consistent with the parental phenotypes as *S. tetraptera* has large leaves, straighter nodes and wider stems with generally less branching than *S. prostrata* and suggests that these traits may be under common genetic control. Internode length and branch angle were positively correlated with each other but did not display correlations with other traits in the PN F_2 population, suggesting that they may be under control of different genes. However, internode length did correlate

with the other traits in the Lincoln F_2. Differences between the two populations may result from the age difference of the plants or environmental factors differing between the two populations. Branch angle did not correlate with any traits in the Lincoln F_2; however, in the PN F_2, it did correlate with node angle which could suggest these traits may be co-regulated in young plants or that the age of the plants may affect the angles of branches.

The branch number traits in the PN F_2 population indicate similar correlations, except with node angle, to the branch number trait (BP10) in the Lincoln F_2 even with the difference in how the trait was measured. A statistically significant association between total branch number and *RMS1* was indicated in the PN F_2 but was not indicated with branch number in the Lincoln F_2 and may result from the difference in the way the trait was measured. Branch number in the PN F_2 was measured by counting primary, secondary and tertiary branching, separately and combined to a total branch number. In the Lincoln population, branch number was measured from outer branches, which would not include primary or secondary branches and therefore may not be comparable. Further, there would have been more variability in the branch distance from primary branches in the Lincoln F_2 population. Trait correlations are often observed for traits that are mapped to the same loci which can indicate these traits are controlled by a pleiotropic gene or genes (Xu et al., 2015; Chunlian et al., 2016; Wang et al., 2018). The correlations between the leaf traits and branch traits were also generally observed for the Lincoln F_2 providing further support for the correlations seen in that population and suggesting a genetic basis exists for these correlated traits.

3.5.3 RMS genes

RMS1 showed a similar segregation ratio in the PN F_2 as in the Lincoln F_2; in both populations, the gene was segregating as expected according to Mendelian principles. As

such, differences between the populations for association between *RMSI* and divaricate phenotypes may be due to either age-related differences, different environments, or both. In both populations, rachis length, leaflet length, leaflet width, and node angle were consistently associated with variation at *RMSI* (Table 3.19), indicating that these traits may be influenced by *RMSI* in *Sophora* and therefore the strigolactone pathway. Branch number was only statistically significant in the PN F_2 population for secondary and total branch number which may result from the difference in measuring method between the two populations. However, it is possible that these associations are due to a gene linked to the *RMSI* marker that is controlling these phenotypes and not the *RMSI* gene directly. Other *RMS* genes also displayed significance between parental alleles in the Lincoln F_2 population for some traits also suggesting the strigolactone pathway may contribute to the divaricate phenotype.

In comparisons between *S. prostrata* and *S. tetraptera*, amino acid replacements were observed in all the *RMS* sequences. Some of these did not result in a change in the class of amino acid, e.g. alanine to glycine, which are both classified as non-polar aliphatic amino acids. However, some substitutions resulted in amino acids with very different properties, e.g. glutamine (a polar and neutral amide) versus glutamate (acidic and negatively charged). The impact that an amino acid replacement will have on a protein is dependent on the position, whether it is in a functional or structurally important domain, and what type of amino acid change is made. Programs such as SIFT (Ng & Henikoff, 2003) and PROVEAN (Choi & Chan, 2015) can be used to predict if a substitution is likely to have an effect, but experimentation is still required to determine the actual effect on the protein.

Inferred protein sequences for the candidate genes indicated only a few amino acid replacements between *S. prostrata* and *S. tetraptera* for most of the genes and only

one of these, in *RMS1,* was predicted to be deleterious (Table 3.14); this was a change from arginine in *S. tetraptera* to serine in *S. prostrata.* All RMS1 (CCD8) protein sequences downloaded from Genbank from 12 species including monocots and dicots (see methods 3.3.2) have an arginine at this position except *Pisum sativum* which has a lysine. Arginine is positively charged and frequently found in protein binding or active sites. Arginine prefers replacements by the other positively charged amino acid, lysine, but can also tolerate changes to other polar amino acids in some situations; however even a change to lysine cannot be tolerated in all situations (Betts & Russell, 2003). As *RMS1* is functional in *P. sativum,* the lysine in this position likely does not affect the function of the protein, however, as all other sequences (see section 3.3.2), from diverse plant families including monocots and dicots, have arginine at this site this may be an important amino acid site where a replacement from another amino acid, such as serine in *S. prostrata,* may have a negative effect. This suggests that *S. prostrata RMS1* may not be functional or has reduced function and therefore the species may not be able to synthesise strigolactone or is less efficient at synthesis, causing increased branching.

RMS1/*CCD8* is known to occur in multiple copies in some plant species, such as maize, sorghum and rice, and multiple copies could compensate for the lack of function of one copy. However, in the species with multiple *CCD8*s, the *CCD8*bs (the name given to one of the copies of *CCD8*) form distinct groups from the *CCD8*/*RMS1* group with a closer relationship to *CCD1* and may indicate a diverse role for these copies in these species (Vallabhaneni et al., 2010). A single *CCD8* in maize is involved in the strigolactone pathway for axillary branching (Guan et al., 2012) despite the presence of multiple *CCD8*s indicating these copies may have diverse roles and cannot compensate for each other. The copy number in *Sophora* is unknown and it is possible a compensatory gene may be present if the RMS1 serine replacement has a deleterious effect. The initial PCR amplification of *RMS1* did have multiple bands but was performed

with degenerate sites in the primers which may result in non-specific amplification rather than the presence of multiple copies. While the later PCR did not indicate multiple copies in *S. prostrata*, the presence of multiple *RMSI* cannot be ruled out. Alternatively, a compensatory mutation in a different region of the gene could mean RMSI is still functional but further experiments would be required to determine the effect this replacement would have on the protein and the possible effect on strigolactone levels in *S. prostrata*. Full characterization of the complete gene sequences will be an important step in future investigations of the *Sophora* sequences.

Synonymous SNP mutations can also affect proteins even though no amino acid replacements are made. This often occurs through mRNA structure where the SNP can generate a more stable loop or alternatively loosening of the secondary structure of the mRNA. These can have an impact on translation and ribosomal translocation delaying protein synthesis and folding, as well as post-modification of the protein sequences (Nackley et al., 2006; Kudla et al., 2009; Plotkin & Kudla, 2011; Shabalina et al., 2013). A synonymous mutation, causing altered conformation of the protein despite similar mRNA and protein levels, was observed in the *MDR1* gene which showed altered drug and inhibiter interactions (Kimchi-Sarfaty et al., 2007). The presence of multiple rare codons affected the timing of folding and translocation to the membrane, which was hypothesised to be the cause of the altered structure of the protein. From *RMS* gene fragments, base pair substitutions resulting in nonsynonymous and synonymous substitutions were both observed in the *Sophora* orthologs (data not shown). Further research and full sequencing of these genes may reveal other mutations indicative of the involvement of these genes in the evolution of the divaricating habit in *Sophora*.

3.5.4 Strigolactone and *Sophora* divarication

Increased branching is one of several traits that define the divaricate growth form. The strigolactone pathway regulates axillary branching, which is an important part of increasing branching, and the *RMS* genes display associations with the genotypes of the *Sophora* F_2 indicating the strigolacone pathway may be involved in the divaricate phenotype. *RMS1,* which displayed significant differences between genotypes for the majority of the traits phenotyped, including branch number in the PN F_2, is involved in the biosynthesis of strigolactone (Flematti et al., 2016). *Pisum sativum RMS1* mutants are restored to wild-type by the application of synthetic strigolactone, GR24 (Dun et al., 2013). Synthetic strigolactone was applied to *S. prostrata* seedlings which could revert these to the non-divaricate, wild-type, form if the *RMS* pathway contributes to the divaricate form.

Branch number was not significantly different between control plants and GR24 treatments (Table 3.21) suggesting the strigolactone pathway in *S. prostrata* may not be lacking strigolactone. However, it is also possible *S. prostrata* cannot perceive the strigolactone signal as other *RMS* genes, such as *RMS2* and *RMS4,* are also significantly associated with some divaricate traits and these are involved in the feedback response and signal perception respectively. Non-functioning signal perception, for example in *RMS4* or *RMS3,* would prevent the perception of synthetic strigolactone and no change in phenotype would be observed even when sufficient strigolactone is supplied to the plant. None of the amino acid replacements in RMS2, 4 or 5 were indicated as deleterious; however, as the *Sophora* gene sequences are only partial fragments and the prediction analysis is not absolute, there may be other mutations that could affect function in the regions not sequenced that may be deleterious. Sequencing of the full gene sequences would be helpful in identifying all differences in the sequences; however, experimentation would still be required to determine effects in the plant.

The lack of response to GR24 may also be affected by the application of the hormone. The concentration selected for *S. prostrata* was based on the effectiveness of different concentrations of strigolactone in *Arabidopsis* and *Pisum* (Gomez-Roldan et al., 2008; Rasmussen et al., 2012) and from the amount of GR24 that could be obtained for the experiment, however, this may not be the optimum concentration needed for *S. prostrata.* Also, the strigolactone may have been degrading over time and therefore may have been less effective in later experiments. GR24 has been shown to have a half-life of around 9 days (Akiyama et al., 2010) although another study observed a half-life of as little as 4-9 hours in a methanol-water solution (Rasmussen et al., 2013); methanol is known to destabilise strigolactones (Akiyama et al., 2010). In my experiments, GR24 was suspended in acetone and stored below 8°C as suggested by the manufacturer to limit the degradation throughout the experiment; however, the effectiveness of GR24 may still degrade over time. Using freshly made solutions for each treatment to prevent degradation may help, but was not possible to perform in the *Sophora* GR24 experiment.

Despite branch number not showing a significant difference, leaflet length and width were significantly different between the controls and the treated plants indicating application of strigolactone does increase leaf size. However, GR24 typically contains enantiomers including a 2'S configuration of the D ring (Waters et al., 2017). Endogenous strigolactone in plants has only been observed to have a D ring with a 2'R configuration and the 2'S configuration can activate non-strigolactone signalling pathways, such as the karrikin pathway, indicating results from GR24 should be treated with caution (Scaffidi et al., 2014). This could mean that the significant differences in leaf size in the strigolactone experiment on *S. prostrata* seedlings may not be occurring through the *RMS* strigolactone pathway. The smaller trial with eight *S. tetraptera* seedlings, four treated and four as a control, showed a significant difference in leaflet width and, although this is a very small sample size, may suggest the difference in leaf

size in *S. prostrata* when GR24 is applied is not a result of a mutation in the *RMS* pathway, but results from a different mechanism as *S. tetraptera* is likely to produce endogenous strigolactone. It is possible the small sample size of this experiment is an issue as only nineteen total plants survived to treatment even though many more were started resulting in nine control and ten treated plants. The main difficulty in generating plants was caused by a fungus affecting the seeds and seedling despite treatment in H_2SO_4, indicating better conditions for growing *S. prostrata* or newer seeds, as they were collected around two years prior, may be required to get a better sample size. *Sophora prostrata* is also slow growing so it was difficult to get plants of similar ages for the experiment as some took much longer to grow once germinated. As a result, some seedlings were excluded due to a large age and size difference, even though seeds were all treated at the same time, which could have had an effect on the results if they were included.

3.5.5 ZIG

ZIG is a gene that encodes a vSNARE protein that, in *Arabidopsis thaliana* mutants, produces a zig zag shaped stem (Kato et al., 2002). While a zig zag stem is not present in all divaricates, it is a striking feature of the divaricate, *S. prostrata,* and is also seen in other divaricate species (Tomlinson, 1978). A *ZIG* ortholog was sequenced from *Sophora* and a genetic marker developed that distinguishes between the *Sophora* species. Significant differences between the *ZIG* genotype and phenotype was observed for leaflet width and branch width. However, node angle, which is the angle that gives the zig zag stem was not associated with *ZIG,* suggesting this gene is not involved in producing the zig zag stem of *S. prostrata.*

3.6 Conclusion

Strigolactone is a plant hormone involved in the control of axillary branching in land plants with the *RMS*/*MAX* genes responsible for biosynthesis, signal perception and response in this pathway. Mutants of these genes display similar phenotypes to those observed in the divaricate growth form. Many of the genes in this pathway show associations with divaricate phenotypes in two F_2 populations. Results from the PN F_2 population support the correlations among phenotypes from the Lincoln F_2 population. One gene, *RMSI,* displayed associations for multiple divaricate traits and an amino acid replacement was predicted as deleterious. Application of synthetic strigolactone to *S. prostrata* seedlings did not result in significant differences in branch number, indicating strigolactone may not be involved in branching in *S. prostrata* or mutations in other genes in the pathway have led to problems with perceiving the strigolactone signal. Leaflet length and width were significantly different after application of synthetic strigolactone; however, this may not result from the action of the strigolactone pathway. The candidate gene markers developed in this chapter are included in the linkage maps (Chapter 4) and QTL analyses (Chapter 5).

Given the broad taxonomic distribution of divaricating species in NZ and the conservation of the strigolactone pathway, it is conceivable that this pathway is at least partly involved in the independent formations of the divaricating form. Further research on the role of strigolactone in divarication of *Sophora* and in other divaricating species may be important in unravelling the genetic nature of the divaricate growth form. Future research on the potential involvement of strigolactone in the divaricate growth form, may, to our knowledge, be the first evidence of natural variation in regulation of shoot branching by strigolactone.

3.7 References

Akiyama, K., Matsuzaki, K., & Hayashi, H. (2005). Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature, 435*(7043), 824-827. 10.1038/nature03608

Akiyama, K., Ogasawara, S., Ito, S., & Hayashi, H. (2010). Structural requirements of strigolactones for hyphal branching in AM fungi. *Plant and Cell Physiology, 51*(7), 1104-1117. 10.1093/pcp/pcq058

Auldridge, M. E., Block, A., Vogel, J. T., Dabney-Smith, C., Mila, I., Bouzayen, M., Magallanes-Lundback, M., DellaPenna, D., McCarty, D. R., & Klee, H. J. (2006). Characterization of three members of the *Arabidopsis* carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family. *Plant Journal, 45*(6), 982-993. 10.1111/j.1365-313X.2006.02666.x

Batra, R., Agarwal, P., Tyagi, S., Saini, D. K., Kumar, V., Kumar, A., Kumar, S., Balyan, H. S., Pandey, R., & Gupta, P. K. (2019). A study of CCD8 genes/proteins in seven monocots and eight dicots. *Plos One, 14*(3) 10.1371/journal.pone.0213531

Bennett, T., & Leyser, O. (2014). Strigolactone signalling: standing on the shoulders of DWARFs. *Current Opinion in Plant Biology, 22*, 7-13. 10.1016/j.pbi.2014.08.001

Betts, M. J., & Russell, R. B. (2003). *Bioinformatics for Geneticists*. West Sussex, United Kingdom John Wiley & Sons.

Beveridge, C. A. (2006). Axillary bud outgrowth: sending a message. *Current Opinion in Plant Biology, 9*(1), 35-40. 10.1016/j.pbi.2005.11.006

Beveridge, C. A., Symons, G. M., Murfet, I. C., Ross, J. J., & Rameau, C. (1997). The rms1 mutant of pea has elevated indole-3-acetic acid levels and reduced root-sap zeatin riboside content but increased branching controlled by graft-transmissible signal(s). *Plant Physiology, 115*(3), 1251-1258.

Booker, J., Auldridge, M., Wills, S., McCarty, D., Klee, H., & Leyser, O. (2004). MAX3/CCD7 is a carotenoid cleavage dioxygenase required for the synthesis of a novel plant signaling molecule. *Current Biology, 14*(14), 1232-1238. 10.1016/j.cub.2004.06.061

Braun, N., de Saint Germain, A., Pillot, J.-P., Boutet-Mercey, S., Dalmais, M., Antoniadi, I., Li, X., Maia-Grondard, A., Le Signor, C., Bouteiller, N., Luo, D., Bendahmane, A., Turnbull, C., & Rameau, C. (2012). The pea TCP transcription factor PsBRC1 acts downstream of strigolactones to control shoot branching. *Plant Physiology, 158*(1), 225-238. 10.1104/pp.111.182725

Brewer, P. B., Dun, E. A., Ferguson, B. J., Rameau, C., & Beveridge, C. A. (2009). Strigolactone acts downstream of auxin to regulate bud outgrowth in pea and *Arabidopsis*. *Plant Physiology, 150*(1), 482-493. 10.1104/pp.108.134783

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for equality of variances. *Journal of the American Statistical Association, 69*(346), 364-367. 10.2307/2285659

Bulmer, M. G. (1979). *Principles of Statistics* (3rd ed.). Cambridge: Cambridge University Press.

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics, 31*(16), 2745-2747. 10.1093/bioinformatics/btv195

Chomicki, G., Coiro, M., & Renner, S. S. (2017). Evolution and ecology of plant architecture: integrating insights from the fossil record, extant morphology, developmental genetics and phylogenies. *Annals of Botany, 120*(6), 855-891. 10.1093/aob/mcx113

Chunlian, L., Guihua, B., Shiaoman, C., Carver, B., & Zhonghua, W. (2016). Single nucleotide polymorphisms linked to quantitative trait loci for grain quality traits in wheat. *Crop Journal, 4*(1), 1-11. 10.1016/j.cj.2015.10.002

Cook, C. E., Whichard, L. P., Turner, B., Wall, M. E., & Egley, G. H. (1966). Germination of witch-weed (*Striga lutea* Lour.): isolation and properties of a potent stimulant. *American Association for the Advancement of Science. Science, 154*(3753), 1189-1190.

Czarnecki, O., Yang, J., Wang, X., Wang, S., Muchero, W., Tuskan, G. A., & Chen, J.-G. (2014). Characterization of MORE AXILLARY GROWTH Genes in *Populus*. *Plos One, 9*(7) 10.1371/journal.pone.0102757

de Saint Germain, A., Clave, G., Badet-Denisot, M.-A., Pillot, J.-P., Cornu, D., Le Caer, J.-P., Burger, M., Pelissier, F., Retailleau, P., Turnbull, C., Bonhomme, S., Chory, J., Rameau, C., & Boyer, F.-D. (2016). An histidine covalent receptor and butenolide complex mediates strigolactone perception. *Nature Chemical Biology, 12*(10), 787-+. 10.1038/nchembio.2147

Delaux, P. M., Xie, X. N., Timme, R. E., Puech-Pages, V., Dunand, C., Lecompte, E., Delwiche, C. F., Yoneyama, K., Becard, G., & Sejalon-Delmas, N. (2012). Origin of strigolactones in the green lineage. *New Phytologist, 195*(4), 857-871. 10.1111/j.1469-8137.2012.04209.x

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin, 19*, 11-15.

Dun, E. A., de Saint Germain, A., Rameau, C., & Beveridge, C. A. (2013). Dynamics of strigolactone function and shoot branching responses in *Pisum sativum*. *Molecular Plant, 6*(1), 128-140. 10.1093/mp/sss131

Dun, E. A., Hanan, J., & Beveridge, C. A. (2009). Computational modeling and molecular physiology experiments reveal new insights into shoot branching in pea. *Plant Cell, 21*(11), 3459-3472. 10.1105/tpc.109.069013

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association, 56*(293), 52-&. 10.2307/2282330

Eichhorn, S. E., Evert, R. F., & Raven, P. H. (2005). *Biology of Plants* (7th ed.). United States of America: W. H. Freeman and Company.

Flematti, G. R., Scaffidi, A., Waters, M. T., & Smith, S. M. (2016). Stereospecificity in strigolactone biosynthesis and perception. *Planta, 243*(6), 1361-1373. 10.1007/s00425-016-2523-5

Foo, E., Buillier, E., Goussot, M., Foucher, F., Rameau, C., & Beveridge, C. A. (2005). The branching gene *RAMOSUS1* mediates interactions among two novel signals 464 and auxin in pea. *Plant Cell, 17*(2), 464-474. 10.1105/tpc.104.026716

Foo, E., Turnbull, C. G. N., & Beveridge, C. A. (2001). Long-distance signaling and the control of branching in the *rms1* mutant of peal. *Plant Physiology, 126*(1), 203-209. 10.1104/pp.126.1.203

Games, P. A., & Howell, J. F. (1976). Pairwise multiple comparison procedures with unequal N's and/or variances: A Monty Carlo Study *Journal of Educational Statistics, 1*(2), 113-125.

Gomez-Roldan, V., Fermas, S., Brewer, P. B., Puech-Pages, V., Dun, E. A., Pillot, J.-P., Letisse, F., Matusova, R., Danoun, S., Portais, J.-C., Bouwmeester, H., Becard, G., Beveridge, C. A., Rameau, C., & Rochange, S. F. (2008). Strigolactone inhibition of shoot branching. *Nature, 455*(7210), 189-U122. 10.1038/nature07271

Guan, J. C., Koch, K. E., Suzuki, M., Wu, S., Latshaw, S., Petruff, T., Goulet, C., Klee, H. J., & McCarty, D. R. (2012). Diverse Roles of Strigolactone Signaling in Maize Architecture and the Uncoupling of a Branching-Specific Subnetwork. *Plant Physiology, 160*(3), 1303-1317. 10.1104/pp.112.204503

Hashiguchi, Y., Niihama, M., Takahashi, T., Saito, C., Nakano, A., Tasaka, M., & Morita, M. T. (2010). Loss-of-function mutations of retromer large subunit genes suppress the phenotype of an *Arabidopsis* zig mutant that lacks Qb-SNARE VTI11. *Plant Cell, 22*(1), 159-172. 10.1105/tpc.109.069294

Jiang, L., Liu, X., Xiong, G., Liu, H., Chen, F., Wang, L., Meng, X., Liu, G., Yu, H., Yuan, Y., Yi, W., Zhao, L., Ma, H., He, Y., Wu, Z., Melcher, K., Qian, Q., Xu, H. E., Wang, Y., & Li, J. (2013). DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature, 504*(7480), 401-+. 10.1038/nature12870

Johnson, X., Brcich, T., Dun, E. A., Goussot, M., Haurogne, K., Beveridge, C. A., & Rameau, C. (2006). Branching genes are conserved across species. Genes controlling a novel signal in pea are coregulated by other long-distance signals. *Plant Physiology, 142*(3), 1014-1026. 10.1104/pp.106.087676

Kato, T., Morita, M. T., Fukaki, H., Yamauchi, Y., Uehara, M., Niihama, M., & Tasaka, M. (2002). SGR2, a Phospholipase-Like protein, and ZIG/SGR4, a SNARE, are involved in the shoot gravitropism of *Arabidopsis*. *The Plant Cell Online, 14*(1), 33-46. 10.1105/tpc.010215

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649. 10.1093/bioinformatics/bts199

Kelly, D. (1994). Towards a numerical definition of divaricate (interlaced small-leaved) shrubs. *New Zealand Journal of Botany, 32*(4), 509-518.

Kimchi-Sarfaty, C., Oh, J. M., Kim, I.-W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V., & Gottesman, M. M. (2007). A "silent" polymorphism in the *MDR1* gene changes substrate specificity. *Science, 315*(5811), 525-528. 10.1126/science.1135308

Koltai, H., Dor, E., Hershenhorn, J., Joel, D. M., Weininger, S., Lekalla, S., Shealtiel, H., Bhattacharya, C., Eliahu, E., Resnick, N., Barg, R., & Kapulnik, Y. (2010). Strigolactones' effect on root growth and root-hair elongation may be mediated by auxin-efflux carriers. *Journal of Plant Growth Regulation, 29*(2), 129-136. 10.1007/s00344-009-9122-7

Koltai, H., LekKala, S. P., Bhattacharya, C., Mayzlish-Gati, E., Resnick, N., Wininger, S., Dor, E., Yoneyama, K., Yoneyama, K., Hershenhorn, J., Joel, D. M., & Kapulnik, Y. (2010). A tomato strigolactone-impaired mutant displays aberrant shoot morphology and plant interactions. *Journal of Experimental Botany, 61*(6), 1739-1749. 10.1093/jxb/erq041

Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-Sequence Determinants of gene expression in *Escherichia coli*. *Science, 324*(5924), 255-258. 10.1126/science.1170160

Kuroda, H., Takahashi, N., Shimada, H., Seki, M., Shinozaki, K., & Matsui, M. (2002). Classification and expression analysis of *Arabidopsis* F-box-containing protein genes. *Plant and Cell Physiology, 43*(10), 1073-1085. 10.1093/pcp/pcf151

Ligerot, Y., de Saint Germain, A., Waldie, T., Troadec, C., Citerne, S., Kadakia, N., Pillot, J.-P., Prigge, M., Aubert, G., Bendahmane, A., Leyser, O., Estelle, M., Debelle, F., & Rameau, C. (2017). The pea branching *RMS2* gene encodes the *PsAFB4/5* auxin receptor and is involved in an auxin-strigolactone regulation loop. *Plos Genetics, 13*(12) 10.1371/journal.pgen.1007089

Morris, S. E., Turnbull, C. G. N., Murfet, I. C., & Beveridge, C. A. (2001). Mutational analysis of branching in pea. Evidence that *RMS1* and *RMS5* regulate the same novel signal. *Plant Physiology, 126*(3), 1205-1213. 10.1104/pp.126.3.1205

Nackley, A. G., Shabalina, S. A., Tchivileva, I. E., Satterfield, K., Korchynskyi, O., Makarov, S. S., Maixner, W., & Diatchenko, L. (2006). Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science, 314*(5807), 1930-1933. 10.1126/science.1131262

Ng, P. C., & Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research, 31*(13), 3812-3814. 10.1093/nar/gkg509

Peng, J. R., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., Beales, J., Fish, L. J., Worland, A. J., Pelica, F., Sudhakar, D., Christou, P., Snape, J. W., Gale, M. D., & Harberd, N. P. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature, 400*(6741), 256-261.

Plotkin, J. B., & Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics, 12*, 32-42.

Plourde, A., Krause, C., & Lord, D. (2009). Spatial distribution, architecture, and development of the root system of *Pinus banksiana* Lamb. in natural and planted stands. *Forest Ecology and Management, 258*(9), 2143-2152. 10.1016/j.foreco.2009.08.016

Quine, C. P. (1990). Planting at stump - preliminary evidence of its effect on root architecture and tree stability. *Research Information Note - Forestry Commission Research Division*(169), 3 pp.

R Development Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rasmussen, A., Heugebaert, T., Matthys, C., Van Deun, R., Boyer, F.-D., Goormachtig, S., Stevens, C., & Geelen, D. (2013). A fluorescent alternative to the synthetic strigolactone GR24. *Molecular Plant, 6*(1), 100-112. 10.1093/mp/sss110

Rasmussen, A., Mason, M. G., De Cuyper, C., Brewer, P. B., Herold, S., Agusti, J., Geelen, D., Greb, T., Goormachtig, S., Beeckman, T., & Beveridge, C. A. (2012). Strigolactones suppress adventitious rooting in *Arabidopsis* and pea. *Plant Physiology, 158*(4), 1976-1987. 10.1104/pp.111.187104

Reinhardt, D., & Kuhlemeier, C. (2002). Plant architecture. *Embo Reports, 3*(9), 846-851. 10.1093/embo-reports/kvf177

Revelle, W. (2016). psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, Illinois: Northwestern University.

Ruter, J. M., & Ingram, D. L. (1991). Germination and morphology of *Sophora secundiflora* seeds following scarification. *Hortscience, 26*(3), 256-257.

Salmon, J., Ward, S. P., Hanley, S. J., Leyser, O., & Karp, A. (2014). Functional screening of willow alleles in *Arabidopsis* combined with QTL mapping in willow (*Salix*) identifies *SxMAX4* as a coppicing response gene. *Plant Biotechnology Journal, 12*(4), 480-491. 10.1111/pbi.12154

Scaffidi, A., Waters, M. T., Sun, Y. K., Skelton, B. W., Dixon, K. W., Ghisalberti, E. L., Flematti, G. R., & Smith, S. M. (2014). Strigolactone hormones and their stereoisomers signal through two related receptor proteins to induce different physiological responses in *Arabidopsis*. *Plant Physiology, 165*(3), 1221-1232. 10.1104/pp.114.240036

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods, 9*(7), 676-682. 10.1038/nmeth.2019

Shabalina, S. A., Spiridonov, N. A., & Kashina, A. (2013). Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Research, 41*(4), 2073-2094. 10.1093/nar/gks1205

Shimizu-Sato, S., & Mori, H. (2001). Control of outgrowth and dormancy in axillary buds. *Plant Physiology, 127*(4), 1405-1413. 10.1104/pp.010841

Sorefan, K., Booker, J., Haurogne, K., Goussot, M., Bainbridge, K., Foo, E., Chatfield, S., Ward, S., Beveridge, C., Rameau, C., & Leyser, O. (2003). *MAX4* and *RMS1* are orthologous dioxygenase-like genes that regulate shoot branching in *Arabidopsis* and pea. *Genes & Development, 17*(12), 1469-1474. 10.1101/gad.256603

Soto, M. J., Fernandez-Aparicio, M., Castellanos-Morales, V., Garcia-Garrido, J. M., Ocampo, J. A., Delgado, M. J., & Vierheilig, H. (2010). First indications for the involvement of strigolactones on nodule formation in alfalfa (*Medicago sativa*). *Soil Biology & Biochemistry, 42*(2), 383-385. 10.1016/j.soilbio.2009.11.007

Soundappan, I., Bennett, T., Morffy, N., Liang, Y., Stang, J. P., Abbas, A., Leyser, O., & Nelson, D. C. (2015). SMAX1-LIKE/D53 family members enable distinct MAX2-dependent responses to strigolactones and karrikins in *Arabidopsis*. *Plant Cell, 27*(11), 3143-3159. 10.1105/tpc.15.00562

Stirnberg, P., Furner, I. J., & Ottoline Leyser, H. M. (2007). MAX2 participates in an SCF complex which acts locally at the node to suppress shoot branching. *Plant Journal, 50*(1), 80-94. 10.1111/j.1365-313X.2007.03032.x

Takeda, F., Krewer, G., Li, C. Y., MacLean, D., & Olmstead, J. W. (2013). Techniques for increasing machine harvest efficiency in highbush blueberry. *Horttechnology, 23*(4), 430-436.

Tomlinson, P. S. (1978). Some qualitative and quantitive aspects of New Zealand divaricating shrubs. *New Zealand Journal of Botany, 16*, 299-309.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics, 5*(2), 99-114. 10.2307/3001913

Umehara, M., Hanada, A., Yoshida, S., Akiyama, K., Arite, T., Takeda-Kamiya, N., Magome, H., Kamiya, Y., Shirasu, K., Yoneyama, K., Kyozuka, J., & Yamaguchi, S. (2008). Inhibition of shoot branching by new terpenoid plant hormones. *Nature, 455*(7210), 195-U129. 10.1038/nature07272

Vallabhaneni, R., Bradbury, L. M. T., & Wurtzel, E. T. (2010). The carotenoid dioxygenase gene family in maize, sorghum, and rice. *Archives of Biochemistry and Biophysics, 504*(1), 104-111. 10.1016/j.abb.2010.07.019

Van Etten, M. L., Houliston, G. J., Mitchell, C. M., Heenan, P. B., Robertson, A. W., & Tate, J. A. (2014). *Sophora microphylla* (Fabaceae) microsatellite markers and their utility across the genus. *Applications in Plant Sciences, 2*(3) 10.3732/apps.1300081

Vincent, Q. V. (2011). A ggplot2 based biplot (Version 0.55).

Wang, X., Chen, Z., Li, Q., Zhang, J., Liu, S., & Duan, D. (2018). High-density SNP-based QTL mapping and candidate gene screening for yield-related blade length and width in *Saccharina japonica* (Laminariales, Phaeophyta). *Scientific Reports, 8* 10.1038/s41598-018-32015-y

Waters, M. T., Gutjahr, C., Bennett, T., & Nelson, D. C. (2017). Strigolactone Signaling and Evolution. In S. S. Merchant (Ed.), *Annual Review of Plant Biology, Vol 68* (Vol. 68, pp. 291-322). 10.1146/annurev-arplant-042916-040925

Wei, T., & Simko, V. (2016). Visualization of a Correlation Matrix. R package version 0.77: https://CRAN.R-project.org/package=corrplot.

Woo, H. R., Chung, K. M., Park, J. H., Oh, S. A., Ahn, T., Hong, S. H., Jang, S. K., & Nam, H. G. (2001). ORE9, an F-box protein that regulates leaf senescence in Arabidopsis. *Plant Cell, 13*(8), 1779-1790. 10.1105/tpc.13.8.1779

Xie, X., Yoneyama, K., Kisugi, T., Uchida, K., Ito, S., Akiyama, K., Hayashi, H., Yokota, T., Nomura, T., & Yoneyama, K. (2013). Confirming stereochemical structures of strigolactones produced by rice and tobacco. *Molecular Plant, 6*(1), 153-163. 10.1093/mp/sss139

Xie, X. N., Yoneyama, K., & Yoneyama, K. (2010). The Strigolactone Story. In N. K. VanAlfen, G. Bruening, & J. E. Leach (Eds.), *Annual Review of Phytopathology, Vol 48* (Vol. 48, pp. 93-117). 10.1146/annurev-phyto-073009-114453

Xu, Y., Huang, L., Ji, D., Chen, C., Zheng, H., & Xie, C. (2015). Construction of a dense genetic linkage map and mapping quantitative trait loci for economic traits of a doubled haploid population of *Pyropia haitanensis* (Bangiales, Rhodophyta). *BMC Plant Biology, 15* 10.1186/s12870-015-0604-4

Yao, R., Ming, Z., Yan, L., Li, S., Wang, F., Ma, S., Yu, C., Yang, M., Chen, L., Chen, L., Li, Y., Yan, C., Miao, D., Sun, Z., Yan, J., Sun, Y., Wang, L., Chu, J., Fan, S., He, W., Deng, H., Nan, F., Li, J., Rao, Z., Lou, Z., & Xie, D. (2016). DWARF14 is a non-canonical hormone receptor for strigolactone. *Nature, 536*(7617), 469-+. 10.1038/nature19073

3.8 Appendix

Appendix 3.1. ANOVA results for candidate genes in the Lincoln F_2.

Candidate gene	Phenotype	Sum Square	Mean Square	Df	Fvalue	Pr(>F)
RMS1	Leaflet Number	25.9	12.96	2	0.924	0.401
	Rachis Length	3914	1956.9	2	9.716	1.58E-04
	Leaflet Length	155.9	77.94	2	8.909	3.07E-04
	Leaflet Width	7.1	3.551	2	7.705	8.41E-04
	Internode Length	215.5	107.73	2	3.761	2.72E-02
	Branch Width	1.849	0.9247	2	4.453	1.45E-02
	Branch Number	19.7	9.857	2	2.565	8.28E-02
	Node Angle	139.5	69.73	2	4.647	1.21E-02
	Branch Angle	153	76.66	2	1.21	3.03E-01
RMS2	Leaflet Number	172.9	86.44	2	7.026	0.0015
	Rachis Length	2179	1089.4	2	4.911	9.59E-03
	Leaflet Length	80.7	40.34	2	4.188	1.84E-02
	Leaflet Width	5.47	2.7333	2	5.693	4.78E-03
	Internode Length	16.3	8.131	2	0.262	7.70E-01
	Branch Width	0.007	0.00325	2	0.014	9.86E-01
	Branch Number	12.3	6.161	2	1.568	2.14E-01
	Node Angle	107.3	53.67	2	3.489	3.50E-02
	Branch Angle	120	60.23	2	0.945	3.93E-01
RMS4	Leaflet Number	59.1	29.55	2	2.167	0.121
	Rachis Length	1770	885.1	2	3.906	2.38E-02
	Leaflet Length	102.6	51.28	2	5.469	5.83E-03
	Leaflet Width	1.66	0.8301	2	1.581	2.12E-01
	Internode Length	10.4	5.18	2	0.167	8.47E-01
	Branch Width	0.229	0.1146	2	0.506	6.05E-01
	Branch Number	6.4	3.211	2	0.803	4.51E-01
	Node Angle	88.3	44.16	2	2.83	6.46E-02
	Branch Angle	44	21.79	2	0.337	7.15E-01
RMS5	Leaflet Number	9.2	4.598	2	0.323	0.725
	Rachis Length	267	133.3	2	0.545	5.82E-01
	Leaflet Length	51.3	25.64	2	2.569	8.26E-02
	Leaflet Width	4.4	2.1980	2	4.461	1.44E-02
	Internode Length	58.3	29.14	2	0.956	3.89E-01
	Branch Width	0.233	0.1165	2	0.514	6.00E-01
	Branch Number	0.1	0.049	2	0.012	9.88E-01
	Node Angle	41.4	20.73	2	1.283	2.83E-01
	Branch Angle	318	158.76	2	2.585	8.14E-02
ZIG	Leaflet Number	60.8	30.38	2	2.23	0.114
	Rachis Length	1929	964.4	2	4.291	1.68E-02
	Leaflet Length	89.7	44.83	2	4.705	1.15E-02

Candidate gene	Phenotype	Sum Square	Mean Square	Df	Fvalue	Pr(>F)
ZIG	Leaflet Width	5.34	2.6681	2	5.54	5.48E-03
	Internode Length	70	34.99	2	1.152	3.21E-01
	Branch Width	0.833	0.4166	2	1.897	1.56E-01
	Branch Number	11.7	5.859	2	1.488	2.32E-01
	Node Angle	30.6	15.30	2	0.94	3.95E-01
	Branch Angle	2	0.87	2	0.013	9.87E-01

Appendix 3.2. Bonferroni corrected p-values of ANOVA for the candidate genes.

	RMS1 *Lincoln*	*RMS1* *PN F_2*	*RMS2* *Lincoln*	*RMS4* *Lincoln*	*RMS5* *Lincoln*	*ZIG* *Lincoln*
Leaflet Number	1	0.114624	0.0135*	1	1	1
Leaflet Length	0.002763*	1.8936E-07*	0.1656	0.05247	0.7434	0.1035
Leaflet width	0.007569*	6.6156E-07*	0.04302*	1	0.1296	0.04932*
Rachis Length	0.001422*	9.45E-07*	0.08631	0.2142	1	0.1512
Internode Length	0.2448	0.41244	1	1	1	1
Branch Width	0.1305	1	1	1	1	1
Branch Number	0.7452	-	1	1	1	1
Node Angle	0.1089	0.11598	0.315	0.5814	1	1
Branch Angle	1	1	1	1	0.7326	1
Primary Branches	-	0.78456	-	-	-	-
Secondary Branch	-	0.098208	-	-	-	-
Tertiary Branches	-	1	-	-	-	-
Total Branches	-	0.041736*	-	-	-	-

* indicates significance after Bonferroni correction, Bonferroni corrected p-value for all genes of p=0.00556

Appendix 3.3. Levenes test for homogeneity of variances in the Lincoln F_2.

Candidate Gene	Phenotype	DF	F value	P value
RMS1	Leaflet Number	2	0.6282	0.536
	Rachis Length	2	1.225	0.2989
	Leaflet Length	2	1.0323	0.3606
	Leaflet Width	2	0.1914	0.8262
	Internode Length	2	0.3802	0.6848
	Branch Width	2	1.6063	0.2067
	Branch Number	2	0.3222	0.7255
	Node Angle	2	1.0822	0.3435
	Branch Angle	2	1.4627	0.2374
RMS2	Leaflet Number	2	1.5415	0.22
	Rachis Length	2	1.7308	0.1833
	Leaflet Length	2	1.1669	0.3163
	Leaflet Width	2	0.4162	0.6609
	Internode Length	2	0.2525	0.7774
	Branch Width	2	1.2592	0.2891
	Branch Number	2	1.9757	0.145
	Node Angle	2	0.407	0.667
	Branch Angle	2	0.2092	0.8117
RMS4	Leaflet Number	2	0.2437	0.7843
	Rachis Length	2	0.1993	0.8197
	Leaflet Length	2	0.8304	0.4394
	Leaflet Width	2	0.2035	0.8162
	Internode Length	2	0.6223	0.5391
	Branch Width	2	1.3967	0.253
	Branch Number	2	0.4192	0.6589
	Node Angle	2	2.0258	0.1382
	Branch Angle	2	1.7012	0.1887
RMS5	Leaflet Number	2	0.8083	0.449
	Rachis Length	2	1.2333	0.2965
	Leaflet Length	2	0.8702	0.4226
	Leaflet Width	2	0.9313	0.398
	Internode Length	2	1.2663	0.2871
	Branch Width	2	0.9509	0.3904
	Branch Number	2	0.9276	0.3995
	Node Angle	2	0.3377	0.7144
	Branch Angle	2	2.1502	0.1228
ZIG	Leaflet Number	2	2.5355	0.0852
	Rachis Length	2	1.3349	0.2686
	Leaflet Length	2	0.5463	0.5811
	Leaflet Width	2	0.8954	0.4123
	Internode Length	2	0.5851	0.5593
	Branch Width	2	1.5806	0.2118
	Branch Number	2	0.638	0.5309
	Node Angle	2	0.7232	0.4882
	Branch Angle	2	0.7063	0.4964

Appendix 3.4. Tukeys HSD test results for significance in the Lincoln F_2.

Candidate gene	Trait	Comparison	Difference	lwr	upr	P adj
RMS1	Leaflet Length*	T - P	4.166188	1.7744001	6.5579757	0.0002259
		H - P	1.101718	-0.6704973	2.8739340	0.3041697
		H - T	-3.064470	-5.1933795	-0.9355596	0.0026296
	Leaflet Width*	T - P	0.8616382	0.3126973	1.4105791	0.0009492
		H - P	0.1532278	-0.2535146	0.5599703	0.6426437
		H - T	-0.7084104	-1.1970180	-0.2198028	0.0024294
	Rachis Length*	T - P	21.102558	9.626586	32.578530	0.0000967
		H - P	6.646331	-1.856889	15.149551	0.1553463
		H - T	-14.456227	-24.670892	-4.241562	0.0031569
	Internode Length	T - P	-3.9977747	-8.3256247	0.3300753	0.0763318
		H - P	0.3745685	-2.8321891	3.5813261	0.9581268
		H - T	4.3723432	0.5201607	8.2245257	0.0220995
	Branch Width*	T - P	0.44182453	0.07333816	0.81031091	0.0145840
		H – P	0.08340367	-0.1896295	0.35643681	0.7472064
		H - T	-0.3584209	-0.6864075	-0.0304343	0.0287877
	Node Angle*	T - P	3.8870288	0.754867	7.019191	0.0109979
		H - P	0.8611553	-1.459647	3.181957	0.6511174
		H - T	-3.0258736	-5.813784	-0.237963	0.0301342
RMS2	Leaflet Number*	T – P	3.5556319	0.7819216	6.3293421	0.0082950
		H - P	0.6830357	-1.8584812	3.2245526	0.7978805
		H - T	-2.8725962	-4.9100621	-0.8351302	0.0032861
	Leaflet Length	T - P	2.08447802	-0.3699301	4.5388861	0.1121998
		H - P	-0.01833147	-2.2672757	2.2306127	0.9997916
		H - T	-2.1028095	-3.9057278	-0.2998912	0.0180738
	Leaflet Width*	T - P	0.6459856	0.0980749	1.1938963	0.0166417
		H - P	0.1432943	-0.3587496	0.6453381	0.7752553
		H - T	-0.5026913	-0.9051664	-0.1002162	0.0104136
	Rachis Length	T- P	11.5102747	-0.2673764	23.28793	0.0568778
		H - P	0.8033036	-9.9884144	11.59502	0.9827716
		H - T	-10.7069712	-19.3584021	-2.05554	0.0112560
	Node Angle	T - P	2.45436137	-0.646993	5.5557220	0.1484441
		H - P	0.04385798	-2.7978809	2.8855969	0.9992530
		H - T	-2.41050339	-4.6886493	-0.1323575	0.0355538
RMS4	Leaflet Length*	T - P	2.7196859	0.3513209	5.0880509	0.0202995
		H - P	0.5121215	-1.6540374	2.6782803	0.8395620

Candidate gene	Trait	Comparison	Difference	lwr	upr	P adj
	Rachis Length	H - T	-2.2075644	-3.9928794	-0.4222495	0.0113392
		T - P	9.77979487	-1.863957	21.423546	0.1175332
		H - P	-0.06675177	-10.716384	10.582880	0.9998768
		H - T	-9.8465466	-18.623810	-1.069283	0.0240246
RMS5	Leaflet Width*	T - P	0.5734197	0.09412653	1.0527129	0.0148414
		H - P	0.4371056	-0.0030978	0.8773091	0.0520518
		H - T	-0.1363141	-0.5602415	0.2876133	0.7241602
ZIG	Leaflet Length	T - P	-1.826253	-4.458339	0.8058316	0.2285349
		H - P	-3.130810	-5.721701	-0.5399177	0.0137030
		H - T	-1.304556	-2.976970	0.3678574	0.1564873
	Leaflet Width	T - P	-0.5860590	-1.1778258	0.00570779	0.0528302
		H - P	-0.8059673	-1.3884727	-0.22346184	0.0039946
		H - T	-0.2199082	-0.5959139	0.15609739	0.3479833
	Rachis Length	T - P	-7.523826	-20.30779	5.260133	0.3434479
		H - P	-14.114982	-26.69887	-1.531097	0.0240477
		H - T	-6.591156	-14.71402	1.531707	0.1349954

T – *S. tetraptera* homozygotes, *P* – *S. prostrata* homozygotes, H – heterozygotes, * indicates traits displaying a significant difference between parental alleles.

Appendix 3.5. Means of parental homozygotes for traits in the Lincoln F_2 for *RMS1*.

	S. prostrata mean	*S. tetraptera mean*
Leaflet Number	14.42391	16.01786
Leaflet Length (mm)	7.845598	12.011786
Leaflet width (mm)	3.623451	4.485089
Rachis Length (mm)	30.55342	51.65598
Internode Length (mm)	14.32179	10.32401
Branch Width (mm)	2.514783	2.956607
Branch Number	5.086957	3.714286
Node Angle (°)	164.1796	168.0666
Branch Angle (°)	61.60485	61.77977

Appendix 3.6. ANOVA results for the candidate genes in the maternal direction AB population.

Candidate Gene	Phenotype	Sum square	Mean square	DF	F value	P-value
RMS1	Leaflet Number	2.5	1.239	2	0.067	0.935
	Rachis Length	2718	1359	2	5.49	0.00891
	Leaflet Length	111.0	55.49	2	6.276	0.00502*
	Leaflet Width	4.651	2.3257	2	5.954	0.00633
	Internode Length	37.3	18.64	2	0.71	0.499
	Branch Width	0.189	0.09458	2	0.74	0.485
	Branch Number	21.65	10.827	2	2.614	0.0888
	Node Angle	37.91	18.96	2	2.396	0.107
	Branch Angle	63	31.48	2	0.664	0.522
RMS2	Leaflet Number	135.5	67.77	2	4.768	0.0154
	Rachis Length	1795	897.3	2	3.247	0.052
	Leaflet Length	41.2	20.58	2	1.867	0.171
	Leaflet Width	3.451	1.7253	2	4.03	0.0275
	Internode Length	38.3	19.14	2	0.73	0.49
	Branch Width	0.158	0.07881	2	0.612	0.549
	Branch Number	33.52	16.76	2	4.446	0.0198
	Node Angle	50.19	25.093	2	3.334	0.0484
	Branch Angle	58.6	29.32	2	0.617	0.546
RMS4	Leaflet Number	58.6	29.28	2	1.762	0.188
	Rachis Length	1186	592.9	2	2.007	0.151
	Leaflet Length	44.0	22.01	2	2.013	0.15
	Leaflet Width	0.375	0.1875	2	0.358	0.702
	Internode Length	82.1	41.05	2	1.653	0.207
	Branch Width	0.21	0.1050	2	0.826	0.447
	Branch Number	5.91	2.953	2	0.637	0.535
	Node Angle	35.25	17.627	2	2.205	0.127
	Branch Angle	9.4	4.69	2	0.095	0.909
RMS5	Leaflet Number	11.2	5.612	2	0.31	0.736
	Rachis Length	78	38.8	2	0.118	0.889
	Leaflet Length	15.8	7.891	2	0.668	0.52
	Leaflet Width	1.35	0.6752	2	1.367	0.269
	Internode Length	15.1	7.567	2	0.281	0.757
	Branch Width	0.252	0.1259	2	1	0.379
	Branch Number	0.51	0.256	2	0.053	0.948
	Node Angle	4.34	2.171	2	0.242	0.786
	Branch Angle	201.3	100.66	2	2.345	0.113
ZIG	Leaflet Number	69.9	34.97	2	2.15	0.133
	Rachis Length	567	283.5	2	0.901	0.416
	Leaflet Length	11.4	5.694	2	0.476	0.625
	Leaflet Width	2.068	1.0339	2	2.193	0.128

Candidate Gene	Phenotype	Sum square	Mean square	DF	F value	P-value
	Internode Length	26	13.02	2	0.49	0.617
	Branch Width	0.374	0.1869	2	1.531	0.232
	Branch Number	14.65	7.323	2	1.68	0.202
	Node Angle	20.31	10.15	2	1.2	0.314
	Branch Angle	36.5	18.26	2	0.378	0.688

* indicates significance after Bonferroni correction with corrected p value of 0.00556.

Appendix 3.7. Tukeys post hoc test results for *RMS1* in the AB population of the Lincoln F_2.

Candidate gene	Trait	Comparison	Difference	lwr	upr	P adj
RMS1	Leaflet Length*	T - P	4.8582813	0.9121118	8.804451	0.0131427
		H - P	0.1904958	-2.8453262	3.226323	0.9869822
		H - T	-4.6677827	-8.0502137	-1.285352	0.0051624

* indicates significance between parental homozygote alleles

Appendix 3.8. ANOVA results for candidate genes in the maternal direction CD population.

Candidate Gene	Phenotype	Sum square	Mean square	DF	F value	P-value
RMS1	Leaflet Number	27.22	13.610	2	2.635	0.0817
	Rachis Length	1545	772.7	2	6.013	0.00457*
	Leaflet Length	60.0	30.023	2	3.903	0.0266
	Leaflet Width	2.841	1.4205	2	2.906	0.064
	Internode Length	195.6	97.80	2	3.191	0.0496
	Branch Width	2.406	1.2029	2	5.397	0.00755
	Branch Number	4.5	2.252	2	0.637	0.533
	Node Angle	96.4	48.21	2	2.944	0.0618
	Branch Angle	122	61.23	2	0.82	0.446
RMS2	Leaflet Number	23.1	11.552	2	2.201	0.121
	Rachis Length	283	141.3	2	0.919	0.405
	Leaflet Length	27.0	13.519	2	1.619	0.208
	Leaflet Width	1.882	0.9411	2	1.853	0.167
	Internode Length	2.2	1.10	2	0.032	0.969
	Branch Width	0.23	0.1151	2	0.432	0.651
	Branch Number	13.07	6.535	2	1.943	0.154
	Node Angle	37.1	18.57	2	1.057	0.355
	Branch Angle	276	138.14	2	1.93	0.156
RMS4	Leaflet Number	65.6	32.8	2	7.457	0.00147*
	Rachis Length	1565	782.3	2	6.107	0.00424*

Candidate Gene	Phenotype	Sum square	Mean square	DF	F value	P-value
	Leaflet Length	88.8	44.41	2	6.241	0.0038*
	Leaflet Width	2.214	1.1071	2	2.208	0.12
	Internode Length	9.3	4.65	2	0.135	0.874
	Branch Width	0.23	0.1151	2	0.432	0.652
	Branch Number	3.09	1.544	2	0.433	0.651
	Node Angle	110.1	55.07	2	3.421	0.0405
	Branch Angle	98	49.23	2	0.655	0.524
RMS5	Leaflet Number	3.73	1.864	2	0.331	0.72
	Rachis Length	23	11.28	2	0.071	0.932
	Leaflet Length	18.7	9.342	2	1.097	0.342
	Leaflet Width	2.436	1.2179	2	2.451	0.0965
	Internode Length	101	50.49	2	1.552	0.222
	Branch Width	0.134	0.06716	2	0.25	0.78
	Branch Number	0.85	0.427	2	0.118	0.889
	Node Angle	18.2	9.098	2	0.507	0.605
	Branch Angle	100	49.92	2	0.665	0.519
ZIG	Leaflet Number	16.88	8.439	2	1.571	0.218
	Rachis Length	899	449.6	2	3.179	0.0502
	Leaflet Length	78.3	39.15	2	5.343	0.00789
	Leaflet Width	2.924	1.4620	2	3.001	0.0588
	Internode Length	84.4	42.21	2	1.284	0.286
	Branch Width	0.274	0.1368	2	0.515	0.601
	Branch Number	8.45	4.227	2	1.223	0.303
	Node Angle	1.4	0.711	2	0.039	0.962
	Branch Angle	114	57.14	2	0.764	0.471

* indicates significance after Bonferroni correction with corrected p value of 0.00556.

Appendix 3.9. Levenes test for homogeneity of variance in the AB and CD populations.

Population	Candidate Gene	Phenotype	DF	F value	P-value
AB Lincoln F_2	*RMS1*	Leaflet Length	2	1.0912	0.348
CD Lincoln F_2	*RMS1*	Leaflet length	2	5.3598	0.007781*
	RMS4	Leaflet number	2	0.3176	0.7293
		Rachis length	2	0.0337	0.9668
		Leaflet length	2	0.5351	0.5889

* indicates significance ($p < 0.05$)

Appendix 3.10. Post hoc tests results for the significantly different traits in the AB and CD populations.

Population	Candidate gene	Trait	Comparison	Difference	lwr	upr	P adj
AB^{T}	*RMS1*	Leaflet Length*	T - P	4.8582	0.91212	8.8044	0.01314
			H - P	0.19049	-2.845	3.2263	0.9869
			H - T	-4.6677	-8.050	-1.2853	0.0052
CD^{T}	*RMS4*	Leaflet number*	T - P	1.96315	-0.016	3.9422	0.0523
			H - P	-0.4666	-2.373	1.4400	0.8254
			H - T	-2.4298	-3.985	-0.874	0.0012
		Rachis Length*	T - P	11.1927	0.5120	21.873	0.0381
			H - P	-0.1933	-10.48	10.097	0.9988
			H - T	-11.3861	-19.781	-2.991	0.0053
		Leaflet Length*	T - P	2.9548	0.4375	5.4721	0.0178
			H - P	0.3842	-2.041	2.8094	0.9226
			H - T	-2.5706	-4.549	-0.5919	0.0078

Population	Candidate gene	Trait	Comparison	Difference	lo	hi	P	t	df
CD^{GH}	*RMS1*	Rachis Length*	T - P	16.63	9.36	23.90	<0.001	5.86	17.22
			H - P	8.59	0.98	16.2	0.024	2.74	42.69
			H - T	-8.04	-16.15	0.06	0.052	2.46	27.33

* indicates significant difference between parental allele homozygotes. T – indicates Tukeys post hoc test, GH – indicates Games-Howell post hoc test.

Appendix 3.11. ANOVA results for *RMS1* in the PNF_2.

Phenotype	Sum Square	Mean Square	Df	Fvalue	Pr(>F)
Leaflet Number	42.63	21.315	2	4.995	0.00955
Rachis Length	2103	1051.3	2	21.17	7.88E-08
Leaflet Length	82.98	41.49	2	23.88	1.58E-08
Leaflet Width	12.57	6.284	2	21.76	5.51E-08
Total Branch number	2513	1256.3	2	6.176	0.00348
Primary Branching	101.6	50.82	2	2.843	0.0654
Secondary Branching	1039	519.6	2	5.173	0.00818
Tertiary Branching	77.2	38.6	2	2.555	0.0853
Node Angle	402.1	201.03	2	4.981	0.00966
Branch Angle	260	130.1	2	0.732	0.485
Branch Width	8.96	4.479	2	2.167	0.123
Internode Length	245.6	122.8	2	3.549	0.0344

Appendix 3.12. Levenes test for homogeneity of variances in the PNF_2.

Phenotype	Df	Fvalue	Pr(>F)
Leaflet Number	2	3.3004	0.04304
Rachis Length	2	7.2315	0.001446
Leaflet Length	2	2.8373	0.06575
Leaflet Width	2	1.3825	0.2581
Total Branch number	2	1.5266	0.2248
Primary Branching	2	0.5933	0.5554
Secondary Branching	2	2.5141	0.08866
Tertiary Branching	2	2.5554	0.08532
Node Angle	2	0.9317	0.399
Branch Angle	2	0.0498	0.9515
Branch Width	2	1.4413	0.244
Internode Length	2	0.4549	0.6365

Appendix 3.13. Tukeys HSD test results in the PNF_2 for significant traits.

Trait	Comparison	Difference	lwr	upr	P adj
Leaflet Number*	T - P	2.2058824	0.3383611	4.073404	0.0166471
	H - P	-0.2058824	-1.6399334	1.228169	0.9368348
	H - T	2.0000000	0.3697182	3.630282	0.0123549
Leaflet Length*	T - P	3.3725163	2.180814	4.56421822	0.000000
	H - P	-0.9749608	-1.890057	-0.05986473	0.0341639
	H - T	2.3975556	1.357241	3.43787032	0.0000018
Leaflet Width*	T - P	1.2808864	0.7950866	1.76668627	0.0000001
	H - P	-0.2828915	-0.6559324	0.09014938	0.1715465
	H - T	0.9979949	0.5739084	1.42208150	0.0000011
Rachis Length*	T - P	16.984473	10.614241	23.35470469	0.0000001
	H - P	-4.939577	-9.831215	-0.04793952	0.0472845
	H - T	12.044896	6.483903	17.60588903	0.0000064
Internode Length	T - P	-2.046529	-7.364016	3.270957	0.6279198
	H – P	-2.738421	-6.821666	1.344825	0.2494464
	H - T	-4.784950	-9.426933	-0.142967	0.0418916
Node Angle*	T - P	7.328441	1.5854294	13.071453	0.0088706
	H - P	-1.831141	-6.2411430	2.578861	0.5822418
	H - T	5.497300	0.4838482	10.510752	0.0283212
Secondary Branching*	T - P	-9.8676471	-18.927978	-0.8073163	0.0296262
	H - P	-0.5073529	-7.464691	6.4499851	0.9832913
	H - T	-10.375000	-18.284357	-2.4656428	0.0069395

Trait	Comparison	Difference	lwr	upr	P adj
Total Branching*	T - P	-14.08333	-26.97698	-1.189691	0.0290621
	H - P	-2.32500	-12.22590	7.575900	0.8401268
	H - T	-16.40833	-27.66404	-5.152627	0.0024236

T – *S. tetraptera* homozygotes, *P* – *S. prostrata* homozygotes, H – heterozygotes, * indicates traits displaying a significant difference between parental alleles.

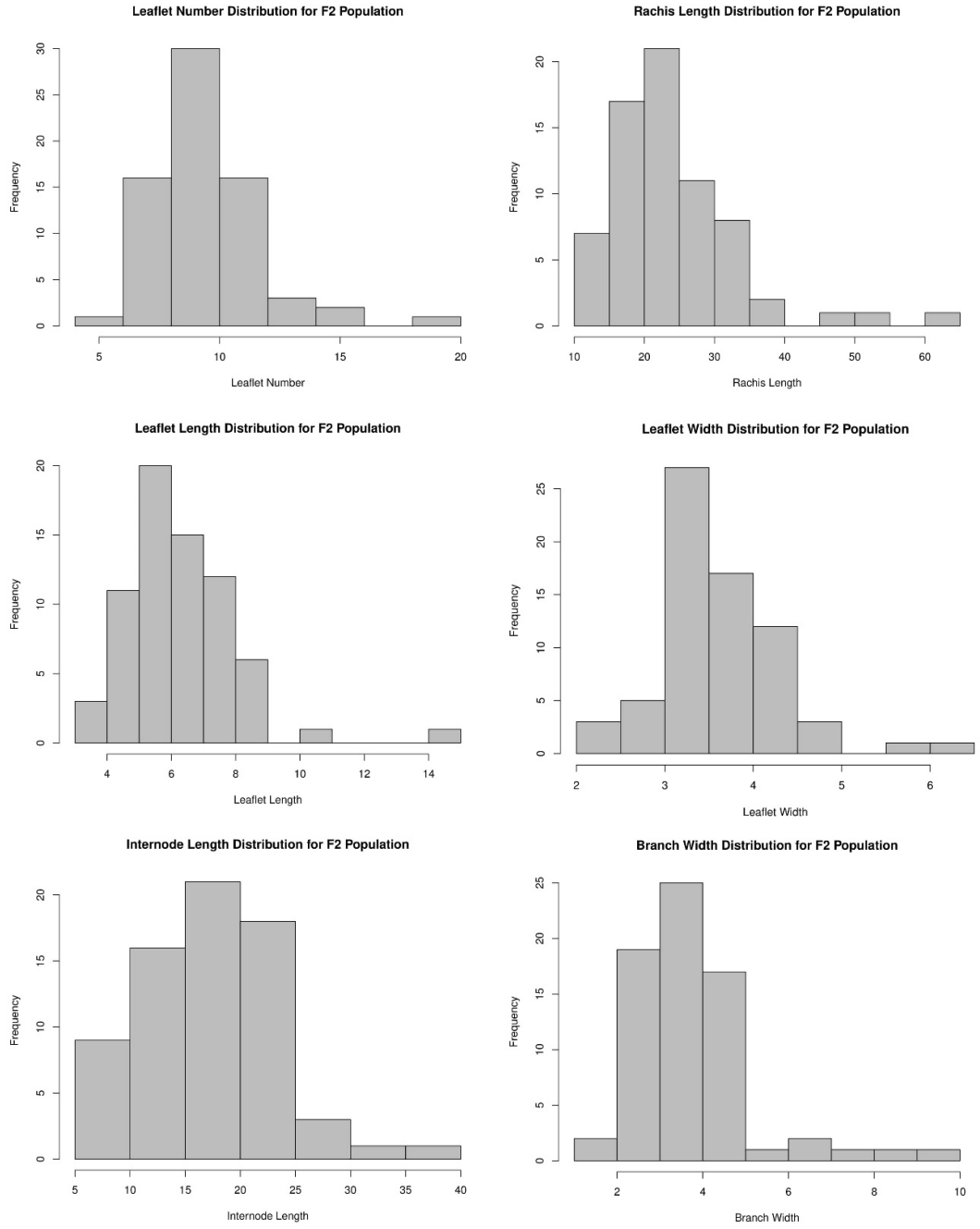
Appendix 3.14. Games-Howell post hoc test results for traits in the PNF_2 with unequal variances.

Trait	Comparison	Difference	lo	hi	t	df	P adj
Leaflet Number	T - P	2.21	-0.38	4.79	2.23	14.42	0.100
	H - P	-0.21	-1.34	0.93	0.44	37.24	0.898
	H - T	2.00	-0.55	4.55	2.07	13.29	0.100
Rachis Length*	T - P	16.98	6.55	27.42	4.30	13.00	0.002
	H - P	-4.94	-8.37	-1.51	3.53	33.05	0.003
	H - T	12.04	1.72	22.37	3.11	12.04	0.023

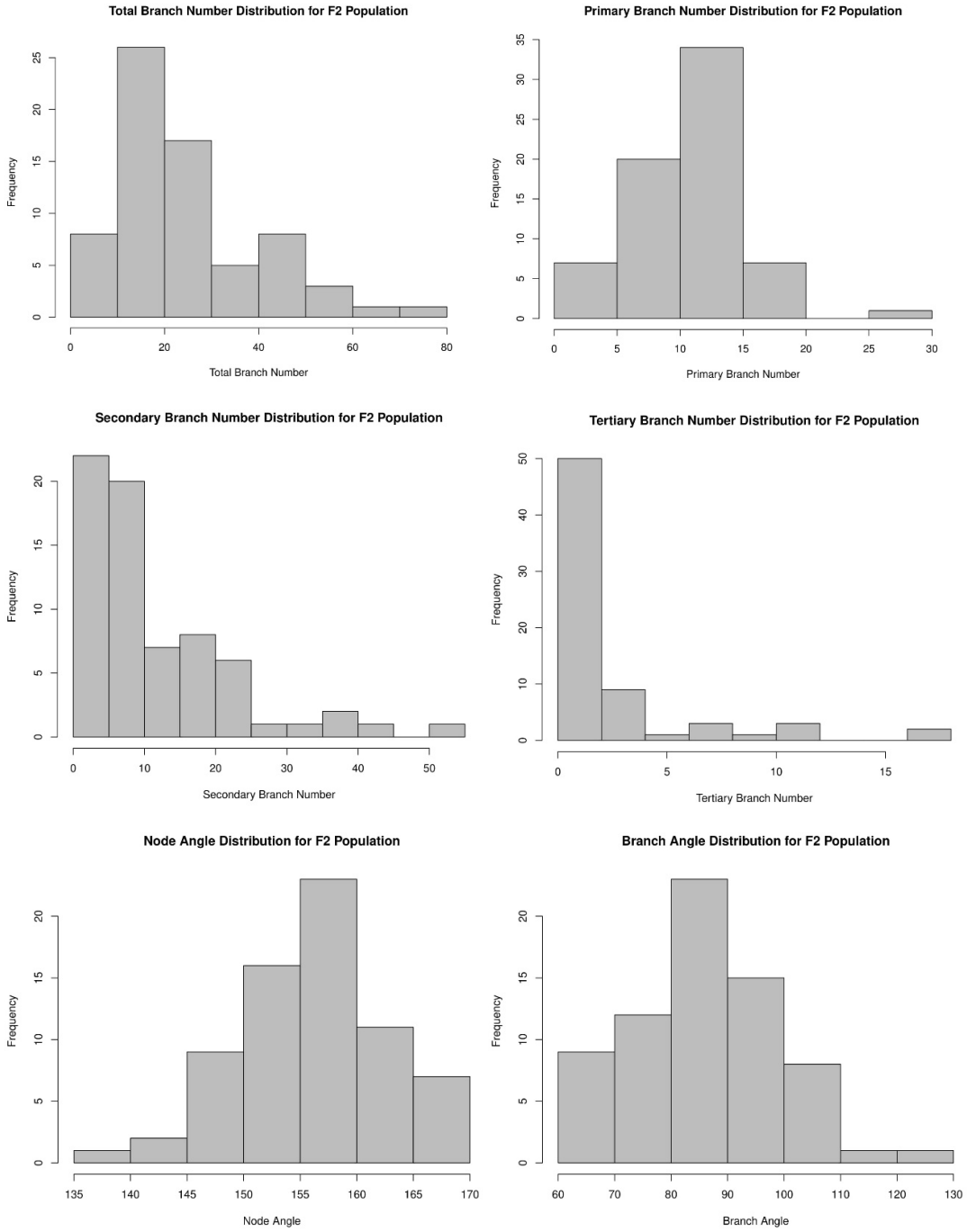
T – *S. tetraptera* homozygotes, *P* – *S. prostrata* homozygotes, H – heterozygotes, * indicates traits displaying a significant difference between parental alleles.

Appendix 3.15. *t*-test p-values for the application of synthetic strigolactone (GR24) to *S. tetraptera* and *S. prostrata* seedlings with a sample size of eight plants each.

Trait	*S. prostrata*	*S. tetraptera*
Leaflet number	0.2757	0.4561
Rachis length	0.1186	0.2667
Leaflet length	0.4317	0.07584
Leaflet width	0.3076	0.04672*
Internode length	0.9343	0.06297
Stem width	0.05596	0.4431
Branch number	0.77	0.391
Branch length	0.4503	0.391



Appendix 3.16. Histogram distributions for six of the nine traits phenotyped in the PN F_2.



Appendix 3.17. Histograms for branch number, node angle and branch angle phenotypes in the PN F_2.

4 Chapter Four – Generating *Sophora* Linkage Maps

4.1 Abstract

Linkage map construction remains a key step in the development of model organisms, allowing for identification of the genomic regions and architecture behind variation and traits and contributing to identifying the genes responsible for these phenotypes. The genus *Sophora* (Fabaceae) consists of species with different growth forms. A previously developed segregating population (Lincoln F_2) from crosses between *S. prostrata,* a divaricating species, with *S. tetraptera,* an arborescent species, provides an ideal resource for generating linkage maps. In this study, linkage maps were generated based on 143 SNP markers developed from MRMEseq and five gene markers using the *Sophora* F_2 segregating population. The SNP markers were assigned to 9 linkage groups, as expected because both *Sophora* species are 2*n*=18, and a total length of 1189.9 cM. The genetic linkage map for *Sophora* will provide a new resource for genetic research in *Sophora*.

4.2 Introduction

The generation of linkage maps has played a key role in characterizing genetic architecture and identifying genes that underlie different phenotypes and has been applied to a wide variety of research. For example, in humans to study genes causing disease (Lander & Schork, 1994; Botstein & Risch, 2003; Altshuler et al., 2008), in animal and plant breeding programs (Collard et al., 2005; Xu & Crouch, 2008; Goddard & Hayes, 2009) and in evolutionary research to identify the genetic basis of natural variation (Slate, 2005). Linkage maps can also be useful in conjunction with *de novo* genome sequencing to help identify mistakes in sequence assemblies and to assemble genome scaffolds (Fierst, 2015). A large limitation to linkage mapping is the generation of mapping populations. Typically, 50-250 individuals are often used (Mohan et al., 1997; Collard et al., 2005); however larger population sizes increase the number of recombination events, which can provide better resolution but are not always practical or possible to develop. Generating mapping populations can be time consuming and expensive (Jin et al., 2004; Fierst, 2015) especially for species that are difficult to grow in lab conditions or have long generation times such as woody plant species. Genetic mapping is often performed in F_1 populations for woody species (Gailing et al., 2009), while segregating F_2 or advanced generations are uncommon. GWAS (genome-wide association studies) can also be used in species with long generational times as a population of related individuals and linkage maps are not required, however, GWAS can suffer from lower detection power, when there are rare alleles, and difficulty identifying epistasis (Korte & Farlow, 2013). Both methods are often used in studies of plant species, particularly in crop species to identify genetic loci in order to improve breeding programs, and these can also be used together to understand the genomic basis of trait variation.

The legume family (Fabaceae) is the third largest angiosperm family with ~ 751 genera and more than 19,000 species (Lewis et al., 2005; Bruneau et al., 2013) and accounts for ~ 27% of crop production around the world (Graham & Vance, 2003). Members of the family are distributed among many of the major habitats, ranging from rainforest, alpine, lowland and savannah biomes (Schrire et al., 2005) and show a huge diversity in morphology from rainforest trees to shrubs, small herbs and even aquatic species (Lewis et al., 2005). The Fabaceae is divided into three subfamilies: Casalpinioideae, Mimisoideae and the Papilionoideae (Figure 4.1). The Papilionoideae is the largest subfamily with ~ 476 genera and 13,800 species (Lewis et al., 2005; Bruneau et al., 2013) and contains economically important species as well as the primary model species for the family (Doyle & Luckow, 2003; Bruneau et al., 2013). These model species include *Glycine max* (L.) Merr, *Medicago truncatula* Gaertn. and *Lotus japonicus* (Regel) K.Larson, which are all herbaceous species and, except for *L. japonicus,* are all annuals. Legumes often form a symbiosis with rhizobia allowing fixation of nitrogen in the soil which is an important part of natural ecosystems as well as agriculture and forestry. Legumes, such as soybean, pea, alfalfa and bean, are also important pasture and food crops and some species also have a role in industry (Graham & Vance, 2003), used to prepare biodegradable plastics (Paetau et al., 1994), dyes, oil, gums and ink (Morris, 1997). Whole genome sequences

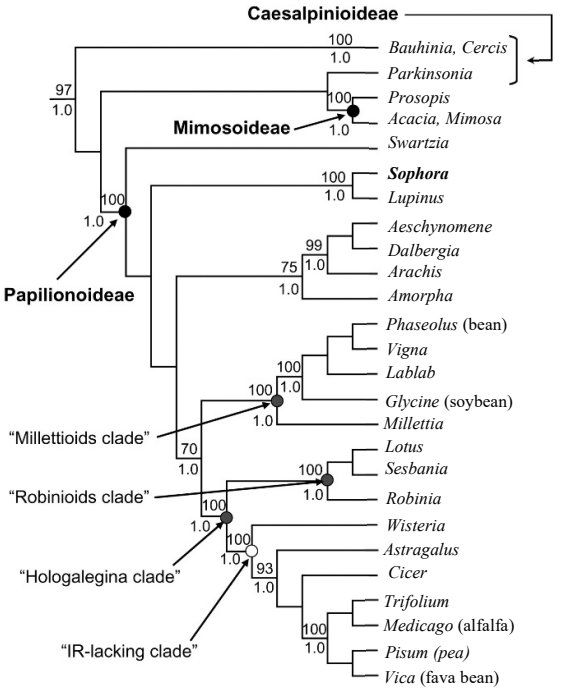


Figure 4.1. Simplified phylogeny of the Fabaceae based on maximum parsimony and Bayesian analyses of *matK* gene sequences (modified from Champagne et al (2007) and Figure 6 of Wojciechowski et al. (2004)).

are publicly available for 17 species in the Fabaceae including *Trifolium pratense, Glycine max, Lotus japonicus, Lupinus angustifolium, Medicago truncatula, Vicia faba, Vigna radiata* and *Arachis duranensis* (Chen et al., 2018). Linkage maps also have been developed for several model species (Cregan et al., 1999; Hayashi et al., 2001; Thoquet et al., 2002) and other Fabaceae species such as *Apios americana* (Singh et al., 2018), *Lupinus albus* (Simioniuc et al., 2011) and *Arachis duranensis* (Nagy et al., 2012). The majority of genetic resources in the legume family are for crop plants or model species that primarily have herbaceous annual or perennial habits; however, the Fabaceae includes a vast diversity in morphology and habits and the genomic resources only cover a small part of the diversity seen in the family.

The genus *Sophora* L. (Fabaceae) is comprised of ~ 50 woody species (Hughes, 2002) and has a relatively wide distribution including southeast Europe, southern Asia, western South America, North America and Australasia. Much of the research on *Sophora* has focused on species relationships, as seen above, or genetic diversity, for example Liu et al. (2006) and Heenan et al. (2018). There have also been a few studies performing transcriptomics on *Sophora* species, for example identifying: genes involved in biosynthesis active compounds in *S. flavescenes* Aiton. (Han et al., 2015), drought stress genes in *S. moorcroftiana* Benth ex. Baker. (Li et al., 2015) and gene expression of the *S. japonica* L. transcriptome (Zhu et al., 2014). The complete mitochondrial (Shi et al., 2018) and chloroplast (Lu et al., 2018) genomes have been sequenced for *S. japonica*. Aside from those listed above and a small set of microsatellite markers (Lee et al., 2012; Van Etten et al., 2014), there are currently no genetic or genomic resources available for the genus.

In New Zealand, eight native *Sophora* species are currently recognised; *S. fulvida* (Allan) Heenan & de Lange*, S. godleyi* Heenan & de Lange*, S. tetraptera* J. S. Mill,

S.chathamica Cockayne, *S. prostrata* Buchanan, *S. longicarinata* G. Simpson & J.S. Thomson, *S. molloyi* Heenan & de Lange and *S. microphylla* Aiton. Despite several studies (Hurr et al., 1999; Maich, 2002; Mitchell & Heenan, 2002; Heenan et al., 2004; Song, 2005; Grierson, 2014; Shepherd et al., 2017) the evolutionary relationships among these species remain undetermined. In many ways, this group exemplifies a common scenario in the New Zealand flora, with several morphologically and ecologically diverged groups with little divergence at traditionally used molecular markers. This is compounded by fairly labile reproductive barriers that lead to interspecific hybridization. While there is some evidence that *S. prostrata* is the most differentiated of the New Zealand *Sophora* species (Song, 2005; Shepherd et al., 2017), this is contradicted by another study using ISSR markers, which showed admixture among *S. prostrata, S. microphylla* and *S. tetraptera* individuals collected from natural populations (Grierson, 2014). A recent paper using microsatellite markers showed some species, such as *S. prostrata, S. chathamica, S. fulvida* and *S. longicarinata,* form distinct groups; however, *S. tetraptera, S. godleyi, S. microphylla* and *S. molloyi* did not (Heenan et al., 2018), indicating species relationships within New Zealand *Sophora* remain problematic.

New Zealand's *Sophora* demonstrate considerable variation in morphology and life history. Among the most notable differences in the group relate to growth form. While most of the species form open tree architectures, two species, *S. prostrata* and *S. microphylla,* are characterised by highly interlaced branching, an architecture termed divarication. Divarication is described as small-leaved shrubs or tree juveniles with wide branch angles, interlacing branches, long internodes and weak apical dominance (Kelly, 1994) and is observed at a relatively high frequency in the New Zealand flora (Greenwood & Atkinson, 1977). *Sophora microphylla* is further distinct as a heteroblast with a divaricating juvenile and adult tree form. New Zealand *Sophora* species also include a weeping tree, *S. godleyi,* and a shrub, *S. molloyi.*

The New Zealand *Sophora* grow in a range of environments from coastal (*S. chathamica* and *S. molloyi*), to lowland and montane (*S. longicarinata*), to forest environments (*S. microphylla*). *Sophora* species in New Zealand have pinnate compound leaves, however, there is considerable variation in the size and shape and number of leaflets among species. *S. tetraptera* is distinguished by its large leaves (100-150 mm long), while *S. prostrata* has the smallest leaves (up to 25 mm long) (Allan, 1961; New Zealand Plant Conservation Network, 2019). Leaflet shapes vary from ovate or elliptical oblong (*S. tetraptera*), to more round or elliptic (*S. molloyi*), to orbicular or obovate (*S. longicarinata*). Leaf colour varies from light green (*S. chathamica*), to grey-green (*S. godleyi*), to dark green (*S. molloyi*). Leaflet density also varies from crowded (*S. chathamica*) to distant (*S. microphylla*). Most species have straight appressed leaf hairs, however, *S. godleyi* and *S. fulvida* both display variation in leaf hairs, which can be decumbent, sometimes spreading, curved or twisted. Flowering time also differs among species with some species flowering October to January (e.g., *S. longicarinata*) and others August to October (e.g., *S. microphylla*) (Heenan et al., 2001). Flower colour is usually yellow in all New Zealand species, but can also be orange in *S. prostrata*.

The origin of the heteroblast, *S. microphylla,* was of interest to Godley (1985) who investigated the hypothesis that *S. microphylla* was formed from hybridization between *S. prostrata* and *S. tetraptera* by creating F_2 populations formed from reciprocal crosses between these two species. This work was carried out in the 1980's and the plants have sat idle for several decades. Upon investigation, it was discovered that they are segregating for flower colour, flowering time, leaf and leaflet size, branching architecture, and more. To develop this rare woody plant resource into a robust tool for genetic work, we have developed molecular markers for this group using next-generation sequencing and assembled the first linkage groups for the genus. This chapter reports on

the generation of markers and linkage maps for the *Sophora* F_2 population for later use in quantitative trait locus (QTL) analysis.

4.3 Materials and Methods

4.3.1 Study Population

Development of the study population started approximately 45 years ago, by Dr. E J Godley, and was initiated with reciprocal crosses between *S. tetraptera* and *S. prostrata*. While a single *S. tetraptera* individual, G3497/6, from Frasertown was used for both cross directions, two *S. prostrata* individuals were used; a plant from Rakaia Gorge, G4928/29 was used as a paternal parent and a plant from Waitahi, G1916 was used as a maternal parent (Figure 4.2). Four F_1 plants, two from each cross, were selfed to create the F_2 populations, A, B, C and D (Figure 4.2). These were planted, on 20 May 1985, at the 'East Block' site outside Lincoln, New Zealand. The Lincoln population is comprised of the remaining parental plants and the F_2 plants arrayed in four blocks, A-D, according to lineage (see Fig. 4.2) and are situated alongside a water race. The F_1 plants are located at a different site in the town of Lincoln. The exact individuals that were used as the

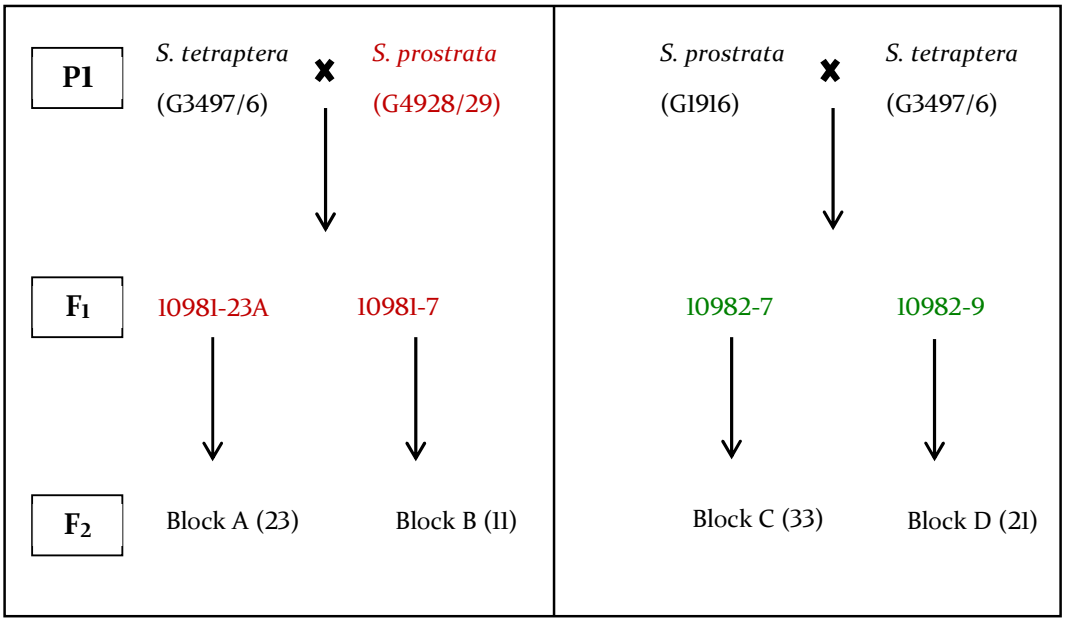


Figure 4.2. Schematic of the crosses used to produce the F_2 blocks; A, B C and D. Plants in red are no longer present so could not be used in sampling. Plants in green are still present and have been sampled. For the parental plants (in black) some plants are still present at the site from the original collection sites and these have been sampled in place of the parental plants. F1 were produced from two reciprocal crosses with different *S. prostrata* used in each cross and a different species as the maternal parent in each cross. Four different F1s were used to produce F2 in each of the four blocks A, B, C and D.

parental plants are not known and some of the original plants are no longer present at the site. However, plants collected from the original source sites are still present to represent the parents and were sampled in place of the parental plants. *Sophora prostrata* from Rakaia Gorge (G4928/29) is the only parent that had no representation remaining at the site. Two F_1 individuals, 10981-23A and 10981-7, produced F_2 blocks A and B, respectively. These two F_1 plants are no longer present at the study site, but two F_1 plants from the same cross, 10981-3 and 10981-18 were sampled as representatives of this cross direction. Individuals 10982-7 and 10982-9 produced the C and D F_2 blocks, respectively; these are both still present and were included in our sampling. Since the F_2 population was established, the site has not always been maintained resulting in overgrowth. Two years prior to our data collection, the site was cleared of all growth other than the F_2 plants. Following a survey of the site, the total remaining number of F_2 plants was 88 with 34 plants from blocks A and B (*S. tetraptera* maternal parent) and 54 plants from blocks C and D (*S. prostrata* maternal parent). The total number of plants per block was: A = 23, B = 11, C = 33 and D = 21.

4.3.2 Collecting samples for DNA extraction

Leaf tissue was collected from each of the 88 F_2 individuals, the four representative F_1 and the four representative parents and dried in silica gel for desiccation and storage. DNAs were extracted from each individual using the Qiagen DNEasy plant mini kit (Qiagen) as per the manufacturer's instructions. DNA, RNA and protein concentration were quantified on a Qubit (Thermo Fisher Scientific) at the Massey Genome Service. All samples were diluted to 10 ng/µl before PCR amplification.

4.3.3 MRMEseq (Modified Random Marker Enrichment Sequencing) – Pilot study

4.3.3.1 Primer pair testing

The MRMEseq protocol is based on the procedure developed by McLay (McLay, 2018), which involves PCR with Random Amplified Polymorphic DNA (RAPD) primers as a method to reduce genome complexity. Eight *Sophora* gDNA samples were initially used to test 66 RAPD primer pair combinations (Figure 4.3). Primers were ordered from IDT (Integrated DNA Technologies, Inc.) and include sequence tails for later indexing. PCR amplification was performed in a 10 µL volume with 1x HF Phusion buffer (Thermo Fisher Scientific), 50 µM of each DNTP, 0.5 µM of forward primer, 0.5 µM of reverse primer, 3% DMSO and 1.0 unit/50 µL Phusion Taq (Thermo Fisher Scientific). Amplification by PCR used: 95°C for 1 minute; 25 cycles of 95°C for 1 minute, 50°C for 90 seconds, ramped at 0.2°C/sec to 72°C for 1 minute, followed by 72°C for 5 minutes. Each PCR was run on a 1% agarose gel at 75 volts for 50 minutes and stained with ethidium bromide for UV visualisation to assess consistency and intensity. Based on those results, sixteen DNA samples were then screened with 16 primer pair combinations from the initial screen that had amplified consistently with multiple bands, to find the most suitable combinations for sequencing. From these, the primer pair combinations that produced multiple consistent bands of similar intensity were used for creation of the sequencing library. Ten of the 16 primer

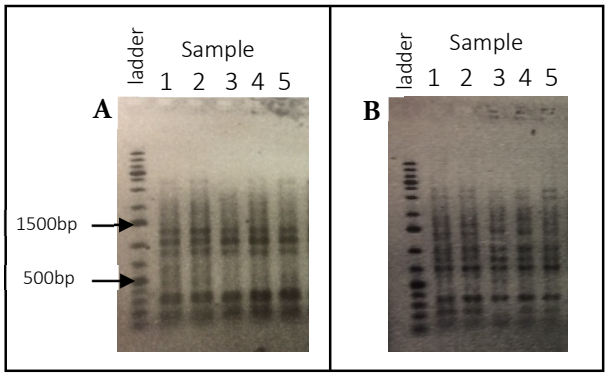


Figure 4.3. Example RAPD gels. Primer pair F14 and R14 (A). Primer pair F19 and R14 (B). Ladder indicates the size standard Generuler 1kb Plus DNA ladder from Thermo Scientific (Thermo Fisher Scientific).

pairs (Table 4.1) were identified that fit these criteria and these ten were used on the sixteen pilot samples (Table 4.2).

Table 4.1. RAPD primer combinations used for Illumina sequencing

Primer	Primer sequence	Tm (°C)	GC content (%)
F10c	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**CCCGTTGCCTC**	70.4	56.8
F11	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**ACATCGCCCAG**	69.3	54.5
F13	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GATGACCGCCACA**	70.6	54.3
F14	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GGACAACGAGACTG**	69.3	53.2
F15	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AAGCGACCTGTCAGC**	70.0	54.2
F18	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AACGGTGACCGTACGTGC**	70.7	54.9
F19	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**ACCCGGTCACTGCGAGTCC**	71.7	57.7
R10b	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**CCCGTTGCCTG**	70.5	57.8
R12	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**TTCCGAACCCAC**	69.3	54.3
R14	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**TGCGCCCTTCCTG**	70.9	57.4
R17	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**GGAAGCCAACAGCTGC**	71	56
R18	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**GGAAGCCAACGACGTGG**	71.2	56.9
R19	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**CCTTGACGCACCGATGAC**	70.7	55.8

Primer combinations used : F10c + R10b, F10c + R1[6], F11 + R14[6], F13 + R12, F14 + R14[3], F14 +R17[6], F14 + R19[6,3], F15 + R10b[6], F18 + R18[6,3], F19 + R14. [6] indicates primer pairs used in multiplex with 6 primer pairs and [3] indicates primer pairs used in multiplex with 3 primer pairs. RAPD sequences indicated in bold, remaining sequence is the primer tail used to attach index and illumina recognition sequence in subsequent PCR.

Table 4.2. List of *Sophora* samples used in pilot study

	Sample	Species or origin	Location	Maternal parent
1	St-L*	*S. tetraptera*	Frasertown	-
2	St-R*	*S. tetraptera*	Frasertown	-
3	SP1916*	*S. prostrata* (G1916)	Waitahi	-
4	Spwai	*S. prostrata*	Waiau	-
5	F_1 1-3*	*S. tetraptera* (G3497/6) x *S. prostrata* (G4928/29)[R]		*S. tetraptera*
6	F_1 1-18*	*S. tetraptera* (G3497/6) x *S. prostrata* (G4928/29)[R]		*S. tetraptera*
7	F_1 2-7 (10982-7)[1]	*S. prostrata* (G1916) x *S. tetraptera (G3497/6)*		*S. prostrata*
8	F_1 2-9 (10982-9)[1]	*S. prostrata* (G1916) x *S. tetraptera (G3497/6)*		*S. prostrata*
9	F_2 A2	10981-23A		*S. tetraptera*
10	F_2 A4	10981-23A		*S. tetraptera*
11	F_2 A10	10981-23A		*S. tetraptera*
12	F_2 A19	10981-23A		*S. tetraptera*
13	F_2 D11	10982-9		*S. prostrata*
14	F_2 D30	10982-9		*S. prostrata*
15	F_2 D31	10982-9		*S. prostrata*
16	F_2 D35	10982-9		*S. prostrata*

* Denotes surviving original parents or F_1, [1] denotes F_1 that are no longer present at site, [R] denotes *S. prostrata* from Rakaia George (no representative of this parent remains).

4.3.3.2 Multiplex PCR trial

Multiplex PCRs were also performed with the MRMEseq primers that would reduce the PCR steps needed to generate the individual libraries. These included two samples with a pool of six primer pairs and two samples with three primer pairs pooled as indicated in Table 4.1. The two samples included a representative of each parental species, *S. tetraptera* (G3497/6) and *S. prostrata* (G1916). PCR amplification was performed in a 20 µL volume with 1x HF Phusion buffer (Thermo Fisher Scientific), 50 µM of each dNTP, 0.5 µM of forward primers, 0.5 µM of reverse primers, 3% DMSO and 1.0 unit/50 µL Phusion Taq (Thermo Fisher Scientific). Amplification by PCR was obtained by: 95°C for 1 minute, then 25 cycles of 95°C for 1 minute, 50°C for 90 seconds, ramp at 0.2°C/sec to 72°C for 1 minute, followed by 72°C for 5 minutes.

4.3.3.3 Bead clean-up and Index PCR

To remove primer dimers and small PCR products a bead clean-up approach was used using Agencourt AMPure XP beads (Beckman Coulter Life Sciences). A range of bead:DNA ratios were tested (0.6x, 0.8x, 1.0x and 1.2x) on 5 µl PCR product in 45 µl water for a consistently amplified single RAPD PCR. The aim was to remove the smallest fragments, below 200 bp, which would be preferentially sequenced by Illumina sequencing (Illumina Inc., 2019). The bead-cleaned samples were run on a gel to determine the best ratio. The 0.8x (Figure 4.4) ratio was found to be the most efficient for clean-up and size selection, as it removed the smaller fragments but still retained the intensity of the rest of the fragments. The 0.8x bead-clean ratio was then tested on

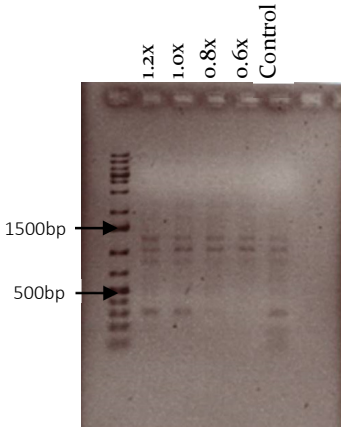


Figure 4.4. Gel photo of the different bead ratios tested. In the 0.8x and 0.6x ratios the smaller bands have been lost leaving only the bands of target size. Size standard is the Generuler 1kb Plus DNA ladder from Thermo Scientific.

pooled samples including: two pooled samples of 10 RAPD PCRs and two pooled samples of five RAPD PCRs to ensure 0.8x was still effective on pooled samples. All bead-clean tests were also run on the bioanalyser to determine the size range of products. Bead-clean was performed at 0.8x for the main sampling. For each individual, the products from the 10 RAPD PCRs were first pooled at 5 µL and then cleaned using Agencourt AMPure XP beads (Beckman Coulter Life Sciences). After bead clean-up, each individual's pool (library) was eluted in 40 µL of Tris-HCl pH 8.

A second round of PCR was performed for two purposes: (1) to add unique sequence IDs (indexing sequences) to each pool for later bioinformatic processing and (2) to add the sequences required for Illumina sequencing. These goals were accomplished through a single round of PCR that used primer combinations that match the sequence tails from the first round of PCR (Figure 4.6). Index sequences were different for each sample, allowing the assignment of each sequence back to individuals. PCR amplification was performed in a 20 µL volume with 1x HF Phusion buffer (Thermo Fisher Scientific), 50 µM of each dNTP, 0.5 µM of forward primer, 0.5 µM of reverse primer, 3% DMSO and 1.0unit/50 µL Phusion Taq (Thermo Fisher Scientific) and 3 µL of template. Amplification by PCR was obtained by: 95°C for 3 minutes; 5 cycles of 95°C for 30 seconds, 62°C for 30 seconds, 72°C for 1 minute; followed by 72°C for 5 minutes. Products were run on a 1% agarose gel to check amplification. Each index PCR product was cleaned up using 0.8x total volume of Ampure beads in a 50 µL volume (15 µl PCR, 35 µL water) and eluted in 25 µL Tris HCl pH 8. Each of these libraries was quantified using a Qubit at the Massey Genome Service and pooled into a single library at a final volume of 197.1 µl, which was bead-cleaned and eluted to a final volume of 40 µl and DNA concentration of ~ 41.5 ng/µl.

4.3.3.4 Pilot study

To test the MRMEseq method (Figure 4.5), the pilot library included 17 samples (with one duplicate sample) including two *S. prostrata* plants, one as the best representative of the parent and the other from a different location, and two *S. tetraptera* (unknown if either are the parent but they are from the same source population) (Table 4.2). To identify the markers derived from a particular primer pair, individual libraries were also developed for each primer pair using the known *S. prostrata* parent. Additionally, to make the creation of libraries easier, multiplex PCR libraries were also tested. The final library consisted of: 8 ng of each of the 10 single RAPD primer libraries from *S. prostrata*, 80 ng of each of the pooled sample libraries, 30 ng of each of the two six primer pair multiplex libraries, and 80 ng of each of the three primer pair multiplex libraries. This final large pool was cleaned and concentrated with 0.8x beads and eluted in 40 µL Tris-HCl. This was quantified using the Qubit at 23.9 ng/µL and run on the 2100 Bioanalyser for quality control checks by the Massey Genome service.

4.3.3.5 Sequencing and marker identification

Sequencing was performed on the Illumina platform with a full 2 x 250 bp paired-end MiSeq run at the Massey Genome Service and reads were processed, including demultiplexing, by the Massey Genome Service. Demultiplexed sequences were processed and further analysed using ipyRAD v0.7.28 (Eaton, 2014; Eaton, 2015) in Python v. 3.6.5 (Python Software Foundation, 2001) to identify SNPs suitable for generating linkage maps.

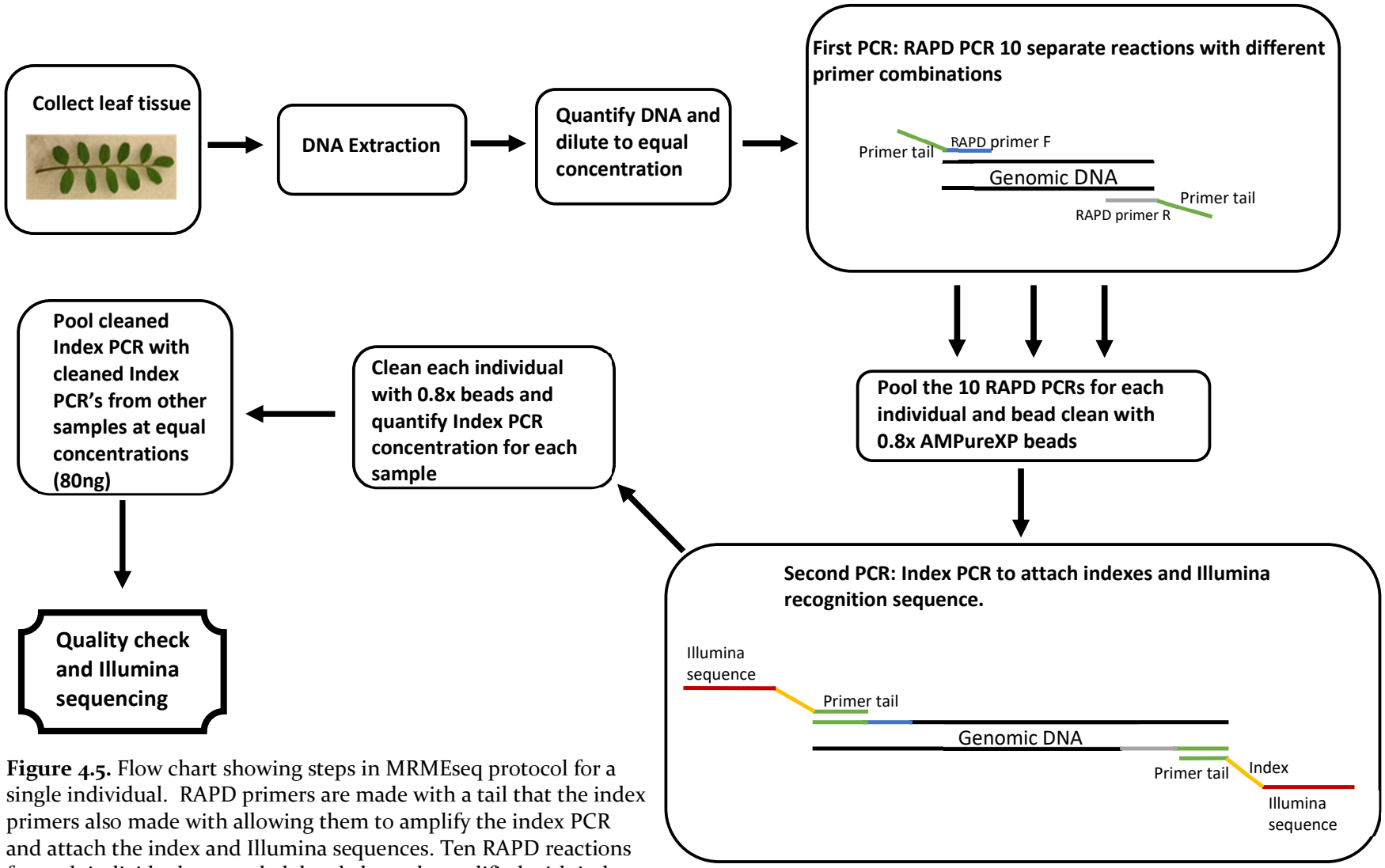


Figure 4.5. Flow chart showing steps in MRMEseq protocol for a single individual. RAPD primers are made with a tail that the index primers also made with allowing them to amplify the index PCR and attach the index and Illumina sequences. Ten RAPD reactions for each individual are pooled, bead cleaned, amplified with index and illumina sequences and bead cleaned again. Then each individual library (with its own index) is pooled equimolar into a single library for sequencing.

4.3.4 Full F_2 population genotyping

4.3.4.1 MRMEseq

Based on results from the pilot study, MRMEseq (Figure 4.5) was performed for the full complement of 88 F_2 individuals, four representative F_1, and two of each parent species. For each sample, DNA was diluted to a concentration of 9 ng/µl and amplified by the 10 RAPD primer pairs, identified in the pilot study (Table 4.1). PCR amplification was performed in a 10 µL volume with 1x HF Phusion buffer (Thermo Fisher Scientific), 50 µM of each DNTP, 0.5 µM of forward primer, 0.5 µM of reverse primer, 3% DMSO and 1.0 unit/50 µL Phusion Taq (Thermo Fisher Scientific). Amplification by PCR was obtained by: 95°C for 1 minute, then 25 cycles of 95°C for 1 minute, 50°C for 90 seconds, ramp at 0.2°C/sec to 72°C for 1 minute, followed by 72°C for 5 minutes. Each sample was run on a 1% agarose gel at 75 volts for 50 minutes to confirm amplification of bands. Any failed reactions were re-run.

For each individual, 4 µL of each of the 10 PCR products were pooled and then cleaned up using Agencourt AMPure XP beads (Beckman Coulter Life Sciences) at 0.8x. A separate clean-up for each primer-pair combination and multiplex PCR was done for each individual. After bead clean-up, samples were eluted in 40 µL of Tris-HCl pH 8. The index PCR to attach Illumina sequencing primers and index tags was performed in 20 µL reactions with 1x HF Phusion buffer (Thermo Fisher Scientific), 50 µM of each DNTP, 0.5 µM of forward primer, 0.5 µM of reverse primer, 3% DMSO and 1.0 unit/50 µL Phusion Taq (Thermo Fisher Scientific) and 3 µL of template. Amplification by PCR was obtained by: 95°C for 2 minutes; 8 cycles of 95°C for 10 seconds, 70°C for 60 seconds, 72°C for 2 minutes; followed by 72°C for 5 minutes. Products were run on a 1% agarose gel to check amplification. Index PCRs were cleaned up using 0.8x total volume of beads in 30 µL volume (25 µl PCR, 5 µL water) and eluted in 25 µL Tris HCl pH 8. All bead-cleaned

samples were quantified on the Qubit (Thermo Fisher Scientific). Eighty nanograms of each sample was pooled into a single tube, which was then checked on a 1% agarose gel. The final pool was diluted to 10 nM and run on a PerkinElmer LabChip® GX Touch HT at the Massey Genome Service. Sequencing was performed on the Illumina platform in a single, full HiSeq 2500 single-end run at the Otago Genomics and Bioinformatics Facility. Initial sequence QC and demultiplexing were carried out by the facility.

4.3.4.2 Marker Identification

The ipyRAD v. 0.7.28 (Eaton, 2014; Eaton, 2015) package in Python v. 3.6.5 (Python Software Foundation, 2001) was used to process the HiSeq sequences. Sequences were processed as the 'rad' data type with a minimum sequence depth of 6, minimum of 20 samples per locus, maximum low quality bases of 10, maximum depth of 100,000, maximum heterozygous bases in consensus at 10 and maximum indels in final locus at 20 with all other parameters using default settings (Appendix 4.1). The vcf file generated was filtered using BrowseVCF v. 2.8 (Salatino & Ramraj, 2017) to identify SNPs that would be suitable for generating linkage maps; specifically, those SNPs where each allele can be traced back to just one parent species and segregate in the F_2. BrowseVCF was first used to search for SNPs that were heterozygous in the F_1 and were homozygous in the parents for alternative alleles. Less strict criteria searched for heterozygous F_1 and only one parent homozygous or alternative homozygous parents and missing F_1 data. The data from each criterion were combined into a single excel file and the number of individuals for each genotype was calculated for the F_2 to identify SNPs that were not segregating; for example, SNPs that only had one or two genotypes in the F_2 or very low numbers of one genotype compared to the other two. This removed markers with large amounts of missing data and those not displaying expected segregation in the F_2. Markers with more than 25% missing data were also discarded. Loci that remained after

this filtering were converted into file format to use in R/qtl. SNP coding was changed using the search and replace feature of excel from 0/0, 1/1 and 0/1 notation from the vcf format into H for heterozygotes, A representing *S. prostrata* allele and, B for *S. tetraptera* allele using the SNP notation in the parent samples for that specific SNP. Missing data were represented by a dash (-). The development of markers from polymorphisms embedded within candidate genes was described in Chapter 3. These were also coded into the A, B, H notation for the file format for R/qtl.

4.3.5 Generating linkage maps

Linkage maps were generated using R v. 3.4.2 and the package R/qtl v. 1.42-8 (Broman et al., 2003) following the procedure described by Broman (2012). Two individuals were excluded from the generation of the map as they had over 90% matching genotypes with another individual as identified by R/qtl and may indicate duplicate samples. Distorted segregation was detected by a chi-square test (X^2) at 5% after Bonferroni correction. LOD score thresholds for the linkage of markers were compared ranging from a LOD of 4 to 10 and a LOD score of 6 was finally used to generate the linkage maps. The requirements for placing markers in the same linkage groups were a maximum recombination fraction of 0.35 and a minimum LOD score of 6 as this gave the expected number of linkage groups based on chromosome number (n=9) (Dawson, 2000; Stiefkens et al., 2003), whereas other LOD scores split linkage groups into multiple smaller groups or combined many linkage groups into one larger linkage group. Map distances were calculated using the Kosambi mapping function (Kosambi, 1944). Markers displaying segregation distortion were initially removed from the map generation to prevent errors in marker order from their inclusion but were later included in generating the map as their inclusion did not affect marker order. Maps were generated in the same way for both sets of markers. Linkage maps were also generated

for individuals for each of the reciprocal cross directions, however, this reduced the populations to smaller sizes below or close to the lower limit suggested for population size for mapping. Maps were generated with the same parameters as the larger population. The final map figure was created with the R package LinkageMapView v. 2.1.2 (Ouellette et al., 2018).

4.4 Results

4.4.1 MRMEseq Pilot Study

Sixty-six primer pairs were initially tested for multiple consistent bands with similar intensity across individuals (Figure 4.6). Of these 66, the 10 primer pairs (Table 4.1) that best fit these criteria were chosen for analysis. A 0.8x bead:volume ratio was determined the best ratio for size selection/sample clean-up (Figure 4.4). Bioanalyser results (Figure 4.7) showed a reduction in the frequency of smaller products (e.g., primer

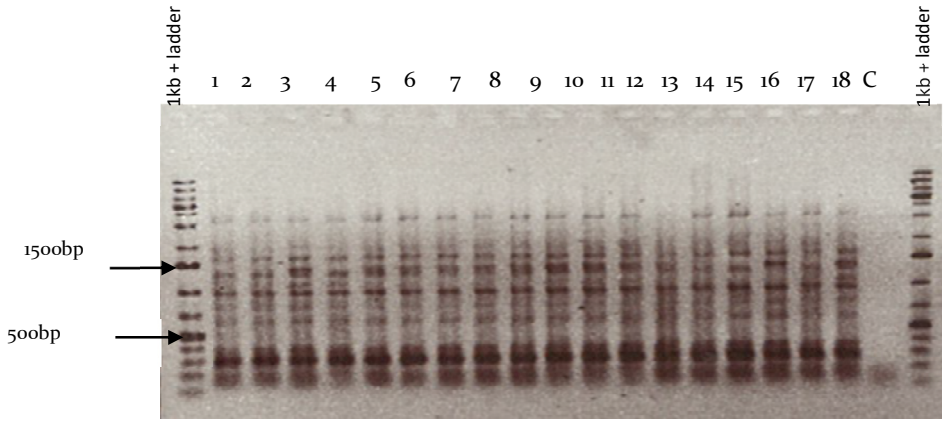


Figure 4.6. Example of the products from RAPD primers F14 and R14. Size standard is the Thermo Scientific Generuler 1kb Plus DNA ladder (Thermo Fisher Scientific)

dimers and products smaller than 200 bp) after clean up but still retained the targeted products larger than 200 bp. To ensure the procedure worked and markers could be developed for QTL analysis, a pilot study was conducted with seventeen samples (Table 4.2) including samples from the individual primer pairs and multiplex PCRs. Multiplex PCRs, using three and six pooled primer pairs, were trialled to test if this was a suitable way to perform the initial RAPD step. Multiplex PCR of the RAPD step would reduce the time and steps needed to create libraries as the initial PCR of each primer could be combined in one step rather than a PCR run for each primer pair and subsequent pooling. The final library had a DNA concentration of 21 µg/mL. The total number of sequences generated from the pilot Illumina Miseq run was 5,581,076 (Table

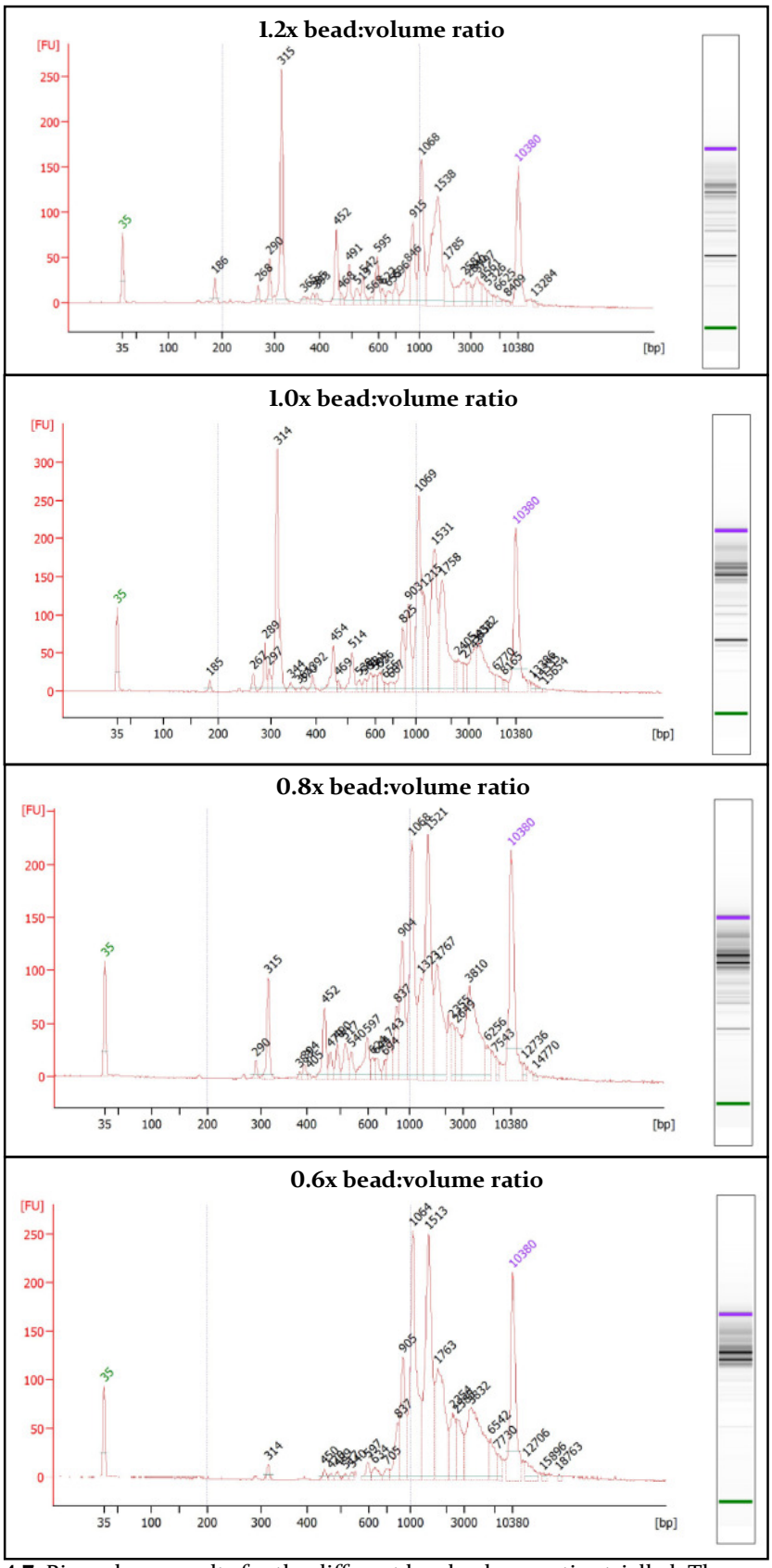


Figure 4.7. Bioanalyser results for the different bead:volume ratios trialled. The graphs show the 0.8x bead:volume had the best decrease in products smaller than 200bp but retained the larger products. The 0.6x ratio had greater loss of larger products than 0.8x. The green and purple numbers correspond to the ladders used. Graphs also show the range of products obtained using the RAPD primer pairs.

4.3). The number of sequences from each library ranged from 7,736 to 353,402 (Table 4.3) and was correlated with the amount of DNA pooled (Figure 4.8). Analysis on multiplex data and single libraries have not been performed.

From these sequences, polymorphisms were identified, using ipyrad and BrowseVCF, as potential markers used to generate linkage maps. The ideal loci are those where the parental species are fixed for alternative alleles so that F_1s will be heterozygous and the F_2 population will be segregating for individuals either homozygous for one of the parental alleles or heterozygous (Figure 4.9). Those preliminary results identified ~ 100 SNPs that matched our criteria.

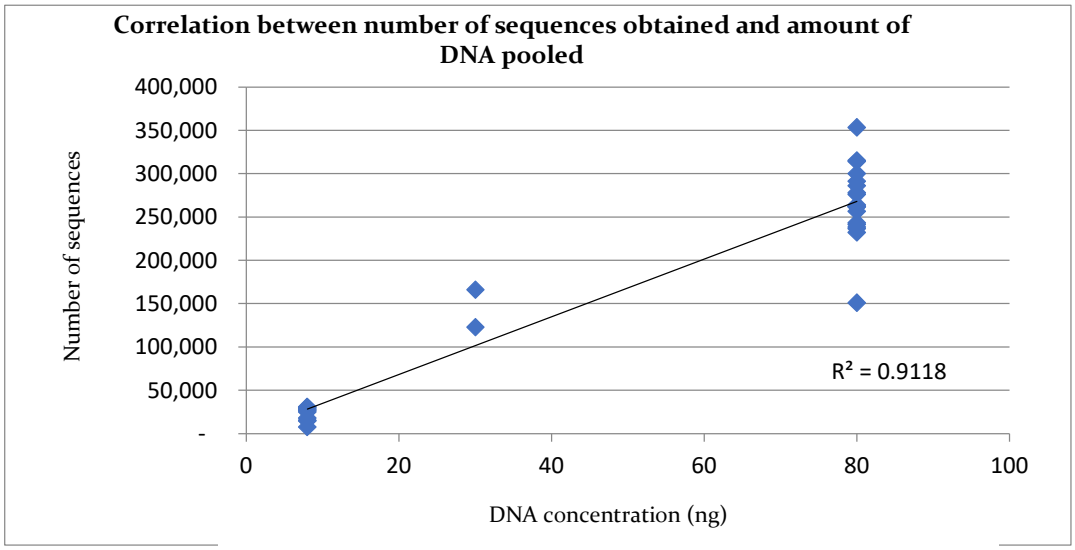


Figure 4.8. Scatterplot displaying the correlation between the number of sequences from the MiSeq run and the concentration of DNA in the final library. The R^2 value is 0.9118. The higher the DNA concentration in the final library, the more sequences were obtained.

4.4.2 Generating linkage maps for *Sophora* F_2

The MRMEseq protocol was completed with the 88 *Sophora* F_2, as the pilot suggested this protocol is suitable to generate markers for linkage map development. Amplification of the RAPD and index primers, previously tested in the pilot, were successful for all 88 F_2 samples. As was observed in the pilot, the 0.8x bead clean was successful at preventing fragments below 200 bp being in the final pool (Figure 4.9).

The concentration of the final pool for sequencing was 15.5 ng/µL. This was diluted to 10 ng/µL as recommended for sequencing. The MRMEseq protocol for the *Sophora* F_2 returned 165,588,214 raw sequences and 147,033,225 sequences with a yield of 35,758 Mbases after initial processing. The percentage equal to or greater than Q30 bases was

Table 4.3. Results of the Miseq run for the pilot *Sophora* libraries.

Number	Sample	Forward Index primer	Reverse Index Primer	Amount pooled (ng)	# sequences
1	St-L	mpxPE1.bc001	mpxPE2.bc001	80	261,791
2	St-R	mpxPE1.bc001	mpxPE2.bc002	80	315,330
3	SP1916	mpxPE1.bc003	mpxPE2.bc003	80	264,091
4	Spwai	mpxPE1.bc004	mpxPE2.bc004	80	236,451
5	F1 1-3	mpxPE1.bc005	mpxPE2.bc005	80	232,284
6	F1 1-18	mpxPE1.bc006	mpxPE2.bc006	80	262,190
7	F1 2-7	mpxPE1.bc007	mpxPE2.bc007	80	241,494
8	F1 2-9	mpxPE1.bc008	mpxPE2.bc008	80	278,155
9	F2A2	mpxPE1.bc003	mpxPE2.bc001	80	353,402
10	F2A4	mpxPE1.bc004	mpxPE2.bc002	80	262,309
11	F2A10	mpxPE1.bc005	mpxPE2.bc003	80	238,777
12	F2A19	mpxPE1.bc006	mpxPE2.bc004	80	276,008
13	F2D11	mpxPE1.bc007	mpxPE2.bc005	80	243,466
14	F2D30	mpxPE1.bc008	mpxPE2.bc006	80	299,957
15	F2D31	mpxPE1.bc009	mpxPE2.bc007	80	151,276
16	F2D35	mpxPE1.bc0010	mpxPE2.bc008	80	286,016
17	St-L dup	mpxPE1.bc005	mpxPE2.bc001	80	314,470
18	F10C + R12	mpxPE1.bc006	mpxPE2.bc002	8	28,993
19	F14 + R17	mpxPE1.bc007	mpxPE2.bc003	8	14,781
20	F15 + R10b	mpxPE1.bc008	mpxPE2.bc004	8	30,501
21	F11 + R14	mpxPE1.bc009	mpxPE2.bc005	8	25,114
22	F10c + R10b	mpxPE1.bc0010	mpxPE2.bc006	8	7,736
23	F19 + R14	mpxPE1.bc0011	mpxPE2.bc007	8	18,289
24	F14 + R19	mpxPE1.bc0012	mpxPE2.bc008	8	15,390
25	F14 + R14	mpxPE1.bc009	mpxPE2.bc001	8	27,753
26	F13 + R12	mpxPE1.bc0010	mpxPE2.bc002	8	30,991
27	F18 + R18	mpxPE1.bc0011	mpxPE2.bc003	8	26,621
28	St-L 6 pool	mpxPE1.bc0012	mpxPE2.bc004	30	166,256
29	SP1916 6 pool	mpxPE1.bc003	mpxPE2.bc005	30	122,957
30	St- L 3 pool	mpxPE1.bc004	mpxPE2.bc006	80	256,695
31	SP1916 3 pool	mpxPE1.bc001	mpxPE2.bc007	80	291,532
			Total	1660	5,581,076

Numbers 1-17 – pool of 10 RAPD PCRs for each sample, 18-27 – single RAPD PCR for one sample, 28 – 31 – multiplex PCRs. Primer names indicate index primers. All 31 samples pooled prior to sequencing.

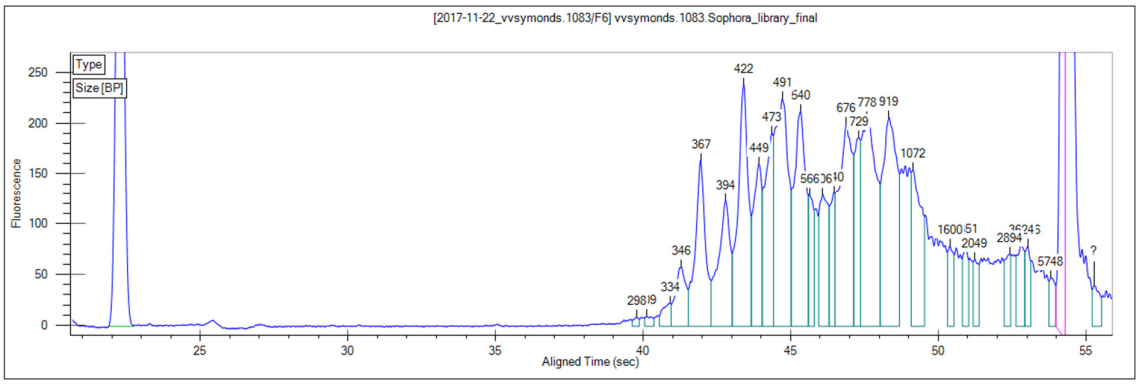


Figure 4.9. Results from the PerkinElmer LabChip® GX Touch HT for the final pool of the MRMEseq protocol for the full 88 samples. After cleaning with 0.8x beads. Figure showing the small fragments below 200bp, which can be preferentially sequenced, are not present in the pool. The large peaks off scale (~22bp and 55bp) are size standards.

91.43% with a mean quality score of 35.2 for all samples (Appendix 4.2) and GC content ranged from 47-50% for each sample. Most samples produced similar numbers of sequences, as expected by pooling an equal concentration, four individual samples had less than one million sequences, the lowest at 696,424 sequences, and two had over two million, the highest with 2,105,505 sequences. Analysis of these sequences in ipyRAD (Eaton, 2014; Eaton, 2015) identified a total of 36,178 SNPs from the sequences. The ideal markers for generating linkage maps are those that are heterozygous in the F_1, homozygous for alternative alleles in the parents and that segregate in the F_2 (Figure 4.10). BrowseVCF (Salatino & Ramraj, 2017) was used to identify SNPs between the parents; 3,170 possible SNPs were identified. After removing markers showing a high degree of missing data and distorted segregation (where one allele combination was not present or at a very low rate) 160 markers remained. From these 160 markers duplicate loci were removed so only one marker from each locus was used in the linkage map leaving 143 total markers. Ten of these markers were identified as having distorted segregation but still showed segregation among the three possible genotypes. As well as the MRMEseq markers, five additional markers were developed from candidate genes (see Chapter 3) identified from phenotypic mutants in *Pisum sativum* (Fabaceae) and

other model plants. These mutants displayed phenotypes similar to the divaricate traits such as the RAMOSUS (RMS) genes, *RMS1, RMS2, RMS4, RMS5* and the gene *ZIG*. These were developed into markers based on SNPs between *S. prostrata* and *S. tetraptera* sequences as described in Chapter 3.

R/qtl was used to generate the linkage maps using the genotype data for all 148 markers. Maps were generated first without the markers displaying distorted segregation and then including them to ensure these markers were not affecting the marker order. Maps were also generated for each of the reciprocal cross directions in the F_2 population, however, the sample sizes of these, especially the AB population, were small which made mapping in the AB population difficult. The CD population genetic map was similar to the full F_2 map with some differences including linkage group 5 split into two linkage groups, changes to marker order within some linkage groups and one marker that did not link to any other markers. The final linkage map for the *Sophora* F_2 (Figure 4.11) consisted of nine linkage groups and a total length of 1189.9 cM (Table 4.4). The lengths of the linkage groups varied from 72 cM to 176.2 cM and the average marker spacing varied from 7.9 cM to 10.9 cM per linkage group with an overall average marker spacing of 8.6 cM. The maximum marker spacing was on linkage group 4 at 27.7 cM. The number of markers varied from 8 to 21 per linkage group.

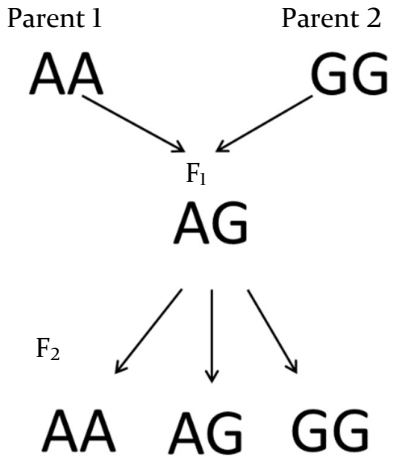


Figure 4.10. *Example of a useful marker as identified from the sequencing for generating linkage maps. Parents are homozygous for different alleles leading to a heterozygous F1. If the marker is segregating, the F2 population will consist of homozygotes of each parental allele and heterozygotes.*

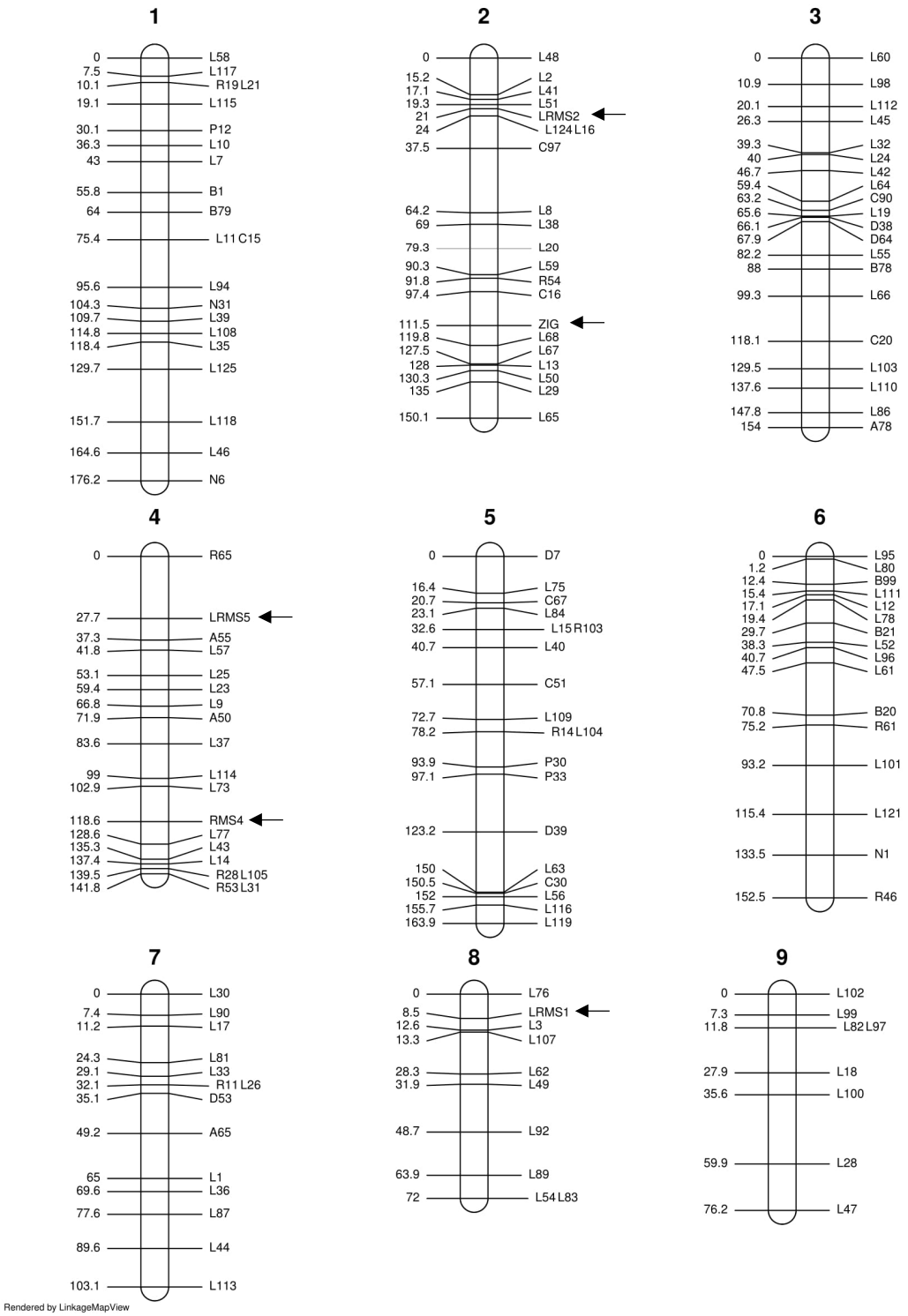


Figure 4.11. Genetic linkage maps constructed for the Lincoln *Sophora* F_2 of crosses between *S. tetraptera* and *S. prostrata* with 148 markers across nine linkage groups. Marker names indicated on right of each linkage group and position (cM) indicated on the left. Gene markers indicated with black arrows.

Table 4.4. Description of linkage groups developed for *Sophora* using MRMEseq

Linkage Group	Number of markers	Length (cM)	Average spacing	Max spacing
1	21	176.2	8.8	22.1
2	21	150.1	7.5	26.7
3	20	154.0	8.1	18.9
4	19	141.8	7.9	27.7
5	19	163.9	9.1	26.8
6	16	152.5	10.2	23.3
7	14	103.1	7.9	15.8
8	10	72.0	8.0	16.8
9	8	76.2	10.9	24.2
Overall	**148**	**1189.9**	**8.6**	**27.7**

4.5 Discussion

Linkage maps represent the order and relative genetic distance of markers on linkage groups or chromosomes based on recombination frequency between markers. They are an important resource for species without a sequenced genome by providing a basic framework for marker spacing and ordering. One of the most important roles of linkage maps is in QTL mapping, to identify regions of the genome that contain genes for the traits of interest, and are particularly important in mapping quantitative traits, which do not follow traditional Mendelian inheritance as they are influenced by multiple genes. Linkage maps are also valuable when used as a reference for sequence assembly and validation of sequenced genomes (Fierst, 2015) and can provide information on linked markers that can be used in breeding programs or to aid identification of candidate genes from the sequenced genomes. These aspects of linkage map usage mean linkage maps remain an important tool in genetic studies today and, with the advances of next-generation sequencing, generating new, and higher density, linkage maps are possible.

4.5.1 MRMEseq

The advances in next-generation sequencing have reduced costs such that genome-level sequencing, using MRMEseq, is possible allowing for large scale sequence analyses. MRMEseq (Figure 4.5) constructs a reduced library of random DNA fragments, using RAPD primers, that are representative of the genome, for sequencing on the Illumina next-generation platform. Large amounts of sequence data are generated where SNPs, and other useful markers, can be identified for genotyping and analysis. RAPDs are relatively faster and cheaper than many other markers (Semagn et al., 2006; Kumar et al., 2009), need no prior sequence knowledge and require less DNA to successfully

amplify. They have high genome abundance and good genome coverage (Kumar et al., 2009) making RAPDs advantageous in reducing the genome for sequencing.

The pilot study showed MRMEseq produces a large number of sequences and a suitable number of markers can be generated from these sequences. The number of sequences from each library in the pilot ranged from 7,736 to 353,402, however, as samples were pooled at different concentrations (Table 4.3), in order to test different methods, the variation in sequence number was expected. The number of sequences obtained reflected the difference in DNA concentration in the final pool with a higher DNA concentration resulting in more sequences than lower DNA concentration (Table 4.3, Figure 4.8). Samples that were pooled at the same concentration had largely comparable sequence numbers. Overall about 100 markers suitable for generating linkage maps were identified from the pilot study using the strictest criteria with more markers possibly identifiable using more relaxed criteria.

Previous studies have suggested the number of markers needed varies with the number and length of chromosomes however 100-200 markers (Lynch & Walsh, 1998; Schneider, 2005; Ferreira et al., 2006) are often suggested as a suitable number to balance cost/resources and mapping QTL. A density of less than 10 cM is often preferred as the further a marker is from a QTL, the chance of detecting the QTL and the magnitude of its effect decrease (Singh & Singh, 2015). However, it is not possible to know the spacing of markers before generating the maps and simulations have suggested the power to detect QTL is only slightly decreased for marker spacings of 20 cM (Darvasi et al., 1993). Using more markers can be helpful for fine mapping however the usefulness is dependent on the potential for recombination in the mapping population and therefore a large number of markers in a smaller population is still limited by the number of recombination events in that population (Liu, 1998).

The pilot study indicated that the MRMEseq method was suitable to generate markers for the development of linkage maps for the *Sophora* F_2 population. Multiplex PCR was not chosen due to the possibility of different primer combinations in the reaction that may not form consistent products between samples and competition from primers may affect the products generated. A Hiseq run was chosen for the full population as the number of samples pooled was increased from 31 to 96 which would result in a decrease in the number of reads per sample if a Miseq run was used. The Hiseq run yielded ~ 1 million sequences per samples with an average quality score of 35.2 (Appendix 4.2) providing, as expected, more sequences per sample than the Miseq run. From these sequences, a total of 143 markers, from MRMEseq, plus five gene-specific markers were used to generate the linkage maps. Compared to the number of SNPs generated from the Hiseq run, there was a considerable reduction in the number of SNPs developed into useful markers. This reduction was caused by (1) sequencing coverage, causing too much missing data at a locus, (2) not having the original parents of the crosses and different parents for the reciprocal crosses. A large number of markers were removed due to deviation from the expected inheritance, reducing the total SNP number from 3,170 to 160. Many of these markers showed a high number of one of the parental alleles and very low or none of the other parent which could be caused by a heterozygous locus in one parent where the same allele is inherited from both parents or a locus that is not showing Mendelian segregation. The use of a different *S. prostrata* parent, from different populations, may also have reduced the number of markers if one of the cross directions doesn't show segregation which can alter the overall genotype ratio.

4.5.2 Generating Linkage maps in *Sophora*

Prior to this work, very little information was available for genetic studies in *Sophora.* The mitochondrial (Shi et al., 2018) and chloroplast (Lu et al., 2018) genomes of

some species have been assembled but there is no reference genome or linkage maps available for the group. New Zealand *Sophora* are woody species with long generational times where generating a segregating population, including an F_2 population, is resource and time consuming, taking years to grow plants to maturity, however the existing F_2 population developed by Godley is a valuable resource to begin development of genetic resources in *Sophora*.

In total 148 markers were developed to generate linkage maps for the *Sophora* F_2 population. Four of these were developed from candidate genes, known to be involved in the strigolactone branching pathway, the *RAMOSUS* genes. Mutants of these genes (*RMSI, RMS2, RMS4, RMS5*) show similar phenotypes to the divaricate trait, including smaller leaves and increased branching (Morris et al., 2001; Sorefan et al., 2003; Auldridge et al., 2006; Johnson et al., 2006), in *Pisum* (Fabaceae) and therefore are potential candidate genes. The fifth marker was developed from a gene named *ZIG* from *Arabidopsis*. *ZIG* results in zig-zag stems in *Arabidopsis* mutants (Zheng et al., 1999; Kato et al., 2002), a feature seen in *S. prostrata,* and therefore is also a potential candidate gene to include here. These five markers displayed ratios similar to the expected 1:2:1 ratio (see Chapter 3) in an F_2 population generated from a cross between two homozygote parents indicating they are showing Mendelian segregation in this population.

Distorted segregation was calculated using a chi-squared test for each of the markers and was observed at 6.8% of markers in the *Sophora* map. Distorted segregation is commonly reported for segregating populations (Xu, 2008; Zhang et al., 2010), and can be caused by many factors including genes under selection, scoring error, sample size or self-incompatible alleles (Rick, 1969; Faris et al., 1998; Khodaeiaminjan et al., 2018). Often QTL are found in these regions of distorted segregation as they are

under selection (Wang et al., 2005) and it has been suggested these should be included in the maps (Wang et al., 2005; Xu, 2008; Van Ooijen, 2011). One method of doing this is by adding markers with distorted segregation after first generating a map without them to ensure these markers do not cause changes in marker order. This may allow recovery of QTL in these regions (Wang et al., 2005; Xu, 2008) or increase understanding of these regions (Van Ooijen, 2011). The *Sophora* map was generated without the ten distorted segregating markers first and their later inclusion didn't result in any changes of original marker order.

Map distances for the *Sophora* map were calculated using the Kosambi function (Kosambi, 1944) as it includes assumptions on crossover interference, where a crossover in one region affects the likelihood of a crossover in another, therefore the Kosambi function has a smaller bias than Haldane's function which assumes no cross-over interference (Huehn, 2011). The number of linkage groups expected was nine as the chromosome number for both *S. tetraptera* and *S. prostrata* is 2*n*=18 (Dawson, 2000; Stiefkens et al., 2003), which is also the most common chromosome number in *Sophora*. The physical chromosomes of *S. tetraptera* are homogeneous and range in size from 1.3 to 2 µm with one chromosome pair showing satellite arms. The *Sophora* linkage map (Figure 4.11, Table 4.4), ranged from 72 cM to 176.2 cM. The range in linkage map lengths does not correspond to the estimates for the physical sizes of the *Sophora* chromosomes, however, physical distance is not often correlated with genetic distance. The size of linkage groups can also be affected by (1) the distribution of markers, (2) areas with low recombination, where the ends and centromeres of the chromosomes are known to have little or no recombination (True et al., 1996; Collard et al., 2005; Myers et al., 2005), and (3) sample size.

The maximum spacing between markers in any of the linkage groups was 27.7 cM with six linkage groups showing at least one gap over 20 cM (Table 4.4). This could be due to (1) regions that are mostly homozygous between the parents, (2) the sample size, as this will affect the number of recombination events observed, or (3) regions not covered by marker generation. Large gaps can reduce the ability to detect QTL in those regions, as the further a QTL is from a marker the lower the power to detect it (Singh & Singh, 2015), and so further study to identify markers from these regions may be useful for identifying further QTL. However, QTL with large effect are still possible to detect with marker spacings of 20 cM (Darvasi et al., 1993). A maize population with 114 markers and 187 individuals generated maps with 5-10 cM among most markers however a population of 1,776 individuals with only 17 markers had only 40% of the genome within 20 cM of a marker (Edwards et al., 1987; Edwards et al., 1992), but was suggested to have greater power to detect QTL (Lynch & Walsh, 1998) indicating population size is also important for QTL mapping. A larger population size has a greater chance of detecting QTL with smaller effects and more precise mapping, however, it is not always practical to generate. Although, the population size should at least be larger than 50 individuals (Singh & Singh, 2015).

Creation of independent maps for each of the cross directions was also attempted. The CD population (with maternal parent *S. prostrata),* showed similar results to the overall map with one linkage group split into two and some change in marker positions within linkage groups. The AB population (with maternal parent *S. tetraptera)* had many orphaned markers and the overall linkage groups were broken into shorter fragments. This is most likely caused by the small sample size of the AB population (33 individuals compared to 55 individuals in the CD population). A population size of at least 50 is generally recommended to generate linkage maps (Young, 1994; Collard et al., 2005). Marker inversions and dividing of linkage groups are

a known problem with smaller sample sizes (Ferreira et al., 2006) where a drop in sample size from 100 to 50 caused a drop from 90% to 60% confidence of marker order in simulations (Liu, 1998). It is also likely that some markers were not segregating as well in each population due to the smaller sample size. Ideally each cross direction would have a suitable sample size individually to generate and compare maps from each cross direction. However, generating segregating populations from tree species requires a much longer time than generating those with short generation times, such as *Arabidopsis,* and resources and time are often limited so such populations are often not possible to generate. The *Sophora* F_2 population developed by Godley is an excellent resource to use as an initial step towards developing the first linkage maps for *Sophora* and attempting to identify QTL related to divaricate traits, despite the limitations such as a small size and incomplete history on the exact parental plants.

4.6 Conclusion

The difference in growth forms between the parent species used to generate the *Sophora* F_2 population, *S. prostrata* (divaricate) and *S. tetraptera* (non-divaricate), offer an excellent opportunity to investigate the genetic architecture behind this growth form as well as other distinguishing features of these species. The pilot study of a selection of the *Sophora* F_2 population showed the MRMEseq method suitable to generate a large amount of sequence data from which suitable markers can be developed. Overall 148 markers were identified from the *Sophora* F_2 population and were used to generate the first linkage groups. Nine linkage groups were obtained for this population, which is the expected number based on the chromosome numbers of *S. tetraptera* and *S. prostrata*. This genetic map provides the first map developed for *Sophora* and will provide an excellent resource to be used in QTL mapping of the divaricate habit and other traits that differ between the parental species. As well, the *Sophora* linkage map is further useful for comparing with other species among the Fabaceae and for any studies, including phylogenetic work, in *Sophora*.

4.7 References

Allan, H. H. (1961). *Flora of New Zealand. Vol. 1. Indigenous Tracheophyta: Psilopsida, Lycopsida, Filicopsida, Gymnospermae, Dicotyledones.* Government Printer, Wellington.

Altshuler, D., Daly, M. J., & Lander, E. S. (2008). Genetic mapping in human disease. *Science, 322*(5903), 881-888. 10.1126/science.1156409

Auldridge, M. E., Block, A., Vogel, J. T., Dabney-Smith, C., Mila, I., Bouzayen, M., Magallanes-Lundback, M., DellaPenna, D., McCarty, D. R., & Klee, H. J. (2006). Characterization of three members of the *Arabidopsis* carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family. *Plant Journal, 45*(6), 982-993. 10.1111/j.1365-313X.2006.02666.x

Botstein, D., & Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics, 33*, 228-237. 10.1038/ng1090

Broman, K. W. (2012). *Genetic map construction with R/qtl* (No. 214). University of Wisconsin-Madison. Retrieved from http://www.rqtl.org/tutorials/geneticmaps.pdf

Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics, 19*(7), 889-890. 10.1093/bioinformatics/btg112

Bruneau, A., Doyle, J. J., Herendeen, P., Hughes, C., Kenicer, G., Lewis, G., Mackinder, B., Pennington, R. T., Sanderson, M. J., Wojciechowski, M. F., Boatwright, S., Brown, G., Cardoso, D., Crisp, M., Egan, A., Fortunato, R. H., Hawkins, J., Kajita, T., Klitgaard, B., Koenen, E., Lavin, M., Luckow, M., Marazzi, B., McMahon, M. M., Miller, J. T., Murphy, D. J., Ohashi, H., de Queiroz, L. P., Rico, L., Saerkinen, T., Schrire, B., Simon, M. F., Souza, E. R., Steele, K., Torke, B. M., Wieringa, J. J., van Wyk, B.-E., & Legume Phylogeny Working, G. (2013). Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species-rich clades. *Taxon, 62*(2), 217-248.

Champagne, C. E. M., Goliber, T. E., Wojciechowski, M. F., Mei, R. W., Townsley, B. T., Wang, K., Paz, M. M., Geeta, R., & Sinhaa, N. R. (2007). Compound leaf development and evolution in the legumes. *Plant Cell, 19*(11), 3369-3378. 10.1105/tpc.107.052886

Chen, F., Dong, W., Zhang, J., Guo, X., Chen, J., Wang, Z., Lin, Z., Tang, H., & Zhang, L. (2018). The sequenced angiosperm genomes and genome databases. *Frontiers in Plant Science, 9* 10.3389/fpls.2018.00418

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica, 142*(1-2), 169-196. 10.1007/s10681-005-1681-5

Cregan, P. B., Jarvik, T., Bush, A. L., Shoemaker, R. C., Lark, K. G., Kahler, A. L., Kaya, N., VanToai, T. T., Lohnes, D. G., Chung, L., & Specht, J. E. (1999). An integrated genetic linkage map of the soybean genome. *Crop Science, 39*(5), 1464-1490. 10.2135/cropsci1999.3951464x

Darvasi, A., Weinreb, A., Minke, V., Weller, J. I., & Soller, M. (1993). Detecting marker-qtl linkage and estimating qtl gene effect and map location using a saturated genetic-map. *Genetics, 134*(3), 943-951.

Dawson, M. I. (2000). Index of chromosome numbers of indigenous New Zealand spermatophytes. *New Zealand Journal of Botany, 38*(1), 47-150.

Doyle, J. J., & Luckow, M. A. (2003). The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiology, 131*(3), 900-910. 10.1104/pp.102.018150

Eaton. (2015). ipyrad:interactive assembly and analysis of RADseq data sets. Retrieved from https://ipyrad.readthedocs.io/#

Eaton, D. A. R. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics, 30*(13), 1844-1849. 10.1093/bioinformatics/btu121

Edwards, M. D., Helentjaris, T., Wright, S., & Stuber, C. W. (1992). Molecular-marker-facilitated investigations of quantitative trait loci in maize .4. analysis based on genome saturation with isozyme and restriction-fragment-length-polymorphism markers. *Theoretical and Applied Genetics, 83*(6-7), 765-774. 10.1007/bf00226696

Edwards, M. D., Stuber, C. W., & Wendel, J. F. (1987). Molecular-marker-facilitated investigations of quantitative-trait loci in maize .I. numbers, genomic distribution and types of gene-action. *Genetics, 116*(1), 113-125.

Faris, J. D., Laddomada, B., & Gill, B. S. (1998). Molecular mapping of segregation distortion loci in *Aegilops tauschii*. *Genetics, 149*(1), 319-327.

Ferreira, A., da Silva, M. F., Silva, L., & Cruz, C. D. (2006). Estimating the effects of population size and type on the accuracy of genetic maps. *Genetics and Molecular Biology, 29*(1), 187-192. 10.1590/s1415-47572006000100033

Fierst, J. L. (2015). Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. *Frontiers in Genetics, 6* 10.3389/fgene.2015.00220

Gailing, O., Vornam, B., Leinemann, L., & Finkeldey, R. (2009). Genetic and genomic approaches to assess adaptive genetic variation in plants: forest trees as a model. *Physiologia Plantarum, 137*(4), 509-519. 10.1111/j.1399-3054.2009.01263.x

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics, 10*(6), 381-391. 10.1038/nrg2575

Godley, E. J. (1985). Paths to maturity. *New Zealand Journal of Botany, 23*, 687-706.

Graham, P. H., & Vance, C. P. (2003). Legumes: Importance and constraints to greater use. *Plant Physiology, 131*(3), 872-877. 10.1104/pp.017004

Greenwood, R. M., & Atkinson, I. A. E. (1977). Evolution of divaricating plants in New Zealand in relation to moa browsing. *Proceedings of the New Zealand Ecological Society, 24*

Grierson, E. R. P. (2014). *The Development and Genetic Variation of Sophora prostrata - A New Zealand Divaricating Shrub.* (Master's Thesis), The University of Waikato, Hamilton, New Zealand.

Han, R., Takahashi, H., Nakamura, M., Bunsupa, S., Yoshimoto, N., Yamamoto, H., Suzuki, H., Shibata, D., Yamazaki, M., & Saito, K. (2015). Transcriptome analysis of nine tissues to discover genes involved in the biosynthesis of active ingredients in *Sophora flavescens*. *Biological & Pharmaceutical Bulletin, 38*(6), 876-883. 10.1248/bpb.b14-00834

Hayashi, M., Miyahara, A., Sato, S., Kato, T., Yoshikawa, M., Taketa, M., Hayashi, M., Pedrosa, A., Onda, R., Imaizumi-Anraku, H., Bachmair, A., Sandal, N., Stougaard, J., Murooka, Y., Tabata, S., Kawasaki, S., Kawaguchi, M., & Harada, K. (2001). Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F-2 population. *DNA Research, 8*(6), 301-310. 10.1093/dnares/8.6.301

Heenan, P., Mitchell, C., & Houliston, G. (2018). Genetic variation and hybridisation among eight species of kowhai (*Sophora*: Fabaceae) from New Zealand revealed by microsatellite markers. *Genes, 9*(2) 10.3390/genes9020111

Heenan, P. B., Dawson, M. I., & Wagstaff, S. J. (2004). The relationship of *Sophora sect. Edwardsia* (Fabaceae) to *Sophora tomentosa*, the type species of the genus *Sophora*, observed from DNA sequence data and morphological characters. *Botanical Journal of the Linnean Society, 146*(4), 439-446. 10.1111/j.1095-8339.2004.00348.x

Heenan, P. B., de Lange, P. J., & Wilton, A. D. (2001). *Sophora* (Fabaceae) in New Zealand: Taxonomy, distribution, and biogeography. *New Zealand Journal of Botany, 39*(1), 17-53. 10.1080/0028825x.2001.9512715

Huehn, M. (2011). On the bias of recombination fractions, Kosambi's and Haldane's distances based on frequencies of gametes. *Genome, 54*(3), 196-201. 10.1139/g10-109

Hughes, D. (2002). *Sophora* — The Kowhais of New Zealand. *Combined Proceedings International Plant Propagators' Society, 52*

Hurr, K. A., Lockhart, P. J., Heenan, P. B., & Penny, D. (1999). Evidence for the recent dispersal of *Sophora* (Leguminosae) around the Southern Oceans: molecular data. *Journal of Biogeography, 26*(3), 565-577. 10.1046/j.1365-2699.1999.00302.x

Jin, C. F., Lan, H., Attie, A. D., Churchill, G. A., Bulutuglo, D., & Yandell, B. S. (2004). Selective phenotyping for increased efficiency in genetic mapping studies. *Genetics, 168*(4), 2285-2293. 10.1534/genetics.104.027524

Johnson, X., Brcich, T., Dun, E. A., Goussot, M., Haurogne, K., Beveridge, C. A., & Rameau, C. (2006). Branching genes are conserved across species. Genes controlling a novel signal in pea are coregulated by other long-distance signals. *Plant Physiology, 142*(3), 1014-1026. 10.1104/pp.106.087676

Kato, T., Morita, M. T., Fukaki, H., Yamauchi, Y., Uehara, M., Niihama, M., & Tasaka, M. (2002). SGR2, a Phospholipase-Like Protein, and ZIG/SGR4, a SNARE, Are Involved in the Shoot Gravitropism of *Arabidopsis. The Plant Cell Online, 14*(1), 33-46. 10.1105/tpc.010215

Kelly, D. (1994). Towards a numerical definition of divaricate (interlaced small-leaved) shrubs. *New Zealand Journal of Botany, 32*(4), 509-518.

Khodaeiaminjan, M., Kafkas, S., Motalebipour, E. Z., & Coban, N. (2018). In silico polymorphic novel SSR marker development and the first SSR-based genetic linkage map in pistachio. *Tree Genetics & Genomes, 14*(4) 10.1007/s11295-018-1259-8

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods, 9* 10.1186/1746-4811-9-29

Kosambi, D. D. (1944). The estimation of map distances from recombination values. *Ann Eugenics, 12*((3)), 172-175.

Kumar, P., Gupta, V. K., Misra, A. K., Modi, D. R., & Pandey, B. K. (2009). Potential of Molecular Markers in Plant Biotechnology. *Plant Omics, 2*(4), 141-162.

Lander, E. S., & Schork, N. J. (1994). Genetic dissection of complex traits. *Science, 265*(5181), 2037-2048. 10.1126/science.8091226

Lee, J.-Y., Lee, D.-H., & Choi, B.-H. (2012). Isolation and characterization of 13 microsatellite loci from a Korean endemic species, *Sophora koreensis* (Fabaceae). *International Journal of Molecular Sciences, 13*(9), 10765-10770. 10.3390/ijms130910765

Lewis, G., Schrire, B., Mackinder, B., & Lock, M. e. (2005). *Legumes of the world*. Richmond. U.K.: Royal Botanic Gardens. Kew.

Li, H., Yao, W., Fu, Y., Li, S., & Guo, Q. (2015). De novo assembly and discovery of genes that are involved in drought tolerance in Tibetan *Sophora moorcroftiana*. *Plos One, 10*(1) 10.1371/journal.pone.0111054

Liu, H. (1998). *Statistical Genomics, Linkage, Mapping and QTL Analysis*. Boca Raton, Florida: CRC.

Liu, Z. M., Zhao, A. M., Kang, X. Y., Zhou, S. L., & Lopez-Pujol, J. (2006). Genetic diversity, population structure, and conservation of *Sophora moorcroftiana* (Fabaceae), a shrub endemic to the Tibetan Plateau. *Plant Biology, 8*(1), 81-92. 10.1055/s-2005-872889

Lu, Y., Li, W., Xie, X., Zheng, Y., & Li, B. (2018). The complete chloroplast genome sequence of *Sophora japonica* var. *violacea*: gene organization and genomic resources. *Conservation Genetics Resources, 10*(1), 1-4. 10.1007/s12686-017-0748-7

Lynch, M., & Walsh, B. (1998). *Genetics and Analysis of Quantative Traits*. Sunderland, MA, U.S.A: Sinauer Associates, Inc.

Maich, B. (2002). *A biochemical genetic evaluation of taxonomy of Sophora microphylla (Fabaceae Subfamily Papilionoideae).* (Master's thesis), Victoria University, Wellington, New Zealand.

McLay, T. (2018). Hybridisation between species of *Xanthorrhoea* in Yuraygir National Park revealed using a newly developed next-generation sequencing method. *Australian Systematic Botany Society Newsletter, 175,* 13-16.

Mitchell, A. D., & Heenan, P. B. (2002). *Sophora* sect. *Edwardsia* (Fabaceae): further evidence from nrDNA sequence data of a recent and rapid radiation around the Southern Oceans. *Botanical Journal of the Linnean Society, 140*(4), 435-441. 10.1046/j.1095-8339.2002.00101.x

Mohan, M., Nair, S., Bhagwat, A., Krishna, T. G., Yano, M., Bhatia, C. R., & Sasaki, T. (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding, 3*(2), 87-103. 10.1023/a:1009651919792

Morris, J. B. (1997). Special-purpose legume genetic resources conserved for agricultural, industrial, and pharmaceutical use. *Economic Botany, 51*(3), 251-263. 10.1007/bf02862094

Morris, S. E., Turnbull, C. G. N., Murfet, I. C., & Beveridge, C. A. (2001). Mutational analysis of branching in pea. Evidence that *RMS1* and *RMS5* regulate the same novel signal. *Plant Physiology, 126*(3), 1205-1213. 10.1104/pp.126.3.1205

Myers, S., Bottolo, L., Freeman, C., McVean, G., & Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. *Science, 310*(5746), 321-324. 10.1126/science.1117196

Nagy, E. D., Guo, Y., Tang, S., Bowers, J. E., Okashah, R. A., Taylor, C. A., Zhang, D., Khanal, S., Heesacker, A. F., Khalilian, N., Farmer, A. D., Carrasquilla-Garcia, N., Penmetsa, R. V., Cook, D., Stalker, H. T., Nielsen, N., Ozias-Akins, P., & Knapp, S. J. (2012). A high-density genetic map of *Arachis duranensis*, a diploid ancestor of cultivated peanut. *Bmc Genomics, 13* 10.1186/1471-2164-13-469

New Zealand Plant Conservation Network. (2019, 6 December 2014). *Sophora prostrata*. Retrieved 13 March 2019 from http://www.nzpcn.org.nz/flora_details.aspx?ID=1303

Ouellette, L. A., Reid, R. W., Blanchard, S. G., & Brouwer, C. R. (2018). LinkageMapView-rendering high-resolution linkage and QTL maps. *Bioinformatics, 34*(2), 306-307. 10.1093/bioinformatics/btx576

Paetau, I., Chen, C. Z., & Jane, J. L. (1994). Biodegradable plastic made from soybean products .1. effect of preparation and processing on mechanical-properties and water-absorption. *Industrial & Engineering Chemistry Research, 33*(7), 1821-1827. 10.1021/ie00031a023

Python Software Foundation. (2001). Python. Retrieved from https://www.python.org/

Rick, C. M. (1969). Controlled introgression of chromosomes of *Solanum pennellii* into *Lycopersicon esculentum* - segregation and recombination. *Genetics, 62*(4), 753-&.

Salatino, S., & Ramraj, V. (2017). BrowseVCF: a web-based application and workflow to quickly prioritize disease-causative variants in VCF files. *Briefings in Bioinformatics, 18*(5), 774-779. 10.1093/bib/bbw054

Schneider, K. (2005). *Mapping populations and principles of genetic mapping*. Weiheim: WILEY_VCH Verlag GmbH & Co KGaA.

Schrire, B. D., Lavin, M., & Lewis, G. P. (2005). Global distribution patterns of the Leguminosae: insights from recent phylogenies. *Kongelige Danske Videnskabernes Selskab Biologiske Skrifter, 55*, 375-422.

Semagn, K., Bjornstad, A., & Ndjiondjop, M. N. (2006). An overview of molecular marker methods for plants. *African Journal of Biotechnology, 5*(25), 2540-2568.

Shepherd, L. D., de Lange, P. J., Perrie, L. R., & Heenan, P. B. (2017). Chloroplast phylogeography of New Zealand *Sophora* trees (Fabaceae): extensive hybridization and widespread Last Glacial Maximum survival. *Journal of Biogeography, 44*(7), 1640-1651. 10.1111/jbi.12963

Shi, Y., Liu, Y., Zhang, S., Zou, R., Tang, J., Mu, W., Peng, Y., & Dong, S. (2018). Assembly and comparative analysis of the complete mitochondrial genome sequence of *Sophora japonica* 'JinhuaiJ2'. *Plos One, 13*(8) 10.1371/journal.pone.0202485

Simioniuc, D., Burlacu-Arsene, M., Simioniuc, V., & Lipsa, F. (2011). A genetic linkage map for white lupin (*Lupinus albus* L.). *Lucrari Stiintifice, Universitatea de Stiinte Agricole Si Medicina Veterinara "Ion Ionescu de la Brad" Iasi, Seria Agronomie, 54*(2), 96-98.

Singh, B. D., & Singh, A. K. (2015). Mapping Populations. In *Marker-Assisted Plant Breeding: Principles and Practices* (pp. 125-150). New Delhi: Springer India. 10.1007/978-81-322-2316-0_5

Singh, J., Kalberer, S. R., Belamkar, V., Assefa, T., Nelson, M. N., Farmer, A. D., Blackmon, W. J., & Cannon, S. B. (2018). A transcriptome-SNP-derived linkage map of *Apios americana* (potato bean) provides insights about genome re-organization and synteny conservation in the phaseoloid legumes. *Theoretical and Applied Genetics, 131*(2), 333-351. 10.1007/s00122-017-3004-3

Slate, J. (2005). Quantitative trait locus mapping in natural populations: progress, caveats and future directions. *Molecular Ecology, 14*(2), 363-379. 10.1111/j.1365-294X.2004.02378.x

Song, J. (2005). *Genetic diversity and flowering in Clianthus and New Zealand Sophora (Fabaceae).* (Ph.D. Thesis), Massey University, Palmerston North, New Zealand.

Sorefan, K., Booker, J., Haurogne, K., Goussot, M., Bainbridge, K., Foo, E., Chatfield, S., Ward, S., Beveridge, C., Rameau, C., & Leyser, O. (2003). MAX4 and RMS1 are orthologous dioxygenase-like genes that regulate shoot branching in Arabidopsis and pea. *Genes & Development, 17*(12), 1469-1474. 10.1101/gad.256603

Stiefkens, L. B., Bernardello, G., & Anderson, G. J. (2003). The karyotype of *Sophora tetraptera* (Fabaceae). *New Zealand Journal of Botany, 41*(4), 731-735.

Thoquet, P., Gherardi, M., Journet, E.-P., Kereszt, A., Ane, J.-M., Prosperi, J.-M., & Huguet, T. (2002). The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes. *BMC Plant Biology, 2*(1 Cited May 5, 2002), 1-13. 10.1186/1471-2229-2-1

True, J. R., Mercer, J. M., & Laurie, C. C. (1996). Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics, 142*(2), 507-523.

Van Etten, M. L., Houliston, G. J., Mitchell, C. M., Heenan, P. B., Robertson, A. W., & Tate, J. A. (2014). *Sophora microphylla* (Fabaceae) microsatellite markers and their utility across the genus. *Applications in Plant Sciences, 2*(3) 10.3732/apps.1300081

Van Ooijen, J. W. (2011). Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genetics Research, 93*(5), 343-349. 10.1017/s0016672311000279

Wang, C. M., Zhu, C. S., Zhai, H. Q., & Wan, J. M. (2005). Mapping segregation distortion loci and quantitative trait loci for spikelet sterility in rice (*Oryza sativa* L.). *Genetical Research, 86*(2), 97-106. 10.1017/s0016672305007779

Wojciechowski, M. F., Lavin, M., & Sanderson, M. J. (2004). A phylogeny of legumes (Leguminosae) based on analyses of the plastid matK gene resolves many well-supported subclades within the family. *American Journal of Botany, 91*(11), 1846-1862. 10.3732/ajb.91.11.1846

Xu, S. Z. (2008). Quantitative trait locus mapping can benefit from segregation distortion. *Genetics, 180*(4), 2201-2208. 10.1534/genetics.108.090688

Xu, Y. B., & Crouch, J. H. (2008). Marker-assisted selection in plant breeding: From publications to practice. *Crop Science, 48*(2), 391-407. 10.2135/cropsci2007.04.0191

Young, N. D. (1994). *Constructing a plant genetic linkage map with DNA markers*.

Zhang, L. Y., Wang, S. Q., Li, H. H., Deng, Q. M., Zheng, A. P., Li, S. C., Li, P., Li, Z. L., & Wang, J. K. (2010). Effects of missing marker and segregation distortion on QTL mapping in F-2 populations. *Theoretical and Applied Genetics, 121*(6), 1071-1082. 10.1007/s00122-010-1372-z

Zheng, H. Y., von Mollard, G. F., Kovaleva, V., Stevens, T. H., & Raikhel, N. V. (1999). The plant vesicle-associated SNARE AtVTI1a likely mediates vesicle transport from the trans-Golgi network to the prevacuolar compartment. *Molecular Biology of the Cell, 10*(7), 2251-2264.

Zhu, L., Zhang, Y., Guo, W., Xu, X.-J., & Wang, Q. (2014). De novo assembly and characterization of *Sophora japonica* transcriptome using RNA-seq. *BioMed Research International* 10.1155/2014/750961

4.8 Appendix

Appendix 4.1. Parameters used in analysing Illumina Hiseq data using ipyrad.

	Parameter	Description	Setting
0	Assembly name	Prefix for all output files	Sip6, Sip8
1	Project dir	Project directory	
2	Raw fastq path	Path to raw non-demultiplexed fastq files	NA
3	Barcodes path	Location of barcodes file	
4	Sorted fastq path	Location of sorted fastq data	/Users/09030727/Desktop/Sophorabigdataparents/Fastq/*.fastq.gz
5	Assembly method	Assembly method to use: denovo, reference, denovo+reference, denovo-reference	Denovo
6	Reference sequence	Location of reference sequence	NA
7	Datatype	Type of data	RAD
8	Restriction_overhang	Restriction overhang sequence	NA
9	Max_low_qual_bases	Reads with more then specified low quality are excluded (default is 5)	10
10	Phred_Qscore_offset	Threshold for base call considered low quality. Default is 33 equivalent to minimum qscore of 20	33
11	Mindepth_statistical	Minimum depth at which statistical base calls will be made during consensus base calling. Default is 6	6
12	Mindepth_majrule	Minimum depth majority rule base calls made during consensus base calling. Must be equal or below mindepth_stastical	6
13	Maxdepth	Max cluster depth within samples. Limits repetitive regions clustering as high depth clusters. Default is 10,000	100,000
14	Clust_threshold	Level of sequence similarity two sequences are identified as being homologous and cluster together	0.85
15	Max_barcodes_mismatch	Maximum number of allowed mismatches between barcodes in barcodes and sequence read files	0

	Parameter	Description	Setting
16	Filter_adapters	Sets whether filtering of adapters is needed from sequences. 0 indicates no adapter filtering, 1 filter based on quality scores, 2 strict filtering for adapters	0
17	Filter_min_trim_len	If filter_adapters is >0 reads trimmed to shorter length, sets limit on minimum length of trimming. Default is 35	35
18	Max_alleles_consens	Maximum number of unique alleles allowed in (individual) consensus reads after accounting for sequence errors. Default is 2	2
19	Max_Ns_consens	Maximum number of uncalled bases allowed in consensus sequence. Default is 5, 5	5, 5
20	Max_Hs_consens	Maximum number of heterozygous bases allowed in consensus sequences. Default is 8,8. Helps to remove poor alignments	10, 10
21	Min_sample_locus	Minimum number of samples that must have data at the locus for it to be retained in final data set. Default is 4	20
22	Max_SNPs_locus	Maximum SNPs allowed in a final locus. Helps remove poor alignments in repetitive regions. Default is 20, 20	20, 20
23	Max_Indels_locus	Maximum number of Indels allowed in final locus. Helps filter out poor alignments. Default is 8, 8	20,20
24	Max_shared_Hs_locus	Maximum number of shared polymorphic sites in a locus. Used to identify potential paralogs. Default is 0.5	0.5
25	Edit_cut_sites	Trim N bases from raw read sequence edges	0, 0, 0, 0
26	Trim_overhang	Trim N bases from edges of final aligned loci	0, 0, 0, 0
27	Output_formats	Formats to output results into	p, s, v
28	Pop_assign_file	For creating population output files	NA

Appendix 4.2. Summary results of the Hiseq Illumina sequencing for the 88 *Sophora* F_2.

Sample	Barcode sequence	PF Clusters	% of the lane	% Perfect barcode	% One mismatch barcode	Yield (Mbases)	% PF Clusters	% >= Q30 bases	Mean Quality Score	%GC
Total pool	N/A	147,033,225	90.69	90.94	9.06	36,758	88.79	91.43	35.20	
F2-A2	TACGAAGTC+ GACTTCGTA	708,938	0.44	87.58	12.42	177	88.79	90.09	34.93	48%
F2-A4	GACGAGATT+ GACTTCGTA	1,520,926	0.94	88.08	11.92	380	89.93	91.70	35.24	49%
F2_A8	TAGTGGCAA+ GACTTCGTA	1,696,942	1.05	89.46	10.54	424	89.44	91.44	35.19	48%
F2_A10	CATTAACGC+ GACTTCGTA	1,546,398	0.95	88.41	11.59	387	88.38	90.83	35.08	49%
F2-A11	TCGTTGAAG+ GACTTCGTA	1,416,198	0.87	90.05	9.95	354	89.11	90.90	35.09	48%
F2-A12	TAGTACGCT+ GACTTCGTA	1,496,273	0.92	89.40	10.60	374	88.60	91.30	35.17	48%
F2-A13	TTCACCGTA+ GACTTCGTA	1,636,155	1.01	89.34	10.66	409	89.16	91.60	35.22	48%
F2-A17	AGGACAGTT+ GACTTCGTA	696,424	0.43	84.70	15.30	174	88.44	87.15	34.38	47%
F2-A19	TACGAAGTC+ AATCTCGTC	1,533,703	0.95	91.99	8.01	383	88.99	91.06	35.13	48%
F2-A30	GACGAGATT+ AATCTCGTC	1,425,307	0.88	91.10	8.90	356	89.17	91.36	35.18	49%

Sample	Barcode sequence	PF Clusters	% of the lane	% Perfect barcode	% One mismatch barcode	Yield (Mbases)	% PF Clusters	% >= Q30 bases	Mean Quality Score	%GC
F2-A31	TAGTGGCAA+ AATCTCGTC	1,476,405	0.91	91.54	8.46	369	88.76	91.07	35.13	49%
F2_A39	CATTAACGC+ AATCTCGTC	1,699,172	1.05	92.23	7.77	425	88.66	91.74	35.26	49%
F2_A40	TCGTTGAAG+ AATCTCGTC	1,664,708	1.03	91.85	8.15	416	89.23	91.78	35.26	48%
F2_A42	TAGTACGCT+ AATCTCGTC	1,694,119	1.04	92.14	7.86	424	88.96	91.53	35.22	48%
F2_A43	TTCACCGTA+ AATCTCGTC	1,460,675	0.90	91.60	8.40	365	88.64	91.22	35.16	49%
F2_A44	AGGACAGTT+ AATCTCGTC	1,319,874	0.81	91.14	8.86	330	89.51	90.92	35.10	49%
F2_A46	TACGAAGTC+ TTGCCACTA	1,081,664	0.67	92.16	7.84	270	85.61	90.34	35.02	48%
F2_A52	GACGAGATT+ TTGCCACTA	1,185,589	0.73	91.61	8.39	296	86.39	91.25	35.19	49%
F2_A53	TAGTGGCAA+ TTGCCACTA	1,284,638	0.79	93.09	6.91	321	85.16	91.21	35.19	48%
F2_A54	CATTAACGC+ TTGCCACTA	1,263,089	0.78	92.62	7.38	316	85.85	91.26	35.20	48%
F2_A58	TCGTTGAAG+ TTGCCACTA	1,383,894	0.85	93.21	6.79	346	86.45	91.29	35.20	49%
F2_A59	TAGTACGCT+ TTGCCACTA	1,220,902	0.75	92.87	7.13	305	86.06	91.43	35.23	48%

Sample	Barcode sequence	PF Clusters	% of the lane	% Perfect barcode	% One mismatch barcode	Yield (Mbases)	% PF Clusters	% >= Q30 bases	Mean Quality Score	%GC
F2_A60	TTCACCGTA+ TTGCCACTA	911,884	0.56	91.26	8.74	228	85.17	89.47	34.85	48%
F1_F1 2-7	AGGACAGTT+ TTGCCACTA	1,383,970	0.85	92.29	7.71	346	87.44	91.29	35.20	48%
F2_B2	TACGAAGTC+ GCGTTAATG	1,226,695	0.76	91.08	8.92	307	87.85	89.98	34.95	48%
F2_B4	GACGAGATT+ GCGTTAATG	1,316,563	0.81	90.97	9.03	329	88.31	90.68	35.08	48%
F2_B7	TAGTGGCAA+ GCGTTAATG	1,402,974	0.87	90.77	9.23	351	87.45	90.11	34.97	49%
F2_B8	CATTAACGC+ GCGTTAATG	1,346,147	0.83	91.60	8.40	337	87.16	90.69	35.09	48%
F2_B10	TCGTTGAAG+ GCGTTAATG	1,438,793	0.89	92.22	7.78	360	88.00	90.64	35.07	48%
F2_B29	TAGTACGCT+ GCGTTAATG	1,288,909	0.80	91.36	8.64	322	87.39	90.46	35.04	49%
F2_B36	TTCACCGTA+ GCGTTAATG	1,634,619	1.01	91.84	8.16	409	88.11	91.03	35.14	48%
F2_B39	AGGACAGTT+ GCGTTAATG	1,311,651	0.81	90.91	9.09	328	88.19	90.13	34.98	49%
F2_B42	TACGAAGTC+ CTTCAACGA	1,027,664	0.63	88.35	11.65	257	86.88	90.33	35.02	49%
F2_B54	GACGAGATT+ CTTCAACGA	1,274,412	0.79	87.70	12.30	319	87.67	91.61	35.26	49%

Sample	Barcode sequence	PF Clusters	% of the lane	% Perfect barcode	% One mismatch barcode	Yield (Mbases)	% PF Clusters	% >= Q30 bases	Mean Quality Score	%GC
F2_B59	TAGTGGCAA+ CTTCAACGA	810,497	0.50	88.80	11.20	203	86.24	89.19	34.80	48%
F2_C2	CATTAACGC+ CTTCAACGA	1,263,519	0.78	88.55	11.45	316	86.68	91.47	35.24	49%
F2_C3	TCGTTGAAG+ CTTCAACGA	1,283,094	0.79	89.24	10.76	321	86.37	90.81	35.11	48%
F2_C7	TAGTACGCT+ CTTCAACGA	1,571,738	0.97	88.94	11.06	393	87.30	91.74	35.29	48%
F2_C8	TTCACCGTA+ CTTCAACGA	1,409,934	0.87	88.82	11.18	352	87.55	91.51	35.25	48%
F2_C11	AGGACAGTT+ CTTCAACGA	1,276,063	0.79	88.59	11.41	319	87.83	91.36	35.22	49%
F2_C12	TACGAAGTC+ AGCGTACTA	1,672,034	1.03	92.48	7.52	418	89.87	91.59	35.22	48%
F2_C14	GACGAGATT+ AGCGTACTA	1,782,793	1.10	91.54	8.46	446	90.63	92.13	35.32	48%
F2_C16	TAGTGGCAA+ AGCGTACTA	1,840,882	1.14	92.63	7.37	460	90.19	92.02	35.30	48%
F2_C17	CATTAACGC+ AGCGTACTA	1,929,649	1.19	92.49	7.51	482	90.06	92.18	35.33	47%
F2_C24	TCGTTGAAG+ AGCGTACTA	2,105,505	1.30	93.04	6.96	526	90.14	92.26	35.34	48%
F2_C28	TAGTACGCT+ AGCGTACTA	1,580,987	0.98	91.96	8.04	395	89.21	91.33	35.17	48%

Sample	Barcode sequence	PF Clusters	% of the lane	% Perfect barcode	% One mismatch barcode	Yield (Mbases)	% PF Clusters	% >= Q30 bases	Mean Quality Score	%GC
F2_C32	TTCACCGTA+ AGCGTACTA	2,049,707	1.26	92.42	7.58	512	90.06	92.33	35.35	48%
F2_C33	AGGACAGTT+ AGCGTACTA	1,821,616	1.12	91.92	8.08	455	90.39	92.08	35.31	48%
F2_C34	TACGAAGTC+ TACGGTGAA	1,558,867	0.96	92.78	7.22	390	88.93	91.30	35.17	49%
F2_C41	GACGAGATT+ TACGGTGAA	1,729,492	1.07	92.51	7.49	432	89.62	92.31	35.36	47%
F2_C42	TAGTGGCAA+ TACGGTGAA	1,734,959	1.07	93.51	6.49	434	89.07	91.96	35.29	47%
F2_C44	CATTAACGC+ TACGGTGAA	1,969,573	1.21	93.01	6.99	492	88.61	91.73	35.25	48%
F2_C45	TCGTTGAAG+ TACGGTGAA	1,708,820	1.05	93.03	6.97	427	89.31	91.83	35.27	48%
F2_C46	TAGTACGCT+ TACGGTGAA	1,761,996	1.09	92.99	7.01	440	89.25	91.86	35.28	48%
F2_C47	TTCACCGTA+ TACGGTGAA	1,963,169	1.21	92.97	7.03	491	88.91	92.21	35.34	48%
F2_C49	AGGACAGTT+ TACGGTGAA	1,633,911	1.01	92.56	7.44	408	89.72	92.11	35.32	49%
F2_C51	TACGAAGTC+ AACTGTCCT	1,769,187	1.09	93.20	6.80	442	89.96	92.12	35.32	48%
F2_C52	GACGAGATT+ AACTGTCCT	1,576,396	0.97	92.01	7.99	394	90.25	92.53	35.40	48%

Sample	Barcode sequence	PF Clusters	% of the lane	% Perfect barcode	% One mismatch barcode	Yield (Mbases)	% PF Clusters	% >= Q30 bases	Mean Quality Score	%GC
F2_C53	TAGTGGCAA+ AACTGTCCT	1,657,060	1.02	92.89	7.11	414	89.22	91.81	35.26	49%
F2_C56	CATTAACGC+ AACTGTCCT	1,793,743	1.11	93.05	6.95	448	89.20	92.03	35.31	48%
F2_C58	TCGTTGAAG+ AACTGTCCT	1,650,872	1.02	92.78	7.22	413	89.65	91.88	35.27	48%
F2_C67	TAGTACGCT+ AACTGTCCT	1,602,522	0.99	92.78	7.22	401	89.43	92.04	35.30	49%
F2_C72	TTCACCGTA+ AACTGTCCT	1,673,617	1.03	92.81	7.19	418	89.64	92.21	35.34	48%
F2_C73	AGGACAGTT+ AACTGTCCT	1,891,498	1.17	92.83	7.17	473	90.39	92.44	35.38	47%
F2_C74	TACGAAGTC+ TCCACGATT	1,687,039	1.04	93.67	6.33	422	87.93	91.86	35.30	48%
F2_C75	GACGAGATT+ TCCACGATT	1,640,670	1.01	92.19	7.81	410	88.06	91.43	35.21	48%
F2_C78	TAGTGGCAA+ TCCACGATT	1,496,248	0.92	92.84	7.16	374	86.95	90.37	35.01	49%
F2_C80	CATTAACGC+ TCCACGATT	1,542,903	0.95	93.16	6.84	386	87.08	91.10	35.15	48%
P_SP1916	TCGTTGAAG+ TCCACGATT	1,420,475	0.88	93.68	6.32	355	87.39	90.77	35.09	48%
P_SPWAI	TAGTACGCT+ TCCACGATT	1,252,587	0.77	92.72	7.28	313	87.15	90.65	35.06	50%

Sample	Barcode sequence	PF Clusters	% of the lane	% Perfect barcode	% One mismatch barcode	Yield (Mbases)	% PF Clusters	% >= Q30 bases	Mean Quality Score	%GC
T_ST-R	TTCACCGTA+ TCCACGATT	1,376,116	0.85	93.14	6.86	344	87.55	91.46	35.21	50%
T_ST-L	AGGACAGTT+ TCCACGATT	1,464,987	0.90	92.96	7.04	366	88.46	92.00	35.32	49%
F2_D3	TACGAAGTC+ GACTGATAC	1,424,221	0.88	92.63	7.37	356	88.93	90.82	35.07	49%
F2_D8	GACGAGATT+ GACTGATAC	1,576,184	0.97	92.02	7.98	394	89.53	91.35	35.18	48%
F2_D11	TAGTGGCAA+ GACTGATAC	1,496,976	0.92	92.55	7.45	374	89.45	91.28	35.16	49%
F2_D13	CATTAACGC+ GACTGATAC	1,606,259	0.99	92.89	7.11	402	88.75	91.31	35.17	49%
F2_D14	TCGTTGAAG+ GACTGATAC	1,912,920	1.18	92.91	7.09	478	89.37	91.54	35.21	49%
F2_D15	TAGTACGCT+ GACTGATAC	1,823,843	1.13	93.10	6.90	456	89.65	91.71	35.24	49%
F2_D19	TTCACCGTA+ GACTGATAC	1,742,664	1.07	92.86	7.14	436	89.06	91.15	35.14	48%
F2_D21	AGGACAGTT+ GACTGATAC	1,537,134	0.95	91.97	8.03	384	89.92	91.27	35.16	49%
F2_D26	TACGAAGTC+ ACATCTGCT	1,864,945	1.15	90.89	9.11	466	89.39	92.03	35.31	49%
F2_D29	GACGAGATT+ ACATCTGCT	1,547,223	0.95	90.12	9.88	387	89.29	91.50	35.21	48%

Sample	Barcode sequence	PF Clusters	% of the lane	% Perfect barcode	% One mismatch barcode	Yield (Mbases)	% PF Clusters	% >= Q30 bases	Mean Quality Score	%GC
F2_D31	TAGTGGCAA+ ACATCTGCT	1,713,825	1.06	91.76	8.24	428	89.08	91.79	35.27	48%
F2_D33	CATTAACGC+ ACATCTGCT	1,872,107	1.15	91.21	8.79	468	88.57	91.76	35.26	48%
F2_D35	TCGTTGAAG+ ACATCTGCT	1,522,605	0.94	91.44	8.56	381	88.99	91.32	35.17	48%
F2_D36	TAGTACGCT+ ACATCTGCT	1,731,101	1.07	90.87	9.13	433	89.21	92.05	35.31	48%
F2_D41	TTCACCGTA+ ACATCTGCT	1,622,973	1.00	91.04	8.96	406	88.75	91.46	35.20	48%
F2_D45	AGGACAGTT+ ACATCTGCT	1,559,883	0.96	90.42	9.58	390	89.45	91.55	35.22	48%
F2_D46	TACGAAGTC+ ACGTTAGGA	1,664,985	1.03	82.19	17.81	416	90.00	91.30	35.16	49%
F2_D51	GACGAGATT+ ACGTTAGGA	1,586,122	0.98	82.07	17.93	397	90.72	92.05	35.31	48%
F2_D56	TAGTGGCAA+ ACGTTAGGA	1,633,775	1.01	84.29	15.71	408	90.11	91.52	35.20	49%
F2_D57	CATTAACGC+ ACGTTAGGA	1,591,009	0.98	82.38	17.62	398	89.88	91.40	35.18	48%
F2_D60	TCGTTGAAG+ ACGTTAGGA	1,563,590	0.96	84.05	15.95	391	90.31	91.01	35.11	48%
F1_F1 2-9	TAGTACGCT+ ACGTTAGGA	1,542,194	0.95	82.21	17.79	386	89.84	90.89	35.08	49%

Sample	Barcode sequence	PF Clusters	% of the lane	% Perfect barcode	% One mismatch barcode	Yield (Mbases)	% PF Clusters	% >= Q30 bases	Mean Quality Score	%GC
F1_F1 1-3	TTCACCGTA+ ACGTTAGGA	1,358,918	0.84	81.16	18.84	340	89.70	89.82	34.88	48%
F1_F1 1-18	AGGACAGTT+ ACGTTAGGA	1,606,770	0.99	82.05	17.95	402	91.02	91.77	35.25	49%

PF clusters is the percentage of clusters passing filtering, Yield is the number of bases sequenced, %>=Q30 is the percentage of bases with a quality score of 30 or higher

5 Chapter Five – QTL analysis of the *Sophora* Lincoln F_2 population

5.1 Abstract

Divarication is a form of plant architecture that occurs at a higher frequency in the New Zealand flora than elsewhere, however, little is currently known about the genetic architecture behind this growth form. An existing segregating population was developed ~40 years ago by Dr. E.J. Godley; made from reciprocal crosses between the *Sophora* species *S. prostrata,* a divaricate, and *S. tetraptera,* a non-divaricate. This population is segregating for the divaricate growth form and provides an ideal resource to investigate the genetic architecture of divarication. Phenotyping and genotyping of this population has been performed and linkage maps developed for QTL analysis. Twelve phenotypic traits were analysed and nine of these identified QTL. Multiple QTL were located for seven of the nine traits and each located at least one major QTL (>10% variation explained). Some QTL were located at the same position for multiple traits suggesting they may be pleiotropic loci. The majority of the QTL showed partial or complete dominance for the *S. prostrata* allele. These results provide the first glimpse into the genetic architecture of plant divarication and suggest multiple loci contribute to producing the divaricate growth form.

5.2 Introduction

Plant architecture describes the three-dimensional structure of the plant and includes the branching pattern, size, shape and position of leaves and reproductive structures (Reinhardt & Kuhlemeier, 2002). Plant architecture is an important component of taxonomy, agriculture and horticultural yield, and individual fitness. The development of body architecture in plants is more labile than for animals (Niklas, 2000) and therefore can be more influenced by the environment; however, it is still under strong genetic control. A body of research has identified some of the potential pathways or mechanisms that underlie plant architecture (as reviewed in Reinhardt & Kuhlemeier, 2002, Costes & Gion, 2015, Hollender & Dardick, 2015 and Chomicki et al., 2017). These include the integration of axillary branching pathways with *BRC1*, the *Arabidopsis* genes *SOC1* and *FUL,* which as double mutants, revert *Arabidopsis* from an annual herb to a woody perennial-like plant (Melzer et al., 2008), and microRNAs, known to be a conserved mechanism involved in phase change in plants (Poethig, 2009; Wu et al., 2009; Jung et al., 2011; Wang et al., 2011; Chomicki et al., 2017). However, the genetic basis behind many plant architectures remain uncertain and many different forms of plant architecture are not yet well understood.

One plant architecture, for which little is known of the genetic basis, is the divaricating form. Divarication is a plant growth form generally described as small-leaved shrubs or tree juveniles with wide branch angles, interlacing branches, long internodes and weak apical dominance (Kelly, 1994). Divaricating species are observed throughout the world but are at greatest frequency in New Zealand, where they make up ~10% of the woody flora (Greenwood & Atkinson, 1977; Atkinson & Greenwood, 1989). The relatively infrequent occurrence of divaricating species in 18 different plant families in New Zealand suggests that this growth form has evolved independently multiple

times. Many hypotheses have been proposed to explain the frequency of divaricate species in the New Zealand flora, such as moa browsing (Went, 1971; Carlquist, 1974; Livingstone, 1974; Greenwood & Atkinson, 1977) or climatic-driven hypotheses (Diels, 1897; Cockayne, 1912; McGlone & Webb, 1981; Lusk & Clearwater, 2015; Lusk et al., 2016) but there has been little focus and no published research on the genetic basis of this unique growth form.

Among the New Zealand genera with divaricate species is *Sophora* (Fabaceae). There are currently eight native New Zealand *Sophora* species recognised, mostly consisting of large non-divaricating trees, however, there is one obligate divaricate species, *Sophora prostrata* Buchanan, and a heteroblastic species with a divaricating juvenile form, *S. microphylla* Aiton. The origins of the heteroblast, *S. microphylla,* were of interest to Dr. E. J. Godley who investigated the hypothesis that *S. microphylla* originated from hybridization between *S. prostrata* and *S. tetraptera* J. S. Mill (Godley, 1975, 1985), the latter of which has an open architecture. To this end Godley created an F_2 population beginning with reciprocal crosses between *S. prostrata* and *S. tetraptera.* The F_2 population has been maintained by Manaaki Whenua Landcare Research but has been unutilised for decades. Fortuitously, this population shows segregation for the divaricate growth form, providing an ideal resource for studying the genetic basis of this intriguing plant architecture.

Sophora tetraptera naturally occurs in the North Island of New Zealand ranging from East Cape to Wairarapa, spreading west to Taihape, Lake Taupo and along the Waikato river (de Lange, 2004). It is mainly a lowland species occurring along streams, forest margins, and coastal forested habitats (Allan, 1961). *Sophora prostrata* is restricted to the eastern part of the South Island and typically inhabits grassland and rocky habitats ranging from lowland to montane areas (Allan, 1961).

The phylogenetic relationship between *S. prostrata* and *S. tetraptera,* indeed among all New Zealand *Sophora* species, still remain unresolved. *Sophora prostrata* has been identified as the most differentiated in some studies using chloroplast markers (Song, 2005; Shepherd et al., 2017); however, ISSR markers show admixture between *S. prostrata, S. microphylla* and *S. tetraptera* (Grierson, 2014). A more recent study using 12 microsatellite markers and more than 626 individual samples showed some New Zealand *Sophora* species (*S. prostrata, S. chathamica, S. fulvida* and *S. longicarinata*) form distinct groups, however, other species (*S. tetraptera, S. godleyi, S. microphylla* and *S. molloyi*) do not (Heenan et al., 2018). More work is needed to resolve the phylogenetic relationships among *Sophora* species including the relationship between *S. prostrata* and *S. tetraptera.*

The work described here reports on the first genetic analysis of the divaricating growth form. To characterise the genetic architecture of the divaricating form in *Sophora,* a quantitative trait loci (QTL) analysis was performed utilizing the decades old *Sophora* F_2 population. QTL analysis aims to identify regions of the genome containing polymorphisms involved in the traits of interest. This requires a segregating population with phenotype and genotype data for a population. The genotype data are used to generate linkage maps, which indicate the order and position of markers along chromosomes calculated from recombination frequencies (Chapter 4). QTL analysis is based on detecting associations between the genotype and the phenotype in segregating populations, which indicate that a locus is linked to a causal variant if there is a significant difference in the trait(s) of interest for alternative genotypes (Collard et al., 2005). Potential candidate genes can also be added to linkage maps as markers and could provide evidence of their involvement if they correspond with a QTL location.

As nothing is known about the genetic basis of divarication, it is unknown if this growth form evolved from a single gene, a few genes or from multiple genes. However,

segregation of divarication in the Lincoln and PN F_2 suggest that more than one gene is involved and the correlations among divaricate traits suggests these traits are under control from the same genes (Chapter 2). *RMSI* is a gene in the biosynthesis pathway of strigolactone that displayed associations with many divaricating traits and was predicted to have a deleterious amino acid replacement in the *S. prostrata* ortholog. Based on this, the divaricate growth form may be controlled by a few pleiotropic loci, one of which may be *RMSI*. To the best of our knowledge, this study is the first genetic investigation into the divaricating habit and attempts to characterise the genetic architecture of the divaricate growth form and identify candidate genes. Phenotyping of the *Sophora* $F_{2,}$ genotyping, development of gene markers, and generation of linkage maps was described in Chapters 2, 3 and 4.

5.3 Materials and Methods:

5.3.1 Study population

The F_2 population formed from reciprocal crosses between *S. prostrata* (divaricate) and *S. tetraptera* (non-divaricate) used in this QTL mapping of divaricate-related traits was previously described in Chapters 1 and 2.

5.3.2 Phenotyping and genetic map construction.

The F_2 population was phenotyped for nine traits related to divarication described in Chapter 2. Genotyping of markers was performed for the population and genetic maps generated from these markers as described in Chapters 3 and 4. Three further traits were used in QTL analysis, which included principal components 1 and 2, and a divaricate index score. The principal components PC1 and PC2 were the two main axes of the principal component analyses for the nine traits measured in Chapter 2. PCA reduces the number of variables in the data set by converting it into a new smaller set of uncorrelated values, called principal components, where the first principal component (PC1) accounts for the greatest possible variance and subsequent components each account for the next highest variance following the condition that they are uncorrelated (orthogonal) to the preceding component. This allows the reduction of variables in the dataset while maintaining the majority of the information from the original data. The divaricate index score used the phenotypes rachis length, internode length, branch width and branch number to produce a single value to represent the degree of divarication of each sample. These phenotypes were chosen as representative of the divaricate growth form. To match the direction of difference (larger for *S. tetraptera* and smaller for *S. prostrata*) for the other traits, branch number and internode length were inverted, then all four of these phenotypes were scaled from 0-1 with *S. prostrata* phenotypes represented by the lower values and *S. tetraptera* phenotypes reaching values closer to 1.

Histograms were made using R v. 3.4.2 (R Development Core Team, 2016) and descriptive statistics were performed using the r:psych v. 1.8.3.3 package (Revelle, 2016).

5.3.3 QTL analysis

QTL analysis was carried out using the software R/qtl v. 1.42-8 (Broman et al., 2003) following guidelines from Broman and Sen (2009). Marker regression, interval mapping, non-parametric mapping and multi-QTL mapping methods were performed on the data. Marker regression divides samples into groups, based on the genotypes at a given marker, and analyses markers individually, comparing phenotype averages at each marker. Samples with missing data are excluded in marker regression, which can be a limitation in datasets with high amounts of missing data but was not a factor here as markers with more than 25% missing data were excluded from both generating the linkage maps and in QTL mapping. Interval mapping allows for consideration of missing genotypes by calculating genotype probabilities along linkage groups including genotype probabilities between markers. These two methods assume normal data distributions, however, when data do not follow normal distributions, non-parametric mapping can be applied (Broman & Sen, 2009). These methods map only a single QTL for each linkage group and cannot detect interactions among loci or multiple loci on a given linkage group, meaning only part of the underlying genetic architecture is revealed (Broman & Sen, 2009). Multi-QTL mapping selects a model that describes the set of QTL that is best supported by the data and can provide a more complete picture of the genetic architecture behind a trait (Broman & Sen, 2009) as it allows detection of more than one QTL per linkage group and interactions between QTL. It is performed by utilising a system of forward/backward QTL selection to build a model of significant QTL. The model is built utilising specific penalties for adding new QTL and QTL interactions where the penalties are determined with an *a priori* permutation test. To build the

model, an initial forward selection is applied, starting with a genome scan for individual QTL, then a scan for interacting QTL is performed. This is repeated up to a predetermined number of QTL and then backwards selection is performed back to the null model of no QTL, dropping one of the current main effect QTL or interactions at each step. The final model is chosen from this forward/backward selection based on the model with the highest penalised LOD score. Interval and multi-QTL mapping were performed in the *Sophora* F_2 using Hayley-Knott regression (Haley & Knott, 1992) to allow comparison between all methods, as both single and multi-QTL analyses can be useful to gain understanding of the genetic architecture of a trait. Hayley-Knott regression gives a fast estimate of the results from standard interval mapping; however, the way it treats missing data is not always ideal and the approximation of standard mapping can be poor in areas where widely spaced or incomplete markers occur. However, both standard (EM) and Hayley-Knott regression were performed on the Lincoln population with results showing acceptable approximation of the standard interval mapping by Hayley-Knott (Appendix 5.1) indicating Hayley-Knott regression is acceptable for use in this population. Conditional genotype probabilities were calculated using *calc.genoprob* for use in interval mapping and multi-QTL mapping. Interval mapping was performed using *scanone* for both EM and Hayley-Knott regression methods. Non-parametric mapping using *scanone* was performed for traits that did not identify QTL with interval mapping as the skew and kurtosis of trait distributions may affect mapping in these traits which assumes a normal distribution. The significance level (LOD threshold) for marker regression, interval mapping and non-parametric methods were determined using a permutation test of 1000 permutations of *scanone*. Penalties for multi-QTL mapping were generated by performing permutation test of 1000 permutations of *scantwo* and calculating penalties from these with *calc.penalties* for 0.05, 0.10, 0.15 and 0.20 alpha levels.

Degree of dominance of each QTL was calculated using the formula of Falconer (1964) as summarised in Stone (1968). Additive effects were calculated as half the difference between homozygote means. To characterise the effects of individual and all QTL for a trait, QTL effects were calculated with the function *fitqtl* from R/qtl. As the maternal direction marker consisted of only a single marker, it could not be included in this calculation therefore QTL effect of this marker was calculated by double the additive effect divided by the range in the trait. Confidence intervals were determined using the qtlTools v. 1.2.0 (Lovell, 2018) package function *calcCis* for Bayes credible intervals (95%). As the F_2 population consists of individuals derived from reciprocal crosses, a marker to represent cross direction, CP, was included as linkage group 10. Analyses also were performed with maternal direction as a covariate to account for maternal direction effects and included a 2D scan (Appendix 5.2).

Table 5.1. Methods of analysis applied for each of the traits indicating significant results.

Trait	Marker Regression	Interval Mapping	Non-parametric mapping	2-D scan mapping	Multi-QTL mapping	MQM
Leaflet Number	*	*	*		*	*
Rachis Length	*	*	*	*	*	*
Leaflet Length	*	*	*	*	*	*
Leaflet Width	*	*	*	*	*	*
Internode Length						
Branch Width						
Branch Number						
Node Angle	*	*	*		*	*
Branch Angle			*	*		*
PC1	*	*	*	*	*	*
PC2			*			
Divaricate Index			*			*

* indicates QTL detected at 5% significance

5.3.3.1 MQM analyses

MQM ("multiple-QTL models" or "marker-QTL-marker") (Jansen & Stam, 1994) is an interval mapping method where "important" markers are selected in marker

regression and these, except for those that flank the QTL interval under investigation, are used as cofactors. These cofactors reduce the majority of the variation induced by close QTL and reduces the chance of type I (QTL indicated when there is no actual QTL) and type II (no QTL detected) errors in standard interval mapping (Jansen, 1994). MQM mapping has been incorporated into R/qtl as R/qtl-MQM (Arends et al., 2010) with improved handling of missing data. MQM mapping in R/qtl was also performed on the *Sophora* Lincoln F_2 population as an alternative method for identifying potential QTL. Cofactors were selected using the *mqmautocofactors* function which utilises backwards elimination to identify a limited number of the most informative cofactors. An initial 50 markers, which included the maternal direction marker (CP), were selected, based on marker density, as potential cofactors using this function. The backwards elimination procedure was applied with an alpha of 0.02 which identified 1 – 11 cofactors among the traits that mapped QTL. A step size of 5cM was used for MQM analyses. LOD thresholds were determined for each trait using *mqmpermutation* with 1000 permutations. An MQM run without cofactors was also performed as recommended (Arends et al., 2010). Most QTL located by MQM, using cofactors, were also located in the multi-QTL analyses however, when no cofactors were included QTL identified were those located in R/qtl interval mapping methods indicating MQM may be able to account for effects of nearby QTL in interval mapping. However, MQM does not include model selection, as in the multi-QTL stepwise approach, meaning MQM can only reveal part of the underlying genetic architecture, being only able to identify one QTL for each linkage group and therefore cannot account for all interactions among QTL. Hence, the multi-QTL stepwise results are considered more representative of the genetic architecture.

5.4 Results:

5.4.1 Distributions for the three new summary phenotypes

The nine divaricate traits under investigation here are described in Chapter 2. Three new traits also were used for QTL analysis and include the principal component axes 1 and 2, and a divaricate index developed from phenotypes in Chapter 2. The distributions of PC1, PC2 and the divaricate index scores were nearly normally distributed. PC1 displayed a right skew and the divaricate score displayed a left skew (Figure 5.1).

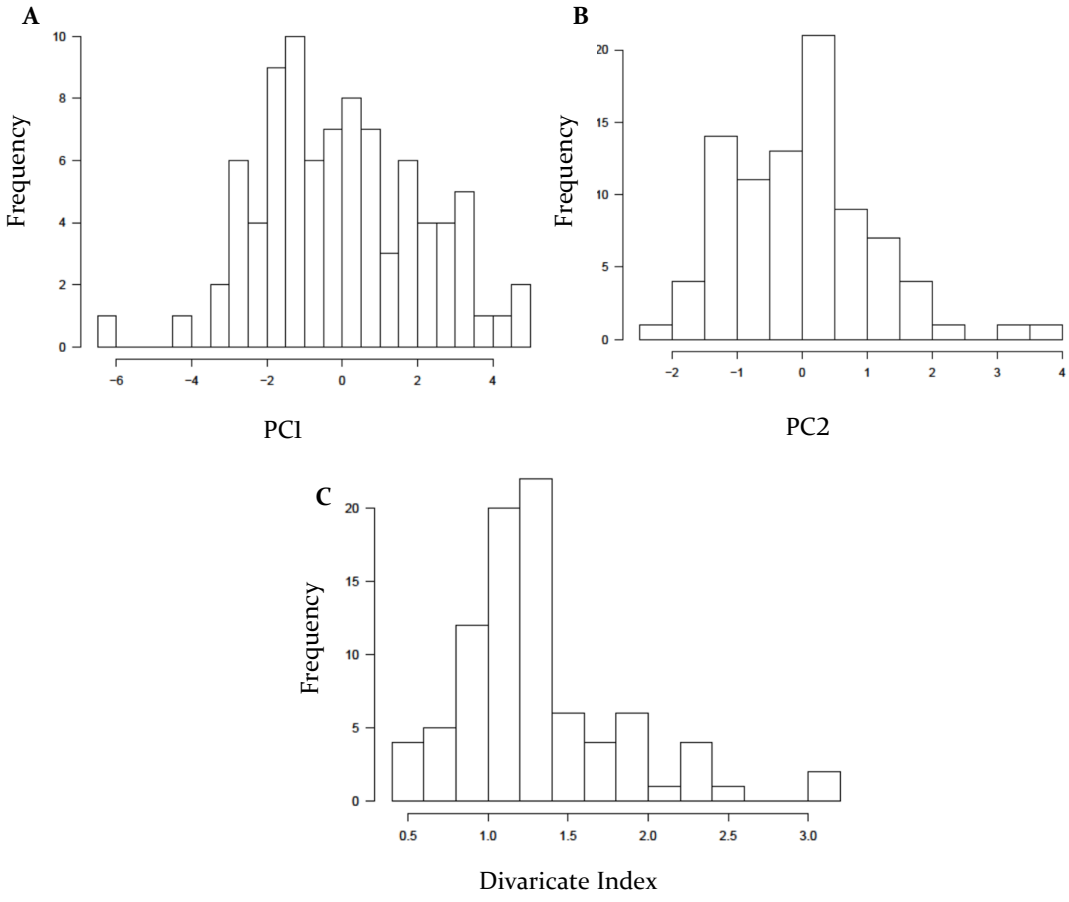


Figure 5.1. Histogram distributions for the new traits in the Lincoln F_2: A – PC1, B – PC2, C- Divaricate index

5.4.2 QTL mapping results

Multiple methods were used for identifying QTL in the *Sophora* F_2 population. These included single QTL analysis methods such as marker regression, interval

mapping, non-parametric mapping, MQM and multi-QTL mapping. These analyses identified a total of 35 QTL at an alpha of 0.05 (Table 5.1), covering all nine linkage groups (Figure 5.2), and five epistatic interactions among QTL. Some QTL co-located for multiple traits. Twenty-eight QTL were mapped for the original nine phenotypes from Chapter 2. Five were identified for the PC1 axis, and two QTL identified from non-parametric mapping, one each for branch angle and the divaricate index score. Most of the QTL identified by single QTL analysis were also observed using multi-QTL mapping (Appendix 5.1). There are a few cases of QTL differing between methods but this is most likely due to the way the methods analyse the data and may result in movement in the position of the QTL. Alternatively, one locus may no longer have support for a QTL when the first QTL is taken into account in multi-QTL models (Broman & Sen, 2009). For the multi-QTL mapping, at an alpha of 0.10, 40 QTL in total were identified and at an alpha of 0.20, 50 QTL were identified. Some QTL revealed allele effects opposite those expected based on the parental phenotypes; for example, some alleles in *S. prostrata* were observed to be associated with larger leaf size than the *S. tetraptera* allele. There are 15 QTL, seven among the phenotypic traits and eight for the combined trait indices, for which this occurred which are indicated in Table 5.2.

5.4.3 Multi-QTL mapping of leaf traits

Multiple QTL were identified for each of the leaf traits (Figure 5.2). For *leaflet number,* six QTL were identified; one of these QTL was mapped to the maternal direction linkage group (CP). The total amount of variation explained for leaflet number was ~70%. 'Major QTL' are often defined as those explaining more than 10% of the variation of a trait (Broman & Sen, 2009). With this, four of the five QTL, excluding the maternal direction, for leaflet number would be classified as major QTL (Table 5.2). The CP marker explains 14.5% of variation based on additive effect (see section 5.3.3).

Table 5.2. Multi-QTL results for the Lincoln *Sophora* F_2 population.

Phenotype	QTL locations	Alpha	Nearest marker	QTL effects (%)	Additive effects	Degree of Dominance
Leaflet number	Q1 – 1@25^{P}	0.05	P12	10.13	0.8604	-4.56890
	Q2 - 4@117^{P}	0.05	RMS4	6.69	0.9992	-0.82084
	Q3 – 5@30^{T}	0.05	L15	25.11	1.5916	0.35962
	Q4 - 5@60^{P}	0.05	C51	26.34	1.3244	-0.92839
	Q5 – 7@18^{P}	0.05	L81	10.61	1.6054	-0.77263
	Q6 – 10	0.05	CP	14.48	1.7824	
Rachis length	Q1 – 1@26^{P}	0.05	P12	12.36	4.1350	-3.91201
	Q2 – 4@118^{P}	0.05	RMS4	4.75	4.8993	-1.04010
	Q3 – 5@32^{T}	0.05	L15	10.43	7.1480	0.28125
	Q4 – 5@57.1^{P}	0.05	C51	10.70	6.0122	-0.23217
	Q5 – 5@150.5	0.05	C30	4.23	13.2284	-0.07642
	Q6 – 7@16^{P}	0.05	L17	9.25	3.7774	-0.39534
	Q7 – 8@1^{P}	0.05	L76	13.34	12.2409	-0.70607
	Q1-Q6	0.05		6.08		
Leaflet length	Q1 – 1@38^{P}	0.05	L10	5.57	0.8256	-2.85262
	Q2 – 2@73^{P}	0.05	L38	4.810	0.7360	-3.66598
	Q3 – 3@67.9^{P}	0.05	D64	8.32	1.0831	-0.29576
	Q4 – 4@127^{P}	0.05	L77	6.51	1.5969	-0.46236
	Q5 – 5@23$^{\circ, P}$	0.05	L84	14.32	0.5829	0.39789*
	Q6 – 5@151^{P}	0.05	C30	19.52	2.5065	-0.11755
	Q7 – 6@25$^{\circ,T}$	0.05	B21	5.03	0.0546	-26.73096*
	Q8 - 8@11^{P}	0.05	L3	9.61	2.0566	-0.38928
	Q5-Q6	0.05		10.83		
Leaflet width	Q1 – 2@75^{P}	0.05	L20	14.87	0.2316	-2.73243
	Q2 – 2@119.8$^{\circ, T}$	0.05	L68	25.89	0.1781	-3.29932*
	Q3 – 3@62^{P}	0.05	L64	14.66	0.1670	-0.26104
	Q4 – 3@134^{T}	0.05	L103	12.07	0.2438	0.35044
	Q5 – 7@103.1^{P}	0.05	L113	16.36	0.3011	-0.64906
	Q6 – 8@8.5^{P}	0.05	RMS1	12.17	0.4308	-0.64429
	Q3-Q5	0.05		10.31		
	Q2-Q4	0.05		11.94		

Phenotype	QTL locations	Alpha	Nearest marker	QTL effects (%)	Additive effects	Degree of Dominance
Branch Width	Q1 – 5@146^{T}	0.10	L63	15.31	0.2397	0.66642
	Q2 – 8@0^{P}	0.20	L76	11.76	0.2743	-0.17685
Node angle	Q1 – 1@144^{T}	0.05	L118	23.56	2.6147	0.45508
	Q2 – 3@107°,^{P}	0.10	L66	11.68	1.8298	1.85766*
	Q3 - 8@0^{P}	0.10	L76	11.14	2.1829	-0.40063
	Q4 – 5@150.5^{T}	0.20	C30	7.47	1.9687	0.29602
Branch Angle	Q1 – 1@62°,^{T}	0.15	B79	13.49	3.9420	-0.14038*
	Q2 – 4@53.1°, ^{P}	0.20	L25	14.74	2.7297	0.52984*
	Q3 – 6@17°, ^{T}	0.20	L12	15.74	3.5823	-0.33973*
	Q4 – 9@7.3^{P}	0.20	L99	10.31	0.1605	-22.3224
PC1	Q1 – 1@109.7 °, ^{T}	0.05	L39	20.26	1.0368	-0.35685*
	Q2 - 2@29°, ^{P}	0.05	L16	15.87	0.5911	0.45089*
	Q3 – 5@139°, ^{T},	0.05	L63	25.52	1.4774	-0.49972*
	Q4 – 8@1°,^{P}	0.05	L76	9.57	1.4319	0.40130*
	Q5 – 9@11.8°, ^{P}	0.05	L82	8.35	0.1754	7.946466*
	Q1-Q2	0.05		11.673		
PC2	Q1- 6@19.4°, ^{T}	0.10	L78	19.49	0.6025	-0.66504*
	Q2 – 9@9^{P}	0.10	L99	14.08	0.2794	-2.44384
	Q3 – 1@61°, ^{P}	0.20	B1	17.73	0.3972	1.281742*
	Q4 - 4@83.6°, ^{T}	0.20	L37	9.82	0.2157	-0.37088*
	Q5 - 9@30^{P}	0.20	L18	13.83	0.0420	-9.65641
Divaricate Index	Q1 – 5@138^{T}	0.10	L63	16.51	0.2487	0.11877
	Q2 – 9@11.8^{P}	0.15	L82	11.80	0.0238	-13.1858
	Q3 – 1@122^{T}	0.20	L35	10.27	0.1980	0.67341

° indicates loci with the opposite effect on phenotype than expected of the parental allele, * indicates sign changed so *S. tetraptera* dominant alleles are always positive for Figure 5.3,^{P} indicates partial or complete dominance of the *S. prostrata* allele, ^{T} indicates complete or partial dominance of the *S.tetraptera allele*, QTL effects are based on additive effects as described in section 5.3.3.

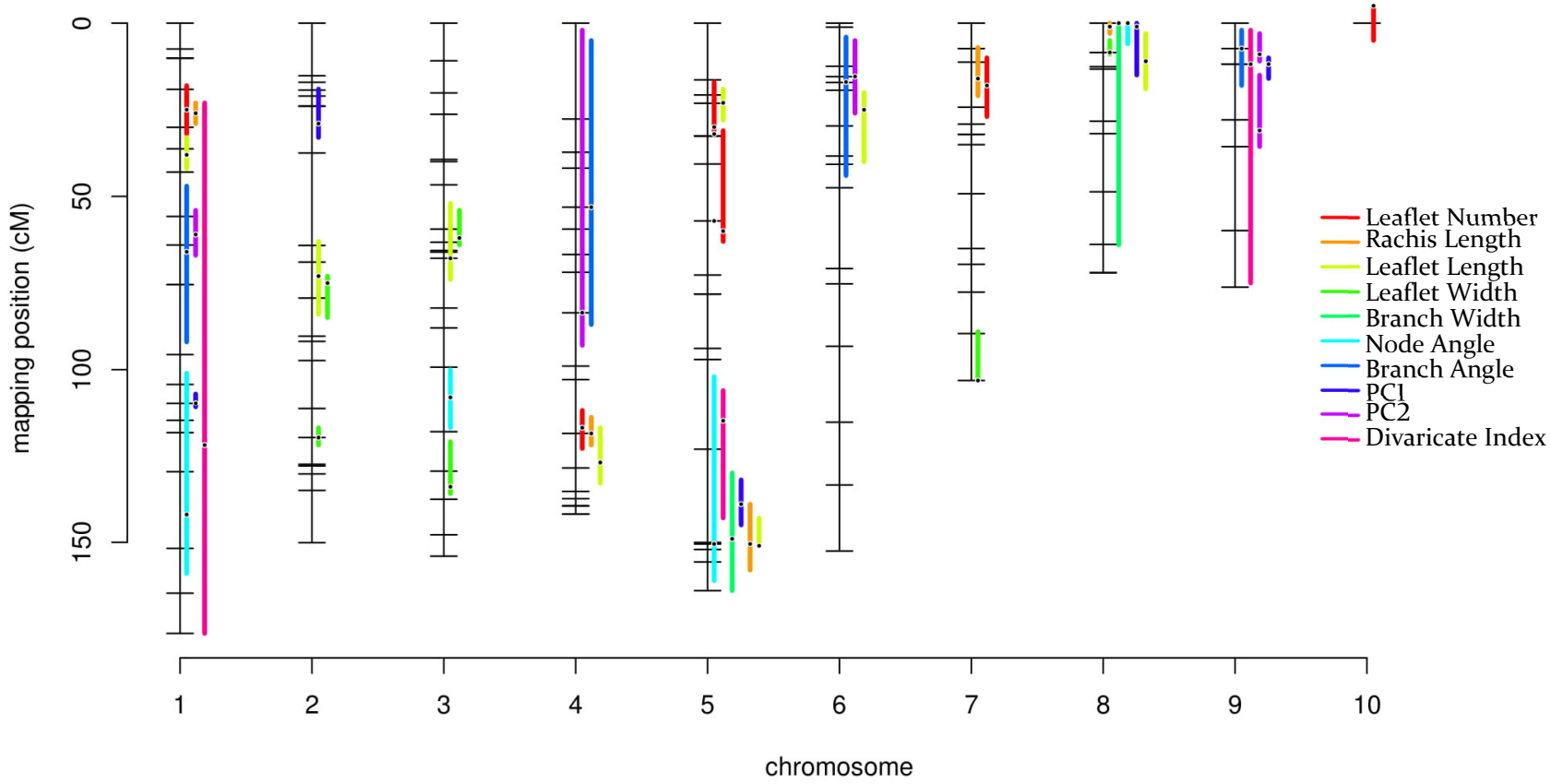


Figure 5.2. Plot of the QTL for alpha levels 0.05-0.20 plotted on the *Sophora* linkage groups. Colours represent phenotypes as indicated on the legend. Coloured lines indicate confidence intervals for a QTL position. Overlapping confidence intervals can be indicative of those traits that map to the same locus.

Multi-QTL mapping for *rachis length* yielded seven QTL that explained over 89% of the total variation for this trait and included an interaction between Q1 and Q6. The high total variation of some traits may result from overestimation of the effects of QTL (Broman & Sen, 2009) and from the small sample size. Four major QTL were identified and the QTL with the highest effect was Q7 at 13.34%.

Eight QTL were mapped for *leaflet length* with two major QTL identified. Q6 explained the highest percent of variation at 19.52%. An interaction between Q5 and Q6 was also observed in *leaflet length*. The total variation explained for leaflet length was ~86%. *Leaflet width* identified six QTL and two interactions between QTL. Each QTL explained more than 10% of the variation with a total explained variation of 82%. One QTL explained more than 20% variation which was Q2 at 25.9%.

5.4.4 Multi-QTL mapping of branch traits

Of the branch traits, no QTL were identified for *internode length*, and *branch number* at any alpha level (Appendix 5.1). *Node angle* identified one major QTL at an alpha of 0.05, on linkage group 1 that explained 23.6% of the variation, two major QTL at alpha of 0.10 and another QTL at an alpha of 0.20. Including the QTL identified at an alpha of 0.20, the total explained variation for *node angle* was 53.96%. *Branch width* and *branch angle* did not identify QTL at alpha of 0.05 but at an alpha of 0.10 *branch width* identified one major QTL with a second major QTL identified at alpha of 0.20. The total explained variation for *branch width* was 27.7%. *Branch angle* mapped one major QTL at 0.15 alpha level and three further QTL at an alpha of 0.20 at which the total variation for *branch angle* was 45.57%. Non-parametric mapping also identified the same major QTL with an alpha of 0.05 as the multi-QTL at alpha of 0.15 for *branch angle* on linkage group 1 that explained 14.48% of the variation (Table 5.3, Figure 5.2).

Table 5.3. QTL mapping results for non-parametric mapping in branch angle and divaricate index score.

Phenotype	QTL location	Nearest marker	Effect (%)	Additive effects	Dominance effects
Branch angle	Q1 – 1@61°, P	B79	14.48	4.0343	0.18710
Divaricate index	Q1 – 1@24P	L115	15.28	0.0780	-4.7975

° indicates loci with opposite effect on phenotype than expected of the parental allele. P indicates partial or complete dominance of the S. prostrata allele

5.4.5 Multi-QTL mapping of combined data traits

The *PC1* axis mapped five QTL (Table 5.2). At an alpha of 0.05, three of these were unique, Q1, Q2 and Q5, and two, Q3 and Q4, that had overlapping confidence intervals with QTL for some of the leaf traits (Figure 5.2). At higher alpha levels only one of these QTL, Q2, was unique as it did not have overlapping confidence intervals with other traits. There was also an interaction observed between Q1 and Q2. The total explained variation for *PC1* was 75.24% and three QTL were major QTL, with Q1 and Q3 each explaining over 20% of the variation at 20.26% and 25.52%, respectively. *PC2* axis failed to identify any QTL at an alpha of 0.05 but mapped two QTL at an alpha of 0.10 and three more at 0.20 alpha level. Four of the five QTL for PC2 were located with QTL for branch angle. The total variation of *PC2* at an alpha level of 0.20 was 55.4% with four major QTL identified. The *divaricate index score* also failed to detect QTL at an alpha of 0.05 but mapped three QTL, at 0.10, 0.15 and 0.20 alpha levels. The total explained variation of the *divaricate index score* was 37.7%. Non-parametric mapping also identified a major QTL for the divaricate index score on linkage group 1 that explained 15.28% variation that overlaps confidence intervals with Q3 at alpha of 0.20 (Table 5.3, Appendix 5.1).

5.4.6 Co-locating loci for divarication in *Sophora*

At an alpha of 0.05, there are nine QTL locations that suggest pleiotropic effects among some leaf traits with QTL that had overlapping confidence intervals for at least two different traits (Figure 5.2). *Leaflet number* and *rachis length* have four such suggestive QTL. *Leaflet length* and *width* suggest three possible pleiotropic QTL. *Leaflet number* and *leaflet length* map to one QTL at the same locus and *rachis length* and *leaflet length* also suggest one possible pleiotropic QTL. Finally, one QTL appears to be in common for *leaflet number, rachis length,* and *leaflet length*.

There were no QTL suggestive of pleiotropy for leaf and branch traits at an alpha level of 0.05. At 0.10 alpha level two loci mapping multiple leaf traits also map with a branch trait. *Branch width* maps with *rachis length* and *leaflet length* on linkage group 5 while *node angle* maps with *rachis length, leaflet length* and *leaflet width* on linkage group 8. *PC1* also maps to both these loci. At an alpha of 0.20 both *node angle* and *branch width* map to these two loci on linkage groups 5 and 8. These two QTL locations, linkage group 5 ~150cM and linkage group 8 ~0cM, were identified as major QTL, explaining over 10% of the variation in most of these traits. *Branch angle* also maps to a locus with *leaflet length* on linkage group 6.

5.4.7 Allelic dominance at QTL

The degree of dominance is indicated for each QTL in Table 5.2. A value of 0 indicates no dominance, a negative value indicates dominance of the *S. prostrata* allele and a positive value indicates dominance of the *S. tetraptera* allele. Values of 1 or -1 indicate complete dominance of an allele and values beyond these indicate overdominance in the heterozygotes. A large proportion of QTL show partial or complete dominance for the *S. prostrata* allele, especially in the leaf traits (Table 5.2 and Figure 5.3). Ten QTL, five for leaf traits and five for branch traits, were dominant

towards the *S. tetraptera* allele, while twenty-six QTL were dominant towards the *S. prostrata* allele; 21 from the leaf traits and five from the branch traits. The QTL located near *RMSI* on linkage group 8 were dominant for the *S. prostrata* allele in all traits mapped there. Fourteen QTL showed overdominance, indicated with a degree of dominance either above 1 or below -1, where the heterozygote phenotype mean is beyond the homozygote means of the parents (Examples in Figure 5.4). Of these, twelve displayed overdominance beyond the *S. prostrata* allele mean and two were overdominant beyond the *S. tetraptera* allele mean.

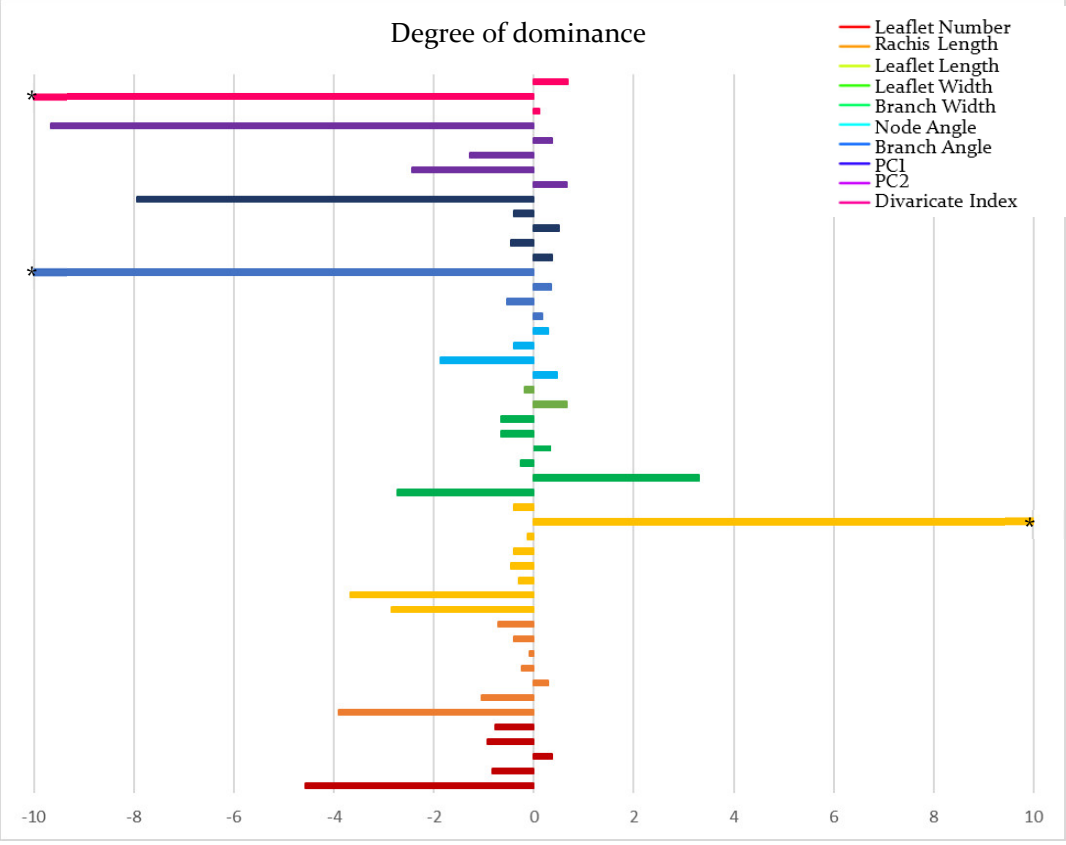


Figure 5.3. Bar chart displaying the degree of dominance for QTL in the *Sophora* Lincoln F_2 population. Leaflet number (red), rachis length (orange), leaflet width (yellow), leaflet width (green), branch width (dark green), node angle (light blue) and branch angle (blue), PC1 (dark blue), PC2 (purple) and the divaricate index (pink). Asterisks indicate dominance values off scale of the chart (values are in Table 5.2) occurring as an effect from the calculation when homozygote phenotypes were similar but heterozygote values much higher or lower.

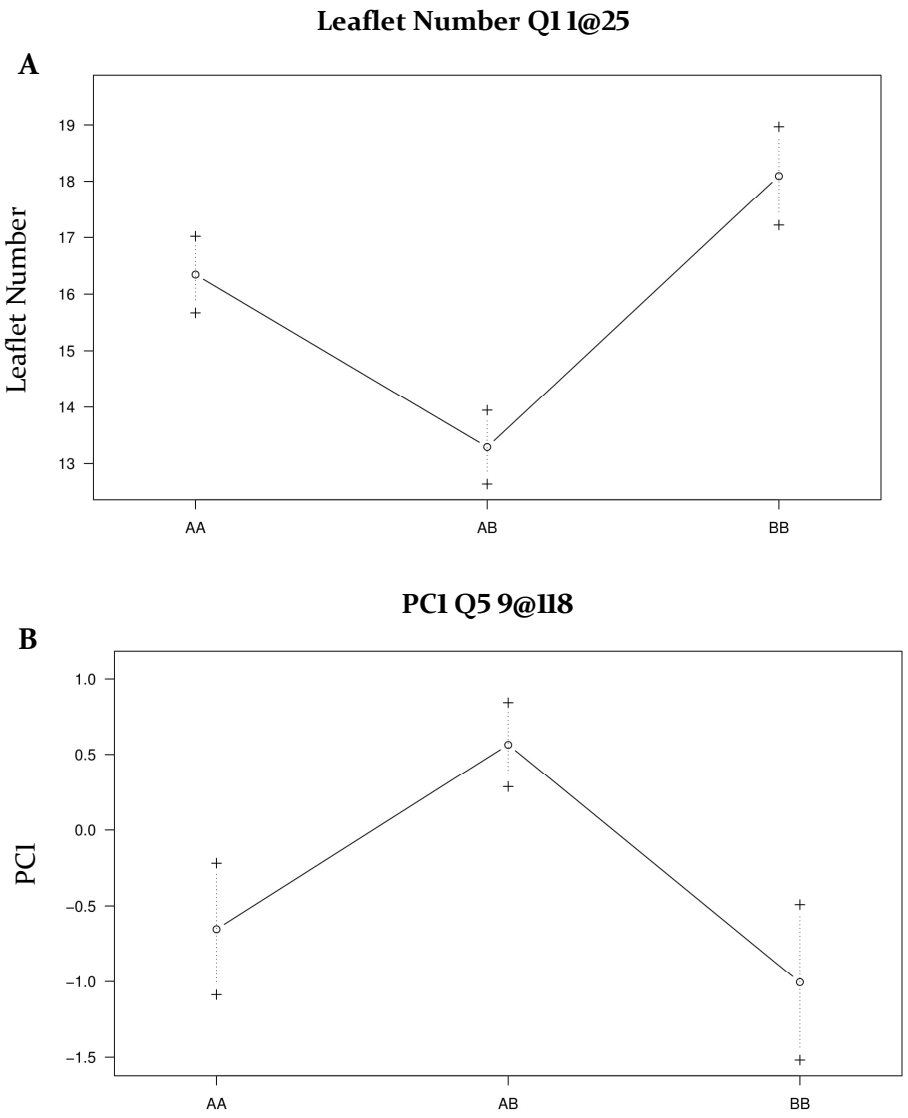


Figure 5.4. Examples of QTL showing under and over dominance. Q1 of leaflet number showing underdominance (A). Q5 of PC1 showing overdominance (B). AA represents the S. prostrata homozygote, AB represents the heterozygote and BB represents the S. tetraptera homozygote.

5.4.8 Gene markers

The candidate genes *RMSI, 2, 4, 5* and *ZIG* were developed into genetic markers (described in Chapter 3) and included in the linkage maps for QTL analysis. The *RMS4* locus was associated with a QTL for both *leaflet number* and *rachis length,* with QTL effects of 6.69% and 4.75%, respectively. This marker was also located within the confidence interval for *leaflet length* which suggests *RMS4,* or a closely linked gene, may

be involved in variation for these traits. The *RMS1* marker was associated with a QTL for *leaflet width* and is within the confidence interval for *leaflet length* with effects of 12.17% and 9.61%, respectively at an alpha level of 0.05. A QTL for *rachis length* also mapped near the *RMS1* marker and explained 13.34% of the variation for this trait. The *PC1* trait also mapped near *RMS1,* which is expected as *PC1* is largely based on the correlated leaf traits so similar loci to the leaf traits would be expected for this trait. At an alpha of 0.10, *node angle* also mapped to *RMS1* which explained 11.14% of the variation for this trait and at an alpha of 0.20 *branch width* also mapped to *RMS1* explaining 11.76% of the variation for this trait. Finally, a single QTL for *PC1* was mapped near *RMS2*. At an alpha of 0.05, no QTL were associated with the *RMS5* or *ZIG* gene markers.

5.5 Discussion

Divarication is a unique growth form found in the New Zealand flora at a relatively high frequency when compared to other floras (~10% of the woody flora). It is defined as a plant with increased branching, wide branch angle and smaller leaves (Kelly, 1994). Research on the evolutionary origins of this growth form has, thus far, focused on why divarication evolved; however little to no research has been performed to investigate the genetic basis of divarication. Here, we have made a start in addressing this gap in understanding by utilising an F_2 population formed from reciprocal crosses between the divaricate, *S. prostrata* and the non-divaricate, *S. tetraptera* to investigate the genetic architecture of divarication in *Sophora.* Multiple QTL were identified that contribute to divarication and many QTL mapped multiple traits suggesting they may be pleiotropic. As well, many QTL displayed dominance of the *S. prostrata* allele. The QTL mapping approach taken here utilised the phenotype data described in Chapter 2 and the genotype and linkage map data described in Chapter 4.

5.5.1 QTL mapping in the Lincoln F_2 population

With essentially no genetic research carried out on the divaricate growth form prior to this work, it has been unclear whether one, few, or multiple loci are involved in the evolution of divaricates. The precise and consistent morphology of divarication has proven difficult to define across all plant groups (Kelly, 1994) due to variability in the growth form among species, perhaps indicating that multiple genes contribute to the growth form in each plant group. The results from the QTL analyses in the Lincoln F_2 population suggest multiple genes are indeed involved in not only the divaricating form but also for each of the traits that contribute to the divaricating form. The implication that multiple traits, some largely under their own genetic control, contribute to the divaricate growth form would indicate that for the divaricate phenotype to evolve,

selection on the traits would have to occur simultaneously. This supports the idea that environmental factors, either climatic or predatory, could have been important factors in the evolution and maintenance of divarication in *Sophora.* The total number of QTL identified at an alpha of 0.05 was 35, with 40 QTL identified at an alpha of 0.10 and 50 at an alpha of 0.20. Many QTL were identified for multiple traits suggesting pleiotropy is important in divarication for *Sophora*. The QTL located at alpha levels 0.10-0.20 still indicate large effects (over 10% variation explained), indicating they contribute a significant amount to each trait but, as they are not identified at an alpha of 0.05, may only be suggestive of QTL.

5.5.2 Genetic basis of leaf trait variation is complex

One of the key traits used in defining divarication is a relatively small leaf size, which has been hypothesised to increase light harvesting (Kelly, 1994), enabling plants to survive in conditions where this may be an issue, for example in forest understory. Alternatively, it is also hypothesised to provide protection to inner leaves by the outer shield of branches, produced in the divaricate form, allowing plants to survive in harsher weather conditions (McGlone & Webb, 1981). *Sophora prostrata* has been described as the only divaricating species to consistently have compound leaves (Tomlinson, 1978); yet these too are still much smaller in size than the leaves of the non-divaricate congener. All leaf traits in the *Sophora* F_2 population were positively correlated with each other (see Chapter 2), suggesting they may be under common genetic control. Supporting this, many QTL co-located for multiple leaf phenotypes; however, no QTL mapped to all leaf phenotypes and many QTL had effects on only one leaf trait.

The leaflet number and rachis length traits map to the same locus at three different QTL (Figure 5.2). As leaflet number is likely to be partially dependent on rachis length, these traits are expected to be developmentally linked, thus this result is not

surprising. However, both traits also mapped multiple independent QTL as well. At an alpha of 0.05, six QTL were mapped for leaflet number, with additive effects ranging from 0.86-1.60 and seven QTL were mapped for rachis length, with additive effects ranging from 4.1-13.2. These findings are consistent with the results from other plant species, where leaf length and width have been shown to be controlled by multiple QTL with smaller to moderate effect sizes (Kim et al., 2005; Gailing, 2008; Tan et al., 2016; Wang et al., 2018). Leaflet number has been studied in Soybean, *Glycine max* (L.) Merr (Fabaceae) (Fehr, 1972) and Mungbean (*Vigna radiata* (L.) Wilczek (Fabaceae)*)* (Soehendi et al., 2007) as well, where only a single locus in soybean and two loci in Mungbean were identified to be involved in leaflet number. The larger number of QTL (five) located for leaflet number in *Sophora* is likely a result of segregation in an F_2 population derived from parents with very different leaflet number phenotypes. Leaflet number was also observed to be affected by cross direction, which explained 14.5% of the variation.

A similar pattern was also observed for leaflet length and width: these two traits also have been shown to be positively correlated (Chapter 2) and have three overlapping QTL. For the eight QTL mapped for leaflet length, the range of additive effects is 0.05 – 2.5 and for the six QTL mapped for leaflet width, the additive range is 0.17 – 0.43%. Unsurprisingly leaf length and width have been shown to map together in other species as well, including *Saccharina japonica* (Wang et al., 2018), *Glycine max* (L.) Merr (Fabaceae) (Kim et al., 2005) and *Camellia sinensis* (L.) Kuntze (Theaceae) (Tan et al., 2016). As was found here, in most of these cases, independent QTL were mapped for each trait, suggesting this complex architecture is often observed for genetic control of leaf traits.

These results show that leaf shape and size determination is genetically complex in *Sophora,* which is consistent with the results of many other studies of complex leaf phenotypes with many genes of small effect contributing to the overall trait variation (Welter et al., 2007; Gailing, 2008; Tan et al., 2016). The co-localisation of QTL for correlated traits is consistent with other studies showing that correlated traits often map together (Xu et al., 2015; Chunlian et al., 2016; Wang et al., 2018). Individual QTL for the leaf traits have effects that are similar to the range of effects of QTL for similar growth traits in *Salix* (Salicaceae) (Tsarouhas et al., 2002; Weih et al., 2006), leaflet length and width in *Glycine max* (L.) Merr (Fabaceae) (Kim et al., 2005), and leaf traits in *Quercus robur* (Fagaceae) (Gailing, 2008).

The total variation explained for each of the leaf traits was large at over 70%, however, the trait variation explained by a QTL is generally overestimated (Broman & Sen, 2009). Regardless of the problems inherent in estimating effect size for individual loci, it is clear that a substantial portion of leaf variation in the *Sophora* F_2 population is under genetic control, with little modification by the environment of the F_2 field plot. Consistent with this, in *Rosa* (Rosaceae), the environmental effects on leaflet number were determined to be not significant based on three different field locations, showing that the genotype was significant in determining leaflet number (Shupert et al., 2007). In addition to the multiple loci involved for *Sophora* and the large proportion of variation explained in individual QTL, QTL-by-QTL interactions are also important contributors to most of the leaf traits, as they were observed for rachis length, leaflet length and leaflet width, indicating an important role for epistasis in leaf size evolution.

As reduced leaf size is one of the key characteristics of the divaricate growth form, the characterisation of the genetic architecture of leaf size represents a critical step toward understanding how the divaricate form has evolved. These results show that leaf

size and leaflet number are genetically complex traits controlled by many individual loci with a wide range of effect sizes. The complexity is further increased by the finding of many QTL that appear to affect multiple traits, thus implicating pleiotropy, and QTL-QTL interactions, suggesting an important role for epistasis.

5.5.3 QTL mapping of branch traits

Similar to leaf size, the divaricating architecture is formed from contributions of multiple traits. As such, the genetic architecture of the branching traits is characterised by similar findings to those for the leaf traits: multiple QTL, a wide range of effect sizes, overlapping QTL among traits, and QTL-QTL interactions. However, far fewer QTL were identified for the branch traits and in many cases, a higher alpha level was required for the initial detection of QTL. It is not uncommon to explore multiple alpha levels in QTL mapping and our results at lower alpha levels would appear to be bolstered by the finding of QTL that often overlap with QTL mapped for other traits at more stringent alpha levels.

The branches of divaricate species are often described as thinner than related non-divaricate species. Two major QTL for branch width were mapped, one at an alpha level of 0.10 and the other at 0.20, both of which occur at QTL positions that overlap with QTL mapped for rachis length, leaflet width and node angle; one of these also maps to leaflet length. Branch width is correlated with the leaf traits and node angle (see Chapter 2) so common QTL for these traits may be expected. The total variance explained for branch width is 27.7% indicating (1) stronger environmental effects on branch width, (2) reduced power to detect QTL for this trait, (3) many smaller effect QTL could not be detected, or (4) some combination of 1-3 contributed to mapping in this trait. Branch diameter also has been mapped in a *Pinus elliotti x Pinus caribaea*

(Pinaceae) F_1 population where two closely linked QTL were identified that explained a combined 33% of the variation (Shepherd et al., 2002) similar to the results for *Sophora*.

Node angle contributes to the interlacing branch architecture of *S. prostrata* by producing the distinctive zig-zag stem which has been described as one of the possible contributors to interlacing of branches in divaricates (Tomlinson, 1978). While a zig-zag patterned stem is not observed in all divaricate species, it is a feature for some such as *S. prostrata* and *Muehlenbeckia astonii* Petrie (Polygonaceae) and is one of the features that distinguishes *S. prostrata* from *S. tetraptera* (Carswell & Gould, 1998). In the *Sophora* F_2 population, node angle was mapped to one major locus using multi-QTL mapping at an alpha of 0.05, explaining a considerable 23.6% variation. The single marker analyses located a second QTL indicating other loci may be involved. Lowering the multi-QTL model to alpha levels of 0.10 and 0.20 identified three more large-effect QTL. At an alpha of 0.10, a QTL, Q3, was identified that overlaps with QTL for rachis length, leaflet length and leaflet width, while at an alpha of 0.20, Q4 overlaps with QTL for rachis length, leaflet length, branch width. As with other QTL that show overlap among traits, such a finding was somewhat anticipated by the trait correlations characterised in Chapter 2; however, these two QTL are major effect QTL in most of these traits suggesting that these loci may have been key to the development of the divaricating growth form as both leaf and branch traits were mapped to these loci. The QTL identified for node angle at an alpha of 0.05 does not co-locate with other QTL indicating node angle is also controlled by major independent QTL. Leaflet number is also correlated with node angle and branch width but does not co-locate loci with these branch traits; however, leaflet number does co-locate at other QTL locations with other leaf traits which could impact the correlation with the branch traits. Including results from all significance levels, the total explained variation for node angle was estimated at 53.9% indicating environmental effects may also be influencing node angles in *Sophora*.

Node angle appears to be a trait not often studied in QTL analyses but is an important feature of divarication in *Sophora* and is also seen in other divaricates. Therefore, the results from the *Sophora* population contribute to increasing knowledge of the genetic loci involved in producing the distinctive zig-zag stem and the genetic control of node angle in general.

Interlacing branching is a key feature used in describing the divaricate growth form, but is difficult to measure (Atkinson, 1992; Kelly, 1994). As clear contributors to interlaced branching, branch angle, branch number, and internode length were chosen. For branch angle, no QTL were identified in multi-QTL mapping at an alpha of 0.05. In fact, it was not until a cut-off of 0.15 that a QTL was located for this trait. This same QTL was identified using non-parametric analyses at 0.05 so a combination of the distribution of the trait and our relatively small sample size may contribute to the difficulty in mapping for this trait. Further, an *effectplot* at this locus shows a clear distinction between the three genotypes (Figure 5.5) and the estimated effect is rather large at 13.49%. Three more fairly large-effect QTL were located for branch angle at an alpha of 0.20. Only one of these co-located with QTL for other traits. This too matches reasonably well with the correlation analyses from Chapter 2, which showed that branch angle did not correlate with other traits.

Mapping results for branch angle in *Brassica napus* L. (Brassicaceae) indicates that it is a complex trait with multiple loci and environmental effects contributing to the variation (Liu et al., 2016; Sun et al., 2016; Shen et al., 2018). Analysis of forest grown *S. prostrata* and *S. tetraptera* have shown that the pattern of shoot development is highly variable within both species and *S. prostrata* does not always branch more frequently or have wider branch angles than *S. tetraptera.* Of the four branch characteristics considered here, branch number, branch angle, node angle and internode length, only

node angle was observed as distinct between *S. prostrata* and *S. tetraptera* in the field study (Carswell & Gould, 1998). This may contribute to the lack of QTL identified in *Sophora* for the branch traits; however, there was a large difference observed for these traits between the parental plants used to generate the F_2 population (see Chapter 2), although only one of the *S. prostrata* individuals, G1916, remains.

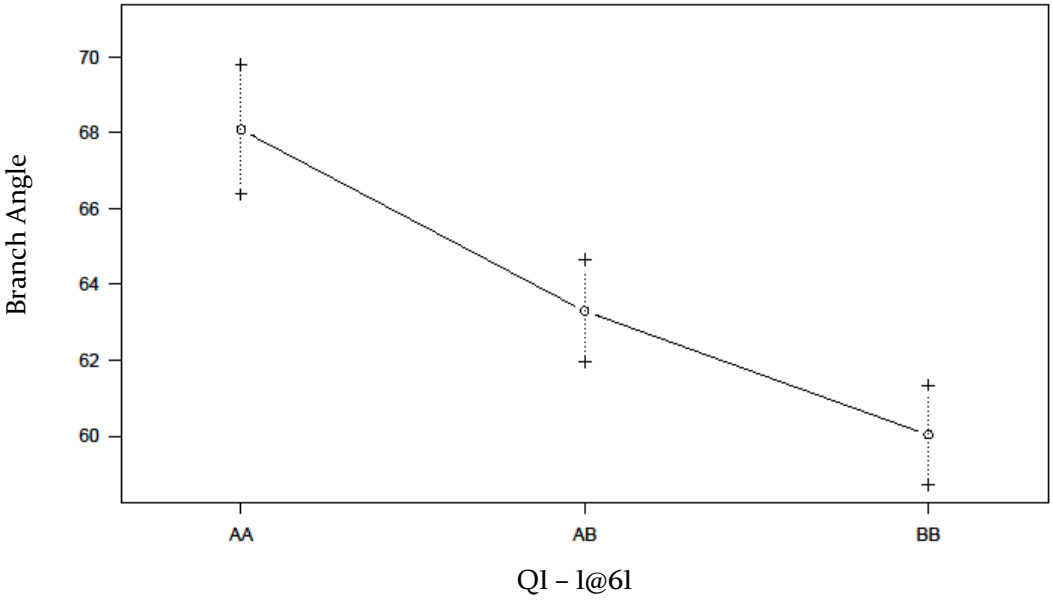


Figure 5.5. *Effectplot* of the QTL located for Branch Angle in non-parametric mapping at an alpha of 0.05.

A study of *Pinus elliotti x P. caribaea* (Pinaceae) F_1 consisting of 89 trees failed to detect QTL for branch angle even though it was one of the more heritable traits in pines. The inability to locate QTL in that study was attributed to a lack of segregation, low power to detect QTL from small sample size, incomplete map coverage or a difference in how the trait was measured compared to other studies (Shepherd et al., 2002). Many of these, such as incomplete map coverage and lack of power from a smaller sample size, which can prevent detection of minor effect QTL (Lynch & Walsh, 1998; Singh & Singh, 2015), may explain our results in *Sophora* also with a sample size of 88 and may be contributing to the lack of QTL in some traits. Despite the obvious difference in the divaricate growth form and arborescent growth form in *Sophora* species, these traits,

(increased branching, wide branch angle and longer internodes) often used to describe divarication, may not best describe the difference between the growth forms well and could suggest different measurements are required to capture the difference in growth form and ultimately identify QTL resulting in the branch architecture of *Sophora*.

No QTL were mapped for branch number in the *Sophora* F_2 population. The genetic architecture of branch number has been analysed in other species such as soybean, (He et al., 2014), Chrysanthemum *(Chrysanthemum morifolium* (Ramat.) Hemsl. (Asteraceae)) (Peng et al., 2015), and *Pinus elliottii* Engelm (Pineaceae) (Shepherd et al., 2002). Results in these species identified multiple QTL (three to eight) for branch number and concluded that both genetic and environmental factors contribute to the final phenotype of branch number. In *Sophora,* branch number was measured in a new F_2 population (PN F_2), grown in a greenhouse to better control for environmental effects (see Chapter 3). In this population, branch number was found to be correlated with other traits and was associated with the *RMSI* marker. Due to the difference in the age of the plants in the Lincoln F_2 and the PN F_2 populations, the branch number phenotype was scored differently, with the Lincoln F_2 measuring branches from the periphery of the plant whereas the glasshouse F_2 measured branching from the main stem, counting primary and secondary branching as well as total branch number.

As branch traits can be affected by branch age (Gort et al., 2010; Thomas, 2014), this may be another factor contributing to lack of power to detect QTL in the Lincoln F_2 population. While attempts were made to measure similar age branches, this necessarily varied with the individual and it was not always possible to be consistent. Finally, as the Lincoln F_2 plants have been growing for over 30 years and the site has not been maintained (manicured) through this time, there may have been environmental effects that influenced the branching of these individuals; e.g. due to shading from competing

plants or differing resources based on variation within the plot. The lower number of QTL detected for branch traits may indicate why there were no loci co-located for the branch and leaf traits at an alpha of 0.05 despite correlations observed for most traits (see Chapter 2).

An alternative method of analysing plant architecture by using 3D digitization has been used in rose bush (Li-Marchetti et al., 2017). The method creates a 3D representation of the plant architecture generating many variables, some that cannot be accurately measured manually. QTL were located for the six most relevant variables, five of which had not been analysed before due to an inability to measure these traits. The number of QTL identified, ranged from three for the number of determined axes to seven for branch angle (Li-Marchetti et al., 2017), indicating this may be a useful and informative approach to differentiate growth forms in *Sophora*.

A second F_2 population of 69 individuals was developed in a greenhouse (described in Chapter 3), to control for the environmental variation that may be present for the *Sophora* F_2. Phenotyping of the nine traits has been completed for these plants. These individuals would need to be genotyped before QTL analysis could be performed but this population may be useful in the future to better identify QTL associated with the branch traits, as it has been grown in controlled conditions, which would reduce environmental differences between samples; however, the smaller population size would remain an issue.

5.5.4 Candidate gene markers

Mutants of the *RAMOSUS* genes, involved in the strigolactone pathway, show increased branching as well as other phenotypes, such as smaller leaves (Arumingtyas et al., 1992; Stirnberg et al., 2002; Sorefan et al., 2003; Booker et al., 2004), which are similar to those traits that define the divaricate form. The strigolactone pathway is

involved in the control of axillary branching and acts to prevent the axillary buds from bud release (Beveridge et al., 1997; Morris et al., 2001; Sorefan et al., 2003; Booker et al., 2004; Auldridge et al., 2006; de Saint Germain et al., 2016). Therefore, this pathway could be at least partially involved in producing the divaricate form. Four genes from the strigolactone pathway were included as markers in the linkage maps (Chapters 3 and 4). Two of these genes, *RMS1* and *RMS4,* were within or near the confidence intervals of QTL for many of the traits in the *Sophora* F_2. *RMS4* is involved in signal perception and response to the strigolactone signal (Woo et al., 2001; Stirnberg et al., 2007; Jiang et al., 2013; Soundappan et al., 2015). *RMS4* was significantly associated, before Bonferroni correction, with leaflet length and rachis length in ANOVA analyses (see Chapter 3). Further supporting an association between these traits and this region of chromosome four, a QTL for each was mapped very close to *RMS4,* as was a QTL for leaflet number.

Leaflet length, leaflet width, rachis length, node angle and branch width mapped QTL near *RMS1*. *RMS1* is one of the genes involved in the biosynthesis of the strigolactone hormone (Flematti et al., 2016). The original ANOVA analyses showed associations in the Lincoln F_2 that showed *RMS1* is associated with all phenotypes except leaflet number, branch number and branch angle before Bonferroni correction (leaflet length, width and rachis length were statistically significant after correction) (see Chapter 3). A sequence analysis of *RMS1* (Chapter 3) indicated a potentially deleterious amino acid replacement in *S. prostrata* which may indicate that *S. prostrata* is lacking a fully functional RMS1, which would be expected to decrease the efficiency of strigolactone synthesis.

The *RMS1* marker is a major QTL having effect sizes of greater than 10% in all but one of the associated traits, leaflet length, which has an effect of 9.61%. Further, results from the preliminary strigolactone application experiments indicate that supplementary

strigolactone had effects on leaflet length and leaflet width traits in *S. prostrata.* In sum, these results including QTL mapping for multiple traits, ANOVA associations, a predicted deleterious mutation in *S. prostrata,* and hormone experiments provide evidence that variation in the *RMS1* locus may be a significant contributor to the divaricate growth form in *S. prostrata* and represents a very good candidate gene for further investigation.

No QTL were mapped near *RMS5* and *ZIG* indicating that they are not significant contributors to the divaricate growth form in *Sophora*. *ZIG* was identified in a mutant of *Arabidopsis* that resulted in a zig-zag-like stem (Kato et al., 2002) however the QTL for node angle in *Sophora* did not map near the *ZIG* marker. In fact, it was on a completely different linkage group, suggesting this gene is not a major locus involved in the zig-zag stem in *Sophora.* Although no associations were made with *RMS2* in the ANOVA analyses, branch angle did map near *RMS2.* Along with *RMS4, RMS2* may also be an important gene that warrants further work.

The promising links between the *RMS* genes and the divarication traits remain just that; as many genes will be linked to any QTL position, it is also possible that other genes are the actual causal genes and further study would be needed to confirm the involvement of any candidates. Tomlinson (1978) suggested divarication largely results from a loss of meristem control and it is likely the loss of axillary branch control that may contribute to the divaricate form, indicating that the strigolactone pathway, which is conserved among plants species (Johnson et al., 2006), remains a good candidate for genes involved in producing or contributing to divarication.

5.5.5 Maternal effects

Linkage group 10 was included as a single marker linkage group in the QTL analyses to account for variation resulting from maternal effects, as the *Sophora* F_2

population consisted of reciprocal crosses. Maternal effects were identified in the *Sophora* F_2 from analysis of the morphological data in Chapter 2. In the QTL analyses, leaflet number mapped to the maternal direction marker, LG10, indicating a maternal effect for this trait. Node angle may also experience a maternal effect as it was indicated linkage group 10 was a QTL in the single marker analyses; however, this was not considered in the multi-QTL model suggesting it may have a minor effect. Maternal effect can result from plastids inherited from the maternal parent, the 3*n* endosperm which has 2*n* majority from the maternal parent, nutrients, mRNAs or proteins supplied from the maternal parent (Byers et al., 1997; Wolf, 2000) and effects resulting from the seed coat which is made from maternal tissue (Donohue, 2009). Maternal effects may influence the phenotype of the offspring enabling them to cope with environmental conditions the maternal parent is exposed to (Mousseau & Fox, 1998). In *Lupinus texensus* (Fabaceae) maternal effects were important in the early growth of the offspring, mostly affecting the growth rate and size of offspring, indicating they can result in increased fitness of offspring (Helenurm & Schaal, 1996).

5.5.6 Multiple trait analyses

While the divaricate growth form results from a combination of traits, the overall form can be described as a single phenotype. In an attempt to represent the overall divaricate growth form, different methods of combining the trait data were used. The first approach used the principal component axes, PC1 and PC2, from a PCA. PCA is essentially a data transformation method that converts many potentially correlated variables into a set of values of uncorrelated variables. For the *Sophora* data, PC1 mainly reflects the variation in the leaf characters, node angle, and branch width, whereas PC2 mainly represents the branch character variation. Five QTL were mapped for PC1, some of which were unique locations and two that co-located with QTL from individual traits,

such as rachis length, leaflet length, branch width and node angle, which is expected as these traits explain most of the variance along this axis. The new QTL locations may represent loci that have small effects on several traits, such that the signal for any one trait is too weak to make a significant association but, when the effects are summed across traits (as represented as a principle component), the signal is strong enough to make an association. The total variation explained for QTL in PC1 was 75.24% and as the leaf characteristics often had high variance this may be as expected. PC2 did not locate any QTL at an alpha of 0.05, which may result from the traits that did not map QTL explaining variation in this direction of the PCA and the effects of these traits may prevent QTL detection as discussed previously. At alpha of 0.10-0.20, five QTL were identified; four of these located near the branch angle QTL.

A second method of combining data was to use a divaricate index score; however the published indices (Atkinson, 1992; Kelly, 1994) could not be calculated for the F_2 population as the individual measurements taken were not the same as those used for either index so an index that better represented the difference between *S. prostrata* and *S. tetraptera* was calculated. This involved the traits rachis length, branch width, internode length and branch number as these traits are often used to describe the divaricate form. Using this new index, only one QTL was mapped at an alpha of 0.05, using non-parametric mapping. The QTL is on linkage group 1 and described 15.28% of the variation in the index. This QTL is located near a QTL for leaflet number and rachis length, supporting this as a real QTL likely resulting from the inclusion of the rachis length data. Multi-QTL mapping at alpha of 0.10-0.20 identified three QTL, one of which overlaps with the QTL identified by the non-parametric method although the confidence intervals for these were both quite long, indicating this could be the same QTL. Two of the traits used in the index, internode length and branch number, did not map QTL on their own and may have affected the index results.

5.5.7 A role for dominance in the evolution of the divaricating form in *Sophora?*

Many of the QTL in the *Sophora* F_2 revealed strong dominance for the *S. prostrata* allele: 21 of the 26 QTL in the leaf traits and 5 of the 10 QTL for the branch traits (Table 5.2 & Figure 5.3). The presence of many loci that are dominant for the *S. prostrata* alleles suggests that dominant mutations may have played an important role in the evolution of the divaricate growth form as the divaricate form would have arisen from a non-divaricate, larger leaved ancestor. The phenotypic effects of dominant alleles are manifested as soon as they arise and, therefore, selection is able to act on them immediately. If these alleles increased fitness in a new or original environment then natural selection may quickly increase their frequency, compared to recessive mutations, in the population. Dominant mutations, in the different traits that contribute to the divaricate form, could allow for a relatively rapid evolution of the divaricate growth form. However, most new mutations arise as recessive mutations (Orr, 1991), meaning a higher frequency of dominant mutations, as is suggested by the majority of dominant *S. prostrata* alleles, is unusual. In a study of dominance in *Arabidopsis,* a higher percentage of morphological mutations generated from mutagenesis were dominant mutations compared to new mutations in 'essential' genes, where most mutations were recessive (Meinke, 2013). While the *Arabidopsis* study utilised artificial mutagenesis and examined variation in a non-natural setting, this work still highlights a finding highly relevant to the current study: a high proportion of new mutations that convey morphological variation are dominant. As the divaricating growth form is essentially a plant morphology, it is conceivable that it has arisen in *Sophora* via a series of largely dominant mutations.

In this context, it must also be noted that the relationship between *S. prostrata* and *S. tetraptera* is not resolved, and though *S. prostrata* likely evolved from a non-divaricate, larger leaved species, *S. tetraptera* may or may not be close to the morphology

of the *S. prostrata* ancestor. It is conceivable then that the typically recessive nature of the *S. tetraptera* alleles relative to *S. prostrata* alleles is an artefact of the crossing history and has little bearing on the evolutionary genetic origins of divarication in *Sophora*. Resolving the inter-specific relationships among the New Zealand *Sophora* should help with interpreting the genetic data in an evolutionary context. Despite this convoluted history, *S. prostrata* likely evolved from a non-divaricate species, with larger leaves, and the possibility of this occurring from multiple dominant mutations will be an interesting topic for future investigation.

The issue of dominance in *Sophora* is made all the more interesting by the finding of positive or negative overdominance for many QTL, again, typically, in the direction of the *S. prostrata* allele (Table 5.2). This was observed at 15 QTL and would have interesting implications for (1) new mutations, as they would always be expected to be in heterozygous form initially and (2) hybridization scenarios. In both cases, such overdominance would be expected to generate transgressive segregation, which may be an important source of phenotypic variation for natural selection to act on.

5.5.8 Transgressive segregation

Transgressive segregation was observed for some of the traits measured and is often attributed to overdominance, complementary gene action (Devicente & Tanksley, 1993) or epistasis (Brem & Kruglyak, 2005). In plants it is often attributed to additive alleles that are present in new combinations in the F_2, which is observed frequently in segregating populations (Rieseberg et al., 2003). Transgressive segregation is also often indicative that multiple loci are involved in variation for a trait. Most of the traits examined here identified multiple QTL and it is likely that transgressive segregation could be caused by new combinations of alleles at these loci in the F_2. Indeed, both *S.*

tetraptera and *S. prostrata* posess alleles at some QTL that have effects opposite those expected based on the parental phenotypes.

As described above, there are several instances of over-dominance at the QTL identified in the current study, where the heterozygote mean was beyond the mean of one of the parental homozygotes. At many of the QTL mapped here, the heterozygotes showed phenotypes much greater or smaller than the two parents (Figure 5.4), indicating there may be effects resulting from the combination of alleles producing a phenotype beyond those of the parents.

5.5.9 Limitations

The work presented here is a novel study on the genetic architecture of the divaricate growth form in *Sophora* and was able to identify QTL associated with several traits contributing to the divaricate growth form. There are some limitations to this study including the total sample size of the population, 88 individuals, which is above the lower limit recommended as sufficient to locate QTL (50-250 individuals) (Mohan et al., 1997; Collard et al., 2005) but is also toward the lower limit. It also consists of reciprocal crosses with a different *S. prostrata* parent for each cross and therefore may be limited in the ability to locate QTL, especially those with minor effects. The number of markers available to generate the linkage maps was 148 and included some gaps of ~20cM (see Chapter 4) which could affect the ability to identify QTL whereas a map with a higher density of markers, and therefore smaller gaps, may have more power to detect QTL. However, in simulations, a gap of 20cM is suggested to only slightly decrease power to detect QTL (Darvasi et al., 1993).

5.6 Conclusion

The genetic basis of the divaricate growth form has represented a significant gap in our understanding of this important component of the New Zealand flora. As divarication is present in many different genera it is possible that the growth form could be produced from the action of a few genes of large effects, however, our results for *Sophora* show that the growth form in this group results from the action of many different loci with a wide range of effects. Initial findings of correlations between leaf traits were supported by the presence of QTL that co-localise for multiple traits; however, many loci were mapped that affect just one or two traits. There were no loci identified in this study that affect all divaricate traits despite most traits being correlated (Chapter 2). Internode length and branch number were the only traits that did not locate QTL. Two of the *RMS* genes, *RMS1* and *RMS4,* were markers that mapped QTL for leaf and branch traits, suggesting they are good candidate genes for further study. The results from the *Sophora* F_2 population provide our first glimpse at the genetic architecture behind the divaricate growth form, suggest the possible involvement of the strigolactone pathway, and provide an important step toward understanding the molecular genetic basis of divarication, the unique growth form of the New Zealand flora.

5.7 References

Allan, H. H. (1961). *Flora of New Zealand. Vol. 1. Indigenous Tracheophyta: Psilopsida, Lycopsida, Filicopsida, Gymnospermae, Dicotyledones.* Government Printer, Wellington.

Arends, D., Prins, P., Jansen, R. C., & Broman, K. W. (2010). R/qtl: high-throughput multiple QTL mapping. *Bioinformatics, 26*(23), 2990-2992. 10.1093/bioinformatics/btq565

Arumingtyas, E. L., Floyd, R. S., Gregory, M. J., & Murfet, I. C. (1992). Branching in *Pisum*: inheritance and allelism tests with 17 *ramosus* mutants. *Pisum Genet, 24*, 17-31.

Atkinson, I. A. E. (1992). *A method for measuring branch divergence and interlacing in woody plants*. Lower Hutt, New Zealand: DSIR Land Resources.

Atkinson, I. A. E., & Greenwood, R. M. (1989). Relationships between moas and plants. *New Zealand Journal of Ecology, 12*, 67-96.

Auldridge, M. E., Block, A., Vogel, J. T., Dabney-Smith, C., Mila, I., Bouzayen, M., Magallanes-Lundback, M., DellaPenna, D., McCarty, D. R., & Klee, H. J. (2006). Characterization of three members of the *Arabidopsis* carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family. *Plant Journal, 45*(6), 982-993. 10.1111/j.1365-313X.2006.02666.x

Beveridge, C. A., Symons, G. M., Murfet, I. C., Ross, J. J., & Rameau, C. (1997). The rms1 mutant of pea has elevated indole-3-acetic acid levels and reduced root-sap zeatin riboside content but increased branching controlled by graft-transmissible signal(s). *Plant Physiology, 115*(3), 1251-1258.

Booker, J., Auldridge, M., Wills, S., McCarty, D., Klee, H., & Leyser, O. (2004). MAX3/CCD7 is a carotenoid cleavage dioxygenase required for the synthesis of a novel plant signaling molecule. *Current Biology, 14*(14), 1232-1238. 10.1016/j.cub.2004.06.061

Brem, R. B., & Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences of the United States of America, 102*(5), 1572-1577. 10.1073/pnas.0408709102

Broman, K. W., & Sen, S. (2009). *A guide to QTL mapping with R/qtl*. United States of America: Springer Science+Business Media, LLC.

Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics, 19*(7), 889-890. 10.1093/bioinformatics/btg112

Byers, D. L., Platenkamp, G. A. J., & Shaw, R. G. (1997). Variation in seed characters in *Nemophila menziesii*: Evidence of a genetic basis for maternal effect. *Evolution, 51*(5), 1445-1456. 10.2307/2411197

Carlquist, S. (1974). *Island Biology*. New York: Columbia University.

Carswell, F. E., & Gould, K. S. (1998). Comparative vegetative development of divaricating and arborescent *Sophora* species (Fabaceae). *New Zealand Journal of Botany, 36*(2), 295-301. 10.1080/0028825x.1998.9512567

Chomicki, G., Coiro, M., & Renner, S. S. (2017). Evolution and ecology of plant architecture: integrating insights from the fossil record, extant morphology, developmental genetics and phylogenies. *Annals of Botany, 120*(6), 855-891. 10.1093/aob/mcx113

Chunlian, L., Guihua, B., Shiaoman, C., Carver, B., & Zhonghua, W. (2016). Single nucleotide polymorphisms linked to quantitative trait loci for grain quality traits in wheat. *Crop Journal, 4*(1), 1-11. 10.1016/j.cj.2015.10.002

Cockayne, L. (1912). Observations concerning evolution, derived from ecological studies in New Zealand. *Transactions and Proceedings of the New Zealand Institute, 44*, 1-30.

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica, 142*(1-2), 169-196. 10.1007/s10681-005-1681-5

Costes, E., & Gion, J. M. (2015). Genetics and Genomics of Tree Architecture. In C. Plomion & A. F. AdamBlondon (Eds.), *Land Plants - Trees* (Vol. 74, pp. 157-200). 10.1016/bs.abr.2015.05.001

Darvasi, A., Weinreb, A., Minke, V., Weller, J. I., & Soller, M. (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic-map. *Genetics, 134*(3), 943-951.

de Lange, P. J. (2004). *Sophora tetraptera. Flora of New Zealand.* Retrieved 9/06/2015 2015 from http://www.nzpcn.org.nz/flora_details.aspx?ID=1304

de Saint Germain, A., Clave, G., Badet-Denisot, M.-A., Pillot, J.-P., Cornu, D., Le Caer, J.-P., Burger, M., Pelissier, F., Retailleau, P., Turnbull, C., Bonhomme, S., Chory, J., Rameau, C., & Boyer, F.-D. (2016). An histidine covalent receptor and butenolide complex mediates strigolactone perception. *Nature Chemical Biology, 12*(10), 787-+. 10.1038/nchembio.2147

Devicente, M. C., & Tanksley, S. D. (1993). QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics, 134*(2), 585-596.

Diels, L. (1897). Vegetations biologie von Nei-Seeland. *Botanische Jahrbucher fur Systematik, Pflanzengeschichte und Pflanzengeographie, 22*, 202-300.

Donohue, K. (2009). Completing the cycle: maternal effects as the missing link in plant life histories. *Philosophical Transactions of the Royal Society B-Biological Sciences, 364*(1520), 1059-1074. 10.1098/rstb.2008.0291

Falconer, D. S. (1964). *Introduction to quantitative genetics*. Edinburgh: Oliver & Boyd.

Fehr, W. R. (1972). Genetic-control of leaflet number in soybeans. *Crop Science, 12*(2), 221-&. 10.2135/cropsci1972.0011183X001200020023x

Flematti, G. R., Scaffidi, A., Waters, M. T., & Smith, S. M. (2016). Stereospecificity in strigolactone biosynthesis and perception. *Planta, 243*(6), 1361-1373. 10.1007/s00425-016-2523-5

Gailing, O. (2008). QTL analysis of leaf morphological characters in a *Quercus robur* full-sib family (*Q-robur x Q-robur ssp slavonica*). *Plant Biology, 10*(5), 624-634. 10.1111/j.1438-8677.2008.00063.x

Godley, E. J. (1975). Kowhai. *New Zealands Nature Heritage, 5 1804-1806.*

Godley, E. J. (1985). Paths to maturity. *New Zealand Journal of Botany, 23*, 687-706.

Gort, J., Zubizarreta-Gerendiain, A., Peltola, H., Kilpelainen, A., Pulkkinen, P., Jaatinen, R., & Kellomaki, S. (2010). Differences in branch characteristics of Scots pine (*Pinus sylvestris* L.) genetic entries grown at different spacing. *Annals of Forest Science, 67*(7) 10.1051/forest/2010030

Greenwood, R. M., & Atkinson, I. A. E. (1977). Evolution of divaricating plants in New Zealand in relation to moa browsing. *Proceedings of the New Zealand Ecological Society, 24*

Grierson, E. R. P. (2014). *The Development and Genetic Variation of Sophora prostrata - A New Zealand Divaricating Shrub.* (Master's Thesis), The University of Waikato, Hamilton, New Zealand.

Haley, C. S., & Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity, 69*, 315-324. 10.1038/hdy.1992.131

He, Q., Yang, H., Xiang, S., Wang, W., Xing, G., Zhao, T., & Gai, J. (2014). QTL mapping for the number of branches and pods using wild chromosome segment substitution lines in soybean *Glycine max* (L.) Merr. *Plant Genetic Resources-Characterization and Utilization, 12*, S172-S177. 10.1017/s1479262114000495

Heenan, P., Mitchell, C., & Houliston, G. (2018). Genetic Variation and Hybridisation among Eight Species of kowhai (*Sophora*: Fabaceae) from New Zealand Revealed by Microsatellite Markers. *Genes, 9*(2) 10.3390/genes9020111

Helenurm, K., & Schaal, B. A. (1996). Genetic and maternal effects on offspring fitness in *Lupinus texensis* (Fabaceae). *American Journal of Botany, 83*(12), 1596-1608. 10.2307/2445836

Hollender, C. A., & Dardick, C. (2015). Molecular basis of angiosperm tree architecture. *New Phytologist, 206*(2), 541-556. 10.1111/nph.13204

Jansen, R. C. (1994). Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics, 138*(3), 871-881.

Jansen, R. C., & Stam, P. (1994). High-resolution of quantitative traits into multiple loci via interval mapping. *Genetics, 136*(4), 1447-1455.

Jiang, L., Liu, X., Xiong, G., Liu, H., Chen, F., Wang, L., Meng, X., Liu, G., Yu, H., Yuan, Y., Yi, W., Zhao, L., Ma, H., He, Y., Wu, Z., Melcher, K., Qian, Q., Xu, H. E., Wang, Y., & Li, J. (2013). DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature, 504*(7480), 401-+. 10.1038/nature12870

Johnson, X., Brcich, T., Dun, E. A., Goussot, M., Haurogne, K., Beveridge, C. A., & Rameau, C. (2006). Branching genes are conserved across species. Genes controlling a novel signal in pea are coregulated by other long-distance signals. *Plant Physiology, 142*(3), 1014-1026. 10.1104/pp.106.087676

Jung, J.-H., Seo, P. J., Kang, S. K., & Park, C.-M. (2011). miR172 signals are incorporated into the miR156 signaling pathway at the SPL3/4/5 genes in *Arabidopsis* developmental transitions. *Plant Molecular Biology, 76*(1-2), 35-45. 10.1007/s11103-011-9759-z

Kato, T., Morita, M. T., Fukaki, H., Yamauchi, Y., Uehara, M., Niihama, M., & Tasaka, M. (2002). SGR2, a Phospholipase-Like Protein, and ZIG/SGR4, a SNARE, Are Involved in the Shoot Gravitropism of *Arabidopsis*. *The Plant Cell Online, 14*(1), 33-46. 10.1105/tpc.010215

Kelly, D. (1994). Towards a numerical definition of divaricate (interlaced small-leaved) shrubs. *New Zealand Journal of Botany, 32*(4), 509-518.

Kim, H. K., Kang, S. T., & Suh, D. Y. (2005). Analysis of quantitative trait loci associated with leaflet types in two recombinant inbred lines of soybean. *Plant Breeding, 124*(6), 582-589. 10.1111/j.1439-0523.2005.01152.x

Li-Marchetti, C., Le Bras, C., Chastellier, A., Relion, D., Morel, P., Sakr, S., Oyant, L. H.-S., & Crespel, L. (2017). 3D phenotyping and QTL analysis of a complex character: rose bush architecture. *Tree Genetics & Genomes, 13*(5) 10.1007/s11295-017-1194-0

Liu, J., Wang, W., Mei, D., Wang, H., Fu, L., Liu, D., Li, Y., & Hui, Q. (2016). Characterizing Variation of Branch Angle and Genome-Wide Association Mapping in Rapeseed (*Brassica napus* L.). *Frontiers in Plant Science, 7* 10.3389/fpls.2016.00021

Livingstone, D. A. (1974). Extract from INQUA report to the U.S. National Academy of Sciences. *Soil News, 22*, 46-48.

Lovell, J. T. (2018). qtlTools: Data Processing and Plotting in Association with R/qtl (Version R package version 1. 2. 0).

Lusk, C. H., & Clearwater, M. J. (2015). Leaf temperatures of divaricate and broadleaved tree species during a frost in a North Island lowland forest remnant, New

Zealand. *New Zealand Journal of Botany, 53*(4), 202-209. 10.1080/0028825x.2015.1086390
Lusk, C. H., McGlone, M. S., & Overton, J. M. (2016). Climate predicts the proportion of divaricate plant species in New Zealand arborescent assemblages. *Journal of Biogeography, 43*(9), 1881-1892. 10.1111/jbi.12814
Lynch, M., & Walsh, B. (1998). *Genetics and Analysis of Quantative Traits*. Sunderland, MA, U.S.A: Sinauer Associates, Inc.
McGlone, M. S., & Webb, C. J. (1981). Selective forces influencing the evolution of divaricating plants. *New Zealand Journal of Ecology, 4*, 20-28.
Meinke, D. W. (2013). A survey of dominant mutations in *Arabidopsis thaliana*. *Trends in Plant Science, 18*(2), 84-91. 10.1016/j.tplants.2012.08.006
Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A., & Beeckman, T. (2008). Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. *Nature Genetics, 40*(12), 1489-1492. 10.1038/ng.253
Mohan, M., Nair, S., Bhagwat, A., Krishna, T. G., Yano, M., Bhatia, C. R., & Sasaki, T. (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding, 3*(2), 87-103. 10.1023/a:1009651919792
Morris, S. E., Turnbull, C. G. N., Murfet, I. C., & Beveridge, C. A. (2001). Mutational analysis of branching in pea. Evidence that RMS1 and RMS5 regulate the same novel signal. *Plant Physiology, 126*(3), 1205-1213. 10.1104/pp.126.3.1205
Mousseau, T. A., & Fox, C. W. (1998). The adaptive significance of maternal effects. *Trends in Ecology & Evolution, 13*(10), 403-407. 10.1016/s0169-5347(98)01472-4
Niklas, K. J. (2000). The evolution of plant body plans - A biomechanical perspective. *Annals of Botany, 85*(4), 411-438. 10.1006/anbo.1999.1100
Orr, H. A. (1991). A test of Fisher's theory of dominance. *Proceedings of the National Academy of Sciences of the United States of America, 88*(24), 11413-11415. 10.1073/pnas.88.24.11413
Peng, H., Zhang, F., Jiang, J., Chen, S., Fang, W., Guan, Z., & Chen, F. (2015). Identification of quantitative trait loci for branching traits of spray cut chrysanthemum. *Euphytica, 202*(3), 385-392. 10.1007/s10681-014-1259-1
Poethig, R. S. (2009). Small RNAs and developmental timing in plants. *Current Opinion in Genetics & Development, 19*(4), 374-378. 10.1016/j.gde.2009.06.001
R Development Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org
Reinhardt, D., & Kuhlemeier, C. (2002). Plant architecture. *Embo Reports, 3*(9), 846-851. 10.1093/embo-reports/kvf177
Revelle, W. (2016). psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, Illinois: Northwestern University.
Rieseberg, L. H., Widmer, A., Arntz, A. M., & Burke, J. M. (2003). The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 358*(1434), 1141-1147. 10.1098/rstb.2003.1283
Shen, Y., Yang, Y., Xu, E., Ge, X., Xiang, Y., & Li, Z. (2018). Novel and major QTL for branch angle detected by using DH population from an exotic introgression in rapeseed (*Brassica napus* L.). *Theoretical and Applied Genetics, 131*(1), 67-78. 10.1007/s00122-017-2986-1
Shepherd, L. D., de Lange, P. J., Perrie, L. R., & Heenan, P. B. (2017). Chloroplast phylogeography of New Zealand *Sophora* trees (Fabaceae): extensive hybridization and widespread Last Glacial Maximum survival. *Journal of Biogeography, 44*(7), 1640-1651. 10.1111/jbi.12963

Shepherd, M., Cross, M., Dieters, M. J., & Henry, R. (2002). Branch architecture QTL for *Pinus elliottii* var. *elliottii* x *Pinus caribaea* var. *hondurensis* hybrids. *Annals of Forest Science, 59*(5-6), 617-625. 10.1051/forest:2002047

Shupert, D. A., Byme, D. H., & Pemberton, H. B. (2007). Inheritance of flower traits, leaflet number and prickles in roses. In H. B. Pemberton (Ed.), *Proceedings of the IVth International Symposium on Rose Research and Cultivation* (pp. 331-+). 10.17660/ActaHortic.2007.751.42

Singh, B. D., & Singh, A. K. (2015). Mapping Populations. In *Marker-Assisted Plant Breeding: Principles and Practices* (pp. 125-150). New Delhi: Springer India. 10.1007/978-81-322-2316-0_5

Soehendi, R., Chanprame, S., Toojinda, T., Ngampongsai, S., & Srinives, P. (2007). Genetics, agronomic, and molecular study of leaflet mutants in mungbean (*Vigna radiata* (L.) Wilczek). *Journal of Crop Science and Biotechnology, 10*(3), 193-200.

Song, J. (2005). *Genetic diversity and flowering in Clianthus and New Zealand Sophora (Fabaceae).* (Ph.D. Thesis), Massey University, Palmerston North, New Zealand.

Sorefan, K., Booker, J., Haurogne, K., Goussot, M., Bainbridge, K., Foo, E., Chatfield, S., Ward, S., Beveridge, C., Rameau, C., & Leyser, O. (2003). *MAX4* and *RMS1* are orthologous dioxygenase-like genes that regulate shoot branching in *Arabidopsis* and pea. *Genes & Development, 17*(12), 1469-1474. 10.1101/gad.256603

Soundappan, I., Bennett, T., Morffy, N., Liang, Y., Stang, J. P., Abbas, A., Leyser, O., & Nelson, D. C. (2015). SMAX1-LIKE/D53 family members enable distinct MAX2-dependent responses to strigolactones and karrikins in *Arabidopsis*. *Plant Cell, 27*(11), 3143-3159. 10.1105/tpc.15.00562

Stirnberg, P., Furner, I. J., & Ottoline Leyser, H. M. (2007). MAX2 participates in an SCF complex which acts locally at the node to suppress shoot branching. *Plant Journal, 50*(1), 80-94. 10.1111/j.1365-313X.2007.03032.x

Stirnberg, P., van de Sande, K., & Leyser, H. M. O. (2002). MAX1 and MAX2 control shoot lateral branching in *Arabidopsis*. *Development, 129*(5), 1131-1141.

Stone, B. F. (1968). A formula for determining degree of dominance in cases of monofactorial inheritance of resistance to chemicals. *Bulletin of the World Health Organization, 38*(2), 325-&.

Sun, C., Wang, B., Wang, X., Hu, K., Li, K., Li, Z., Li, S., Yan, L., Guan, C., Zhang, J., Zhang, Z., Chen, S., Wen, J., Tu, J., Shen, J., Fu, T., & Yi, B. (2016). Genome-wide association study dissecting the genetic architecture underlying the branch angle trait in rapeseed (*Brassica napus* L.). *Scientific Reports, 6* 10.1038/srep33673

Tan, L.-Q., Wang, L.-Y., Xu, L.-Y., Wu, L.-Y., Peng, M., Zhang, C.-C., Wei, K., Bai, P.-X., Li, H.-L., Cheng, H., & Qi, G.-N. (2016). SSR-based genetic mapping and QTL analysis for timing of spring bud flush, young shoot color, and mature leaf size in tea plant (*Camellia sinensis*). *Tree Genetics & Genomes, 12*(3) 10.1007/s11295-016-1008-9

Thomas, P. (2014). *Trees: Their Natural History* (2 ed.). Cambridge CB2 8BS, United Kingdom: University Printing House.

Tomlinson, P. S. (1978). Some qualitative and quantitive aspects of New Zealand divaricating shrubs. *New Zealand Journal of Botany, 16*, 299-309.

Tsarouhas, V., Gullberg, U., & Lagercrantz, U. (2002). An AFLP and RFLP linkage map and quantitative trait locus (QTL) analysis of growth traits in *Salix*. *Theoretical and Applied Genetics, 105*(2-3), 277-288. 10.1007/s00122-002-0918-0

Wang, J.-W., Park, M. Y., Wang, L.-J., Koo, Y., Chen, X.-Y., Weigel, D., & Poethig, R. S. (2011). MiRNA control of vegetative phase change in trees. *Plos Genetics, 7*(2) 10.1371/journal.pgen.1002012

Wang, X., Chen, Z., Li, Q., Zhang, J., Liu, S., & Duan, D. (2018). High-density SNP-based QTL mapping and candidate gene screening for yield-related blade length and width in *Saccharina japonica* (Laminariales, Phaeophyta). *Scientific Reports, 8* 10.1038/s41598-018-32015-y

Weih, M., Ronnberg-Wastljung, A. C., & Glynn, C. (2006). Genetic basis of phenotypic correlations among growth traits in hybrid willow (*Salix dasyclados* x *S-viminalis*) grown under two water regimes. *New Phytologist, 170*(3), 467-477. 10.1111/j.1469-8137.2006.01685.x

Welter, L. J., Gokturk-Baydar, N., Akkurt, M., Maul, E., Eibach, R., Topfer, R., & Zyprian, E. M. (2007). Genetic mapping and localization of quantitative trait loci affecting fungal disease resistance and leaf morphology in grapevine (*Vitis vinifera* L). *Molecular Breeding, 20*(4), 359-374. 10.1007/s11032-007-9097-7

Went, F. W. (1971). Parallel evolution. *Taxon, 20*, 197-226.

Wolf, J. B. (2000). Gene interactions from maternal effects. *Evolution, 54*(6), 1882-1898.

Woo, H. R., Chung, K. M., Park, J. H., Oh, S. A., Ahn, T., Hong, S. H., Jang, S. K., & Nam, H. G. (2001). ORE9, an F-box protein that regulates leaf senescence in Arabidopsis. *Plant Cell, 13*(8), 1779-1790. 10.1105/tpc.13.8.1779

Wu, G., Park, M. Y., Conway, S. R., Wang, J.-W., Weigel, D., & Poethig, R. S. (2009). The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell, 138*(4), 750-759. 10.1016/j.cell.2009.06.031

Xu, Y., Huang, L., Ji, D., Chen, C., Zheng, H., & Xie, C. (2015). Construction of a dense genetic linkage map and mapping quantitative trait loci for economic traits of a doubled haploid population of *Pyropia haitanensis* (Bangiales, Rhodophyta). *BMC Plant Biology, 15* 10.1186/s12870-015-0604-4

1.12 Appendix

Appendix 5.1. Summary of QTL results for the Lincoln F_2 population for different methods of analysis at an alpha of 0.05.

Trait	Marker regression	Interval mapping (Hayley-Knott regression)	Interval mapping (EM standard regression)	Non-parametric mapping	Two-D scan mapping	Multi-QTL mapping	Multi-QTL mapping with Two-D scan
Leaflet number	1@30 (P12) 5@135 (D39) 1@10 (CP)	1@28 (P12) 5@135 (D39) 1@10 (CP)	5@131 (D39) 1@10 (CP)	1@30 (P12) 3@107 (L66) 5@114 (D39) 1@10 (CP)	-	Q1 1@25 (P12) Q2 4@117(RMS4) Q3 5@30 (L15) Q4 5@60(C51) Q5 7@18 (L81) Q6 10 (CP)	Q1 1@30 Q2 1@129.7 Q3 4@114 Q4 5@22 Q5 5@40 Q6 5@90 Q7 6@21 Q8 7@21 Q9 8@22 Q5:Q8
Rachis length	1@7 (L117) 5@150 (C30) 8@0 (L76)	1@26 (P12) 5@142 (L63) 8@2 (L76)	1@25 (P12) 5@143 (L63) 8@2 (L76)	1@26 (P12) 5@139 (L63) 8@2 (L76)	1@40 (L7):8@21 (L62) 5@142 (L63): 8@6 (RMS1)	Q1 1@26 (P12) Q2 4@118(RMS4) Q3 5@32(L15) Q4 5@ 57.1(C51) Q5 5@150.5(C30) Q6 7@16(L17) Q7 8@1 (L76) Q1:Q6	Q1 1@87 Q2 1@113 Q3 3@4 Q4 4@118.6 Q5 5@32 Q6 5@58 Q7 5@151 Q8 7@94 Q9 8@0 Q5:Q8

Trait	Marker regression	Interval mapping (Hayley-Knott regression)	Interval mapping (EM standard regression)	Non-parametric mapping	Two-D scan mapping	Multi-QTL mapping	Multi-QTL mapping with Two-D scan
Leaflet length	2@69 (L38) 5@150 (C30) 8@0 (L76)	2@69 (L38) 5@142 (L63) 8@2 (L76)	2@69 (L38) 5@142 (L63) 8@0 (L76)	2@69 (L38) 5@138 (L63) 8@2 (L76)	2@69 (L38):8@5 (RMS1) 5@143 (L63):8@10 (RMS1)	Q1 1@38 (L10) Q2 2@73 (L38) Q3 3@67.9 (D64) Q4 4@127 (L77) Q5 5@23 (L84) Q6 5@151 (C30) Q7 6@25 (B21) Q8 8@11 (L3) Q5:Q6	Q1 1@126 Q2 2@36 Q3 3@51 Q4 4@139 Q5 5@19 Q6 5@57.1 Q7 5@139 Q8 8@4 Q4:Q5
Leaflet width	2@69 (L38) 5@152 (C30)	2@73 (L20) 5@152 (L56)	2@72 (L20)	2@73 (L20) 5@140 (L63) 8@6 (RMS1)	2@74 (L20):2@115 (ZIG) 2@55 (L8):8@11 (L3)	Q1 2@75 (L20) Q2 2@119.8 (L68) Q3 3@62 (L64) Q4 3@134 (L103) Q57@103.1(L113 Q6 8@8.5 (RMS1) Q3:Q5 Q2:Q4	Q1 2@52 Q2 2@75 Q3 2@118 Q4 3@58 Q5 3@120 Q6 5@88 Q7 7@69.6 Q8 7@103.1 Q9 8@8.5 Q4:Q8 Q3:Q5 Q1:Q6
Internode length	-	-	-	-	-	-	-
Branch width	-	-	-	-	-	-	-
Branch number	-	-	-	-	-	-	-

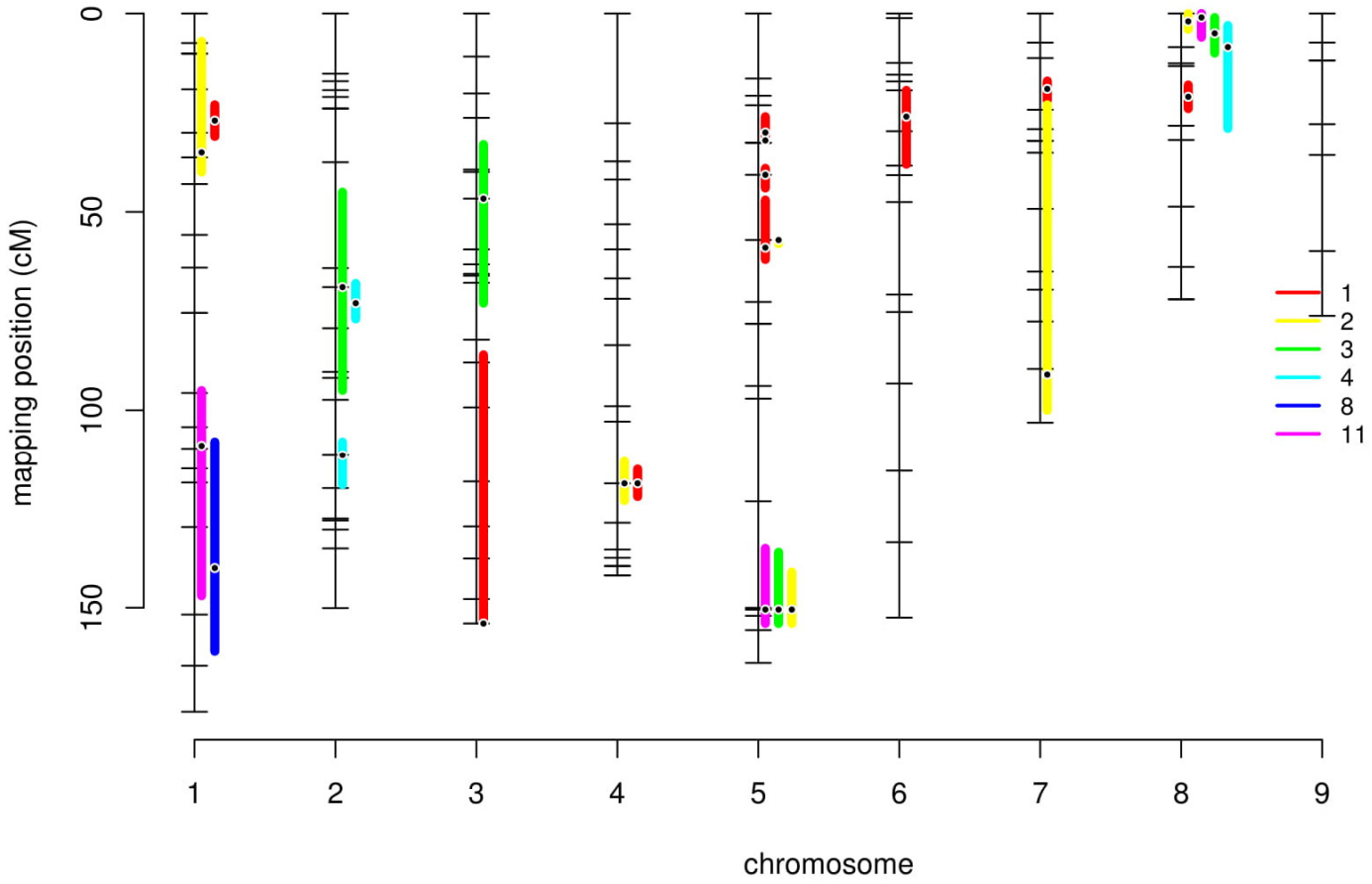
Trait	Marker regression	Interval mapping (Hayley-Knott regression)	Interval mapping (EM standard regression)	Non-parametric mapping	Two-D scan mapping	Multi-QTL mapping	Multi-QTL mapping with Two-D scan
Node angle	CP	1@144 (LI18) 3@104 (L66) 1@10 (CP)	1@141 (LI18) 3@105 (L66)	1@144 (LI18) 1@10 (CP)	-	Q1 1@144 (LI18)	Q1 1@144
Branch angle	-	-	-	1@61 (B79)	3@88 (B78):4@80 (L37)	-	-
PC1	5@150 (C30)	1@26 (P12) 5@140 (L63) 8@3 (L76)	1@26 (P12) 5@137 (L63)	1@26 (P12) 5@138 (L63) 8@3 (L76)	1@142 (LI18):8@0 (L76) 5@139 (D39):8@7 (RMS1) 5@137 (L63):9@12 (L97)	Q1 1@109.7 (L39) Q2 2@29 (LI6) Q3 5@139 (L63) Q4 8@1 (L76) Q5 9@11.8 (L82) Q1:Q2	
PC2	-	-	-	-	-	-	Q1 4@37.3 Q2 6@15.4 Q3 6@70 Q4 9@9 Q3:Q4
Divaricate Index	-	-	-	1@24 (LI15)	-	-	-

Appendix 5.2. QTL results from mapping with maternal cross direction as a covariate

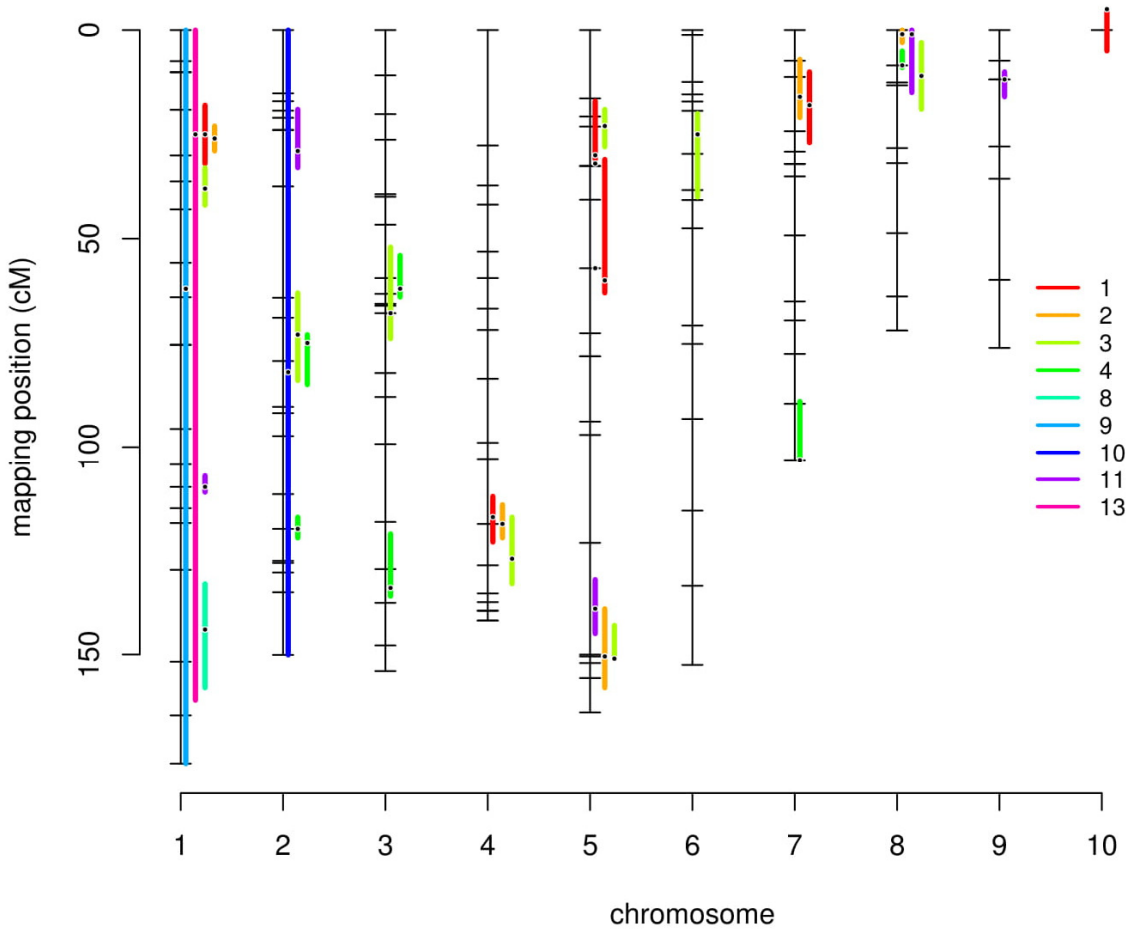
Trait	Marker regression	Interval mapping (Hayley-Knott regression)	Non-parametric mapping	Multi-QTL mapping	QTL effects	Multi-QTL mapping with 2D scan	Multi-QTL mapping at 10% significance
Leaflet number	7@7.35	7@15	1@30 3@107 5@114	Q1 1@27 Q2 3@154 Q3 4@118.6 Q4 5@30 Q5 5@40.7 Q6 5@59 Q7 6@26 Q8 7@19 Q9 8@21 Q5:Q8	Q1 – 7.603 Q2 – 3.351 Q3 – 3.227 Q4 – 10.146 Q5 – 11.626 Q6 – 5.718 Q7 – 2.689 Q8 – 18.373 Q9 – 3.683	Q1 1@24 Q2 2@111.5 Q3 4@117 Q4 5@22 Q5 5@63 Q6 7@20 Q7 8@16	
Rachis length	5@150.46 8@8.47	1@26 5@152 8@3	1@26 5@139 8@2	Q1 1@35 Q2 4@118.6 Q3 5@32 Q4 5@57.1 Q5 5@150.5 Q6 7@91 Q7 8@2 Q3:Q6	Q1 – 2.964 Q2 – 3.562 Q3 – 18.865 Q4 – 14.141 Q5 – 5.177 Q6 – 8.641 Q7 – 13.477	Q1 1@86 Q2 1@113 Q3 3@4 Q4 4@118.6 Q5 5@32 Q6 5@58 Q7 5@151 Q8 7@94 Q9 8@0 Q5:Q8	

Trait	Marker regression	Interval mapping (Hayley-Knott regression)	Non-parametric mapping	Multi-QTL mapping	QTL effects	Multi-QTL mapping with 2D scan	Multi-QTL mapping at 10% significance
Leaflet length	2@69 5@150	2@69 4@123 5@151 8@3	2@69 5@138 8@2	Q1 2@69 Q2 3@46.7 Q3 5@150.5 Q4 8@5	Q1 – 8.966 Q2 – 9.948 Q3 – 16.962 Q4 – 14.431	Q1 1@118 Q2 2@44 Q3 3@49 Q4 4@63 Q5 4@125 Q6 5@21 Q7 5@54 Q8 5@140 Q9 8@3 Q4:Q6	
Leaflet width	2@69	2@73	2@73 5@140 8@6	Q1 2@73 Q2 2@111.5 Q3 8@8.5	Q1 – 15.114 Q2 – 6.996 Q3 – 12.632	Q1 1@93 Q2 1@96 Q3 2@76 Q4 2@116 Q5 3@59 Q6 4@27.7 Q7 5@72.7 Q8 5@146 Q9 7@103.1 Q10 8@10 Q5:Q9 Q7:Q8	
Internode length	-	-	-	-	-	-	-

Trait	Marker regression	Interval mapping (Hayley-Knott regression)	Non-parametric mapping	Multi-QTL mapping	QTL effects	Multi-QTL mapping with 2D scan	Multi-QTL mapping at 10% significance
Branch width	-	-	-	-	-	-	-
Branch number	-	-	-	-	-	-	-
Node angle	-	1@140	1@144	Q1 1@140	Q1 – 22.867	Q1 1@140	Q1 – 1@109.7 Q2 – 3@108 Q3 – 8@0
Branch angle	-	-	-	-	-	-	-
PC1	5@150	5@152	1@26 5@138 8@3	Q1 1@109.0 Q2 5@150.5 Q3 8@1	Q1 – 10.56 Q2 – 17.61 Q3 – 20.67	Q1 - 1@109.7 Q2 – 2@30 Q3 – 4@138 Q4 – 5@26 Q5 – 5@52 Q6 – 5@144 Q7 – 8@12 Q8 – 9@11.8 Q1:Q2 Q3:Q4	
PC2	-	-	-		-	-	Q1 – 6@20 Q2 – 9@9
Divaricate Index	-	-	1@24	-	-	-	



Appendix 5.3. Genetic map of QTL mapping at an alpha of 0.05 for the Lincoln F_2 population including maternal cross direction as a covariate indicating QTL and confidence intervals for divaricate traits. Traits are indicated by the colours as in the legend: 1 – leaflet number, 2 – rachis length, 3 – leaflet length, 4, - leaflet width, 8 – node angle, I1 – PC1.



Appendix 5.4. Plot of the QTL, at an alpha of 0.05, mapped on the *Sophora* linkage maps. Colours represent phenotypes with 1- Leaflet number, 2 – rachis length, 3 – leaflet length, 4 – leaflet width, 8 – node angle, 9 – branch angle, 10 – leaf density, 11 – PC1 and 13 – divaricate index score.

Appendix 5.5. Results from QTL mapping with MQM interval mapping for the Lincoln F_2 population

Phenotype	Rqtl-MQM	MQM with no cofactors
Leaflet number	5@60 (C51) 7@15 (L17)	5@130
Rachis length	5@55 (C51) 8@15 (L3)	5@145 8@15
Leaflet length	5@150 (C30) 8@10 (L3)	5@145
Leaflet width	1@145(L118) 2@60 (L8) 8@0 (L76)	-
Internode length	-	-
Branch width	-	-
Branch number	-	-
Node angle	1@135 (L125)	-
Branch Angle	-	-
Leaf density	1@90(L94) 2@85 (L20)	-
PC1	5@60(C51) 7@20(L81) 8@10(RMS1)	5@135
PC2	-	-
Divaricate index	5@25(C67)	-

Chapter 6 – Thesis summary and future directions

6.1 Thesis summary

Divarication is a plant growth form defined by Kelly (1994) as a small tree or shrub that has "many interlaced branches with wide angles (mean >60°), small leaves (<60mm^2), and widely spaced leaves (mean distance >2x leaf width) with larger leaves towards the interior of the plant (inner leaves >1.4x larger than outer leaves)" (p.509) giving a dense, tangled, shrubby appearance. It is a unique feature of the New Zealand flora occurring at a higher frequency than other floras worldwide. It is found in ~ 18 different angiosperm families (~ 20 genera), making up ~ 10% of the New Zealand flora (Greenwood & Atkinson, 1977; Atkinson & Greenwood, 1989). The frequency of divarication is a long-standing mystery of the New Zealand biota with two main theories proposed on why divarication evolved: as a defence against moa browsing or adaptation to climatic factors; however, the genetic basis behind divarication remains unexplored. In this thesis the genetic architecture of divarication in the genus *Sophora* was investigated.

Sophora (Fabaceae) is one of the many genera in New Zealand with divaricating species: *S. prostrata* and a heteroblastic species, *S. microphylla,* with a divaricating juvenile form. *Sophora* species readily hybridise naturally (Heenan et al., 2001), enabling easy creation of hybrids for experimental work. An F_2 population formed from reciprocal crosses between *S. prostrata* and *S. tetraptera,* a non-divaricate tree, has been maintained by Manaaki Whenua Landcare Research in Lincoln, New Zealand. The creation of this population began ~40 years ago by Dr Eric Godley to investigate the origins of the heteroblast *S. microphylla;* however, the population also displays segregation for divarication. Developing segregating populations in woody, long-lived

species is time consuming, sometimes taking years or even decades, therefore, this existing F_2 population provides an excellent resource to investigate the genetic architecture of divarication in *Sophora.* A quantitative trait loci (QTL) approach was undertaken, using this F_2 population, with the aim to characterize the genetic architecture of the divaricate growth form for the first time. The results presented in this thesis are summarised and future directions considered here.

6.1.1 Phenotyping divaricate traits in the *Sophora* F_2

Nine phenotypic traits were measured for the 88 individuals in the F_2 population and all displayed fairly normal distributions. The leaf traits, branch width, and node angle were all positively correlated and all negatively correlated with other branch traits, such as branch number and branch width. This is consistent with the phenotypes of the parent species where the divaricate *S. prostrata,* has smaller leaves, smaller node angles, thinner stems and increased branching, and vice versa in *S. tetraptera*. Branch angle did not correlate with any other traits. Correlated traits are often observed mapping to the same loci in QTL analyses (Xu et al., 2015; Chunlian et al., 2016; Wang et al., 2018). These trait correlations in the *Sophora* F_2 suggest that many traits are under common genetic control, potentially by pleiotropic loci whereas traits that do not correlate, such as branch angle, may be regulated by different loci. These traits were also measured in a second F_2 population, the Palmerston North F_2 (PNF_2), which was developed in 2017 from the F_1 individual, 10981-2-9, and grown in controlled conditions in a greenhouse. These plants are much younger than the original Lincoln F_2 population, and one trait, branch number, had to be measured differently, however the correlations seen in the Lincoln population were generally consistent with the PNF_2 population.

6.1.2 Strigolactone and divarication

Increased branching is an important feature of the divaricate growth form. One pathway known to be involved in controlling axillary branching involves the hormone strigolactone (McSteen & Leyser, 2005; Johnson et al., 2006; Gomez-Roldan et al., 2008; Mueller & Leyser, 2011; Wang et al., 2013). Five genes have been identified in *Pisum*, and other species, that are involved in control of axillary branching through strigolactone synthesis and perception: *RMS1, RMS2, RMS3, RMS4* and *RMS5*. These have been previously suggested to play a role in divarication (Grierson, 2014). Four of these genes, *RMS1, RMS2, RMS4* and *RMS5,* were partially sequenced in both *S. tetraptera* and *S. prostrata* and developed into genetic markers that were included in the genetic maps and QTL analyses. In the Lincoln F_2 population, all gene markers showed significant associations with some of the divaricate phenotypes. *RMS1* displayed significant associations for leaflet length, width, rachis length, internode length, node angle and branch width, suggesting *RMS1* may play a role in multiple divaricate traits. *RMS1* was also genotyped in the PN F_2 population where associations were again seen with multiple traits, including leaflet number, length, width, rachis length, node angle, internode length and branch number.

RMS1 is one of the genes involved in the biosynthesis of strigolactone and *rms1* mutants lack, or have reduced levels, of strigolactone. The association of multiple traits with *RMS1* in the *Sophora* F_2 suggests the divaricate form of *S. prostrata* may be influenced by a lack of strigolactone. The application of synthetic strigolactone, GR24, can revert *rms1* mutants to wildtype in many species such as *Pisum* and *Arabidopsis* (Gomez-Roldan et al., 2008; Umehara et al., 2008) and, if *RMS1* is involved in the divaricate growth form, application of GR24 to *S. prostrata* may be able to replace the lack of strigolactone and revert plants to a less divaricating form. However, synthetic strigolactone did not prevent branching in *S. prostrata* seedlings, but did result in a

significant difference in leaflet length and width. However, other *RMS* genes, such as *RMS4,* also showed significant associations with some divaricate traits. *RMS4* is involved in signal perception and response to strigolactone (Stirnberg et al., 2007; Jiang et al., 2013; Bennett & Leyser, 2014; Soundappan et al., 2015). If both *RMS1* and *RMS4* were non-functional in *S. prostrata*, no change in branching would be observed following synthetic strigolactone treatment even if there was a lack of strigolactone originally as it could not be perceived.

GR24 has been shown to activate non-strigolactone signalling pathways in plants (Scaffidi et al., 2014; Waters et al., 2017) and this may be responsible for the significant differences in leaflet width and length in treated and control plants. A smaller test for *S. tetraptera* also showed a significant difference in leaflet width between treated and control plants and suggests these results may not be from the strigolactone signalling pathway; however, this was a much smaller sample size so results are not conclusive.

6.1.3 Generating a linkage map for the Lincoln F_2

To perform QTL mapping, a genetic map is required. SNP markers were developed for the Lincoln F_2 population using MRMEseq (McLay, 2018). In total, with the five candidate gene markers, 148 molecular markers were used to generate the first *Sophora* linkage maps. The markers were assigned to nine linkage groups, as expected for *Sophora,* which has a chromosome number of $2n$=18, and had a total length of 1189.9cM. This genetic map provides a new resource for genetic research in *Sophora* and the Fabaceae. The *Sophora* F_2 map and the phenotype data were analysed, using QTL analyses, to identify QTL for divaricate-related traits.

6.1.4 QTL for divaricate traits in the Lincoln F_2

Twelve traits were analysed by QTL mapping for the Lincoln F_2 population including the nine divaricate phenotypes (Chapter 2) and two ways of combining traits

to a single value, including PC1 and PC2 from a PCA, and an index of divarication developed from the nine measured divaricate-related traits. QTL were located for most of the traits except internode length and branch number. The small sample size may have contributed to the lack of QTL, especially if there are many QTL with small effects contributing to these traits (Tanksley, 1993; Lynch & Walsh, 1998; Collard et al., 2005); however, environmental effects and age of the branches measured may also contribute to difficulty in detecting QTL.

Multiple QTL were located across all linkage groups for the remaining traits, indicating that multiple loci contribute to producing the divaricate phenotype. Many of the QTL located for the leaf traits were at the same location, suggesting that pleiotropic loci may control leaf characteristics, which supports the strong correlations observed among leaf traits. Node angle and branch width also were correlated with the leaf traits and also identified QTL that co-located with leaf trait QTL at an alpha of 0.10. Branch angle did not display correlations with other traits or map to QTL with other traits, except for the combined data PC1, PC2 and divaricate index, suggesting it is under genetic control by different loci.

For the combined data, PC1 uses the first principle component of a PCA of all nine traits (Chapter 2) and identified five QTL. Two QTL map near other QTL positions for several individual leaf traits, which may be expected as the leaf characters explain much of the variance along this axis. Node angle and branch width also map to these loci. One QTL co-locates with a QTL for branch angle and the remaining two locate with PC2. PC2, the second principle component, identified 5 QTL at an alpha of 0.20; two of these locate with branch angle, which explains much of the variance of the PC2 axis. The divaricate index mapped three QTL, one mapped to the pleiotropic locus on linkage group 5. These combined traits mostly map to a locus with at least one of the individual

traits, and many map to pleiotropic loci, which provides support for these representing the divaricate form and suggest these loci contribute to multiple traits in the divaricate form.

QTL were mapped to the *RMS1* and *RMS4* loci for several leaf traits, branch width, and node angle, indicating possible involvement of strigolactone signalling in divarication. While rms mutants are often described for the increased branching phenotype, the *Pisum rms1* mutant also displays pleiotropic phenotypes, such as thinner stems, shorter internodes and shorter leaflet lengths (Beveridge et al., 1997), indicating leaf characters may also be influenced by *RMS1*. The *rms4* mutants display fewer pleiotropic phenotypes except for shorter stems and more rounded leaflets (Beveridge et al., 1996). Phenotype associations with *RMS1* were observed for all traits except leaflet number, branch number and branch angle with leaflet number the only leaf trait that does not map near *RMS1,* supporting the associations seen for leaf phenotypes and *RMS1*. Only leaflet length and node angle indicated associations with *RMS4;* however, leaflet number, rachis length and leaflet length all mapped QTL near *RMS4*. The *RMS1* sequence from *S. prostrata* has a predicted deleterious amino acid change at what appears to be a mostly conserved region in the protein (see Chapter 3), indicating that *RMS1* could be non-functional in *S. prostrata* and may be involved in the divaricate growth form in *S. prostrata.* Mapping to a particular gene marker does not necessarily indicate that the gene is involved in the traits, as one of the many genes linked to these markers could be responsible, and therefore, further research is required to determine if the strigolactone pathway is involved in the divaricate growth form.

Overall, many QTL displayed dominance of the *S. prostrata* allele compared to the *S. tetraptera* allele, which suggests that dominant mutations may have played an important role in the evolution of the divaricate form if divarication evolved from a non-

divaricate ancestor. This can result in a more rapid evolution of the phenotype than with recessive alleles as phenotypes conveyed by the dominant allele are immediately exposed to natural selection (Teshima & Przeworski, 2006). New dominant mutations are rare, with the majority of new mutations being recessive compared to the wild-type (Orr, 1991) and would make the large number of dominant QTL for *S. prostrata* unusual. However, the relationships among *Sophora* species are still unresolved and therefore, while *S. prostrata* likely evolved from a non-divaricate species, *S. tetraptera* may not be the closest relative and may have developed new recessive mutations in the evolution of its architecture. This scenario would result in the *S. prostrata* alleles being dominant but not necessarily in the context of new mutations. The presence of multiple loci contributing to the divaricate form indicate that for the growth form to develop and be maintained, selection would be required to act on multiple loci simultaneously.

6.2 Future directions

This thesis investigated the genetic basis of the divaricate growth form in *Sophora*. The results suggest that multiple loci contribute to this architecture in *Sophora;* however, divarication occurs in multiple plant families (~ 18 different plant families) in New Zealand and it is unknown how much the genetic architecture of the divaricate form varies among species. Divarication has proven difficult to define (Kelly, 1994) due to the variation of traits seen among different divaricate species. If multiple loci are controlling the divaricate phenotype, then differences in the divaricate phenotype could arise from differences in the loci or alleles in each species. For example, node angle located a QTL in the *Sophora* F_2 that contributes to the zig-zag shaped stems distinctive of *S. prostrata.* However, not all divaricates have zig-zag stems (Tomlinson, 1978), therefore this locus may not be contributing to divarication in other species, which may explain some of the difficulty in defining the divaricate form among different divaricate

species. To investigate the genetic architecture of divarication in other families, further research is required on divaricating species from other genera.

Generating mapping populations for long-lived woody species, such as divaricates, is time consuming and therefore, often not possible for these species. Generating more markers to create a denser genetic map may increase the power to detect QTL, by decreasing the distance from potential QTL to markers, and allow for fine-mapping of QTL. However, increasing the number of markers may not increase power in a sample size with a limited number of recombination events (Liu, 1998) such as the *Sophora* population. Association mapping, such as GWAS (genome-wide association mapping), is a potential alternative as it uses natural variation and does not require a related mapping population. However, GWAS has other limitations, such as requiring large numbers of markers, bias from selection and population history, small effect loci, rare allele frequencies, and non-random mating that can be problematic, although some of these issues such as relatedness, can be accounted for (Khan & Korban, 2012; Korte & Farlow, 2013). GWAS, in plants, is often used to identify variation within a species, using cultivars, however, it has been performed comparing phenotypic traits for different bird species (Grossen et al., 2016; Silva et al., 2017) and butterflies (Lucas et al., 2018). This method could be used to investigate the genetic basis of divarication in wild divaricate and closely-related non-divaricating populations, in *Sophora* or other divaricating species, where generating mapping populations may not be feasible.

Another potential direction is transcriptome analysis comparing divaricate and non-divaricate species, for example in *S. tetraptera* and *S. prostrata.* Transcriptomics can identify differences in gene expression between these species that may be associated with divarication. However, with the relationships among *Sophora* unresolved, there may be many differences in gene expression that are unrelated to divarication in these

species. The heteroblastic species, *S. microphylla,* provides a unique opportunity to investigate divarication within a single plant which has a divaricating juvenile. Transcriptomics on *S. microphylla* juvenile and adult stages to identify differential gene expression between these different growth stages may be insightful into the genetic control of the divaricate growth form.

Alternatively, investigating candidate genes, identified from the *Sophora* population, in other divaricates may identify their possible involvement in other divaricates. For example, identifying deleterious mutations or changes to expression patterns of *RMS* orthologs in other species may be a fruitful approach. Further work to confirm the involvement of candidate genes in *Sophora* may be required before extending this to other species. The candidate gene, *RMSI,* is indicated to be a strong candidate gene involved in divaricate traits as multiple traits and PC1 mapped QTL at this marker. Significant associations were also seen between phenotypes and *RMSI,* and sequence analyses predicted a deleterious amino acid replacement in *S. prostrata.* Obtaining full gene sequences for these *RMS* genes would help to identify all differences between *S. prostrata* and *S. tetraptera* orthologs in each of these genes and allow for further research on functionality of these genes. More samples from each species would also determine if these mutations are particular to an individual plant or consistent within species.

Transforming *Arabidopsis* or *Pisum,* as it is more closely related, mutants with the alleles from both *S. prostrata* and *S. tetraptera* could reveal if these genes are still functional, or precisely if *S. prostrata* is non-functional and could provide further evidence of *RMS* genes in divarication. The *RMS*/strigolactone pathway is conserved among plants with orthologs found in many diverse species (Johnson et al., 2006) and transformations of *RMS*/*MAX* genes from *Populus* were able to rescue or partially rescue

Arabidopsis mutants (Czarnecki et al., 2014), indicating this could be an ideal experiment to perform using *Sophora RMS* genes. As well, measuring expression levels of the *RMS* genes in *S. tetraptera, S. prostrata* and potentially the juvenile and adult stages of *S. microphylla* may also be helpful in determining the functionality of these genes in these species or life stages. Measuring gene expression of the *RMS* genes in other divaricate and closely related non-divaricates species could also help determine if the *RMS* genes contribute to divarication in other divaricate species.

6.3 Conclusion

The high frequency of the divaricating plant growth form in New Zealand has been a long-standing mystery with nothing known about its genetic architecture. The results, presented in this thesis, suggest divarication occurs from the effects of multiple loci, some acting on individual traits and some that act on multiple traits. *RMSI* was identified as a potential candidate gene for the divaricate growth form. As divarication is present in multiple plant families in New Zealand, one further question is whether the genetic architecture of divarication is consistent among all divaricate species or different for each species that developed this growth form. On a larger scale, this contributes to our general knowledge of plant architecture. Plant architecture is often used in taxonomy and is important in agriculture, affecting factors such as harvest efficiency (Takeda et al., 2013) and plant stability (Quine, 1990; Plourde et al., 2009). An improved understanding of the genetic basis of plant architecture contributes not only to agricultural and horticultural applications but also to issues of phenotype evolution.

6.4 References

Atkinson, I. A. E., & Greenwood, R. M. (1989). Relationships between moas and plants. *New Zealand Journal of Ecology, 12*, 67-96.

Bennett, T., & Leyser, O. (2014). Strigolactone signalling: standing on the shoulders of DWARFs. *Current Opinion in Plant Biology, 22*, 7-13. 10.1016/j.pbi.2014.08.001

Beveridge, C. A., Ross, J. J., & Murfet, I. C. (1996). Branching in pea - Action of genes *RMS3* and *RMS4*. *Plant Physiology, 110*(3), 859-865. 10.1104/pp.110.3.859

Beveridge, C. A., Symons, G. M., Murfet, I. C., Ross, J. J., & Rameau, C. (1997). The rms1 mutant of pea has elevated indole-3-acetic acid levels and reduced root-sap zeatin riboside content but increased branching controlled by graft-transmissible signal(s). *Plant Physiology, 115*(3), 1251-1258.

Chunlian, L., Guihua, B., Shiaoman, C., Carver, B., & Zhonghua, W. (2016). Single nucleotide polymorphisms linked to quantitative trait loci for grain quality traits in wheat. *Crop Journal, 4*(1), 1-11. 10.1016/j.cj.2015.10.002

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica, 142*(1-2), 169-196. 10.1007/s10681-005-1681-5

Czarnecki, O., Yang, J., Wang, X., Wang, S., Muchero, W., Tuskan, G. A., & Chen, J.-G. (2014). Characterization of MORE AXILLARY GROWTH genes in *Populus*. *Plos One, 9*(7) 10.1371/journal.pone.0102757

Gomez-Roldan, V., Fermas, S., Brewer, P. B., Puech-Pages, V., Dun, E. A., Pillot, J.-P., Letisse, F., Matusova, R., Danoun, S., Portais, J.-C., Bouwmeester, H., Becard, G., Beveridge, C. A., Rameau, C., & Rochange, S. F. (2008). Strigolactone inhibition of shoot branching. *Nature, 455*(7210), 189-U122. 10.1038/nature07271

Greenwood, R. M., & Atkinson, I. A. E. (1977). Evolution of divaricating plants in New Zealand in relation to moa browsing. *Proceedings of the New Zealand Ecological Society, 24*

Grierson, E. R. P. (2014). *The Development and Genetic Variation of Sophora prostrata - A New Zealand Divaricating Shrub.* (Master's Thesis), The University of Waikato, Hamilton, New Zealand.

Grossen, C., Seneviratne, S. S., Croll, D., & Irwin, D. E. (2016). Strong reproductive isolation and narrow genomic tracts of differentiation among three woodpecker species in secondary contact. *Molecular Ecology, 25*(17), 4247-4266. 10.1111/mec.13751

Heenan, P. B., de Lange, P. J., & Wilton, A. D. (2001). *Sophora* (Fabaceae) in New Zealand: Taxonomy, distribution, and biogeography. *New Zealand Journal of Botany, 39*(1), 17-53. 10.1080/0028825x.2001.9512715

Jiang, L., Liu, X., Xiong, G., Liu, H., Chen, F., Wang, L., Meng, X., Liu, G., Yu, H., Yuan, Y., Yi, W., Zhao, L., Ma, H., He, Y., Wu, Z., Melcher, K., Qian, Q., Xu, H. E., Wang, Y., & Li, J. (2013). DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature, 504*(7480), 401-+. 10.1038/nature12870

Johnson, X., Brcich, T., Dun, E. A., Goussot, M., Haurogne, K., Beveridge, C. A., & Rameau, C. (2006). Branching genes are conserved across species. Genes controlling a novel signal in pea are coregulated by other long-distance signals. *Plant Physiology, 142*(3), 1014-1026. 10.1104/pp.106.087676

Kelly, D. (1994). Towards a numerical definition of divaricate (interlaced small-leaved) shrubs. *New Zealand Journal of Botany, 32*(4), 509-518.

Khan, M. A., & Korban, S. S. (2012). Association mapping in forest trees and fruit crops. *Journal of Experimental Botany, 63*(11), 4045-4060. 10.1093/jxb/ers105

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods, 9* 10.1186/1746-4811-9-29

Liu, H. (1998). *Statistical Genomics, Linkage, Mapping and QTL Analysis*. Boca Raton, Florida: CRC.

Lucas, L. K., Nice, C. C., & Gompert, Z. (2018). Genetic constraints on wing pattern variation in Lycaeides butterflies: A case study on mapping complex, multifaceted traits in structured populations. *Molecular Ecology Resources, 18*(4), 892-907. 10.1111/1755-0998.12777

Lynch, M., & Walsh, B. (1998). *Genetics and Analysis of Quantative Traits*. Sunderland, MA, U.S.A: Sinauer Associates, Inc.

McLay, T. (2018). Hybridisation between species of *Xanthorrhoea* in Yuraygir National Park revealed using a newly developed next-generation sequencing method. *Australian Systematic Botany Society Newsletter, 175,* 13-16.

McSteen, P., & Leyser, O. (2005). Shoot branching. In *Annual Review of Plant Biology* (Vol. 56, pp. 353-374). 10.1146/annurev.arplant.56.032604.144122

Mueller, D., & Leyser, O. (2011). Auxin, cytokinin and the control of shoot branching. *Annals of Botany, 107*(7), 1203-1212. 10.1093/aob/mcr069

Orr, H. A. (1991). A test of Fisher's theory of dominance. *Proceedings of the National Academy of Sciences of the United States of America, 88*(24), 11413-11415. 10.1073/pnas.88.24.11413

Plourde, A., Krause, C., & Lord, D. (2009). Spatial distribution, architecture, and development of the root system of *Pinus banksiana* Lamb. in natural and planted stands. *Forest Ecology and Management, 258*(9), 2143-2152. 10.1016/j.foreco.2009.08.016

Quine, C. P. (1990). Planting at stump - preliminary evidence of its effect on root architecture and tree stability. *Research Information Note - Forestry Commission Research Division*(169), 3 pp.

Scaffidi, A., Waters, M. T., Sun, Y. K., Skelton, B. W., Dixon, K. W., Ghisalberti, E. L., Flematti, G. R., & Smith, S. M. (2014). Strigolactone hormones and their stereoisomers signal through two related receptor proteins to induce different physiological responses in *Arabidopsis*. *Plant Physiology, 165*(3), 1221-1232. 10.1104/pp.114.240036

Silva, C. N. S., McFarlane, S. E., Hagen, I. J., Ronnegard, L., Billing, A. M., Kvalnes, T., Kemppainen, P., Ronning, B., Ringsby, T. H., Saether, B. E., Qvarnstrom, A., Ellegren, H., Jensen, H., & Husby, A. (2017). Insights into the genetic architecture of morphological traits in two passerine bird species. *Heredity, 119*(3), 197-205. 10.1038/hdy.2017.29

Soundappan, I., Bennett, T., Morffy, N., Liang, Y., Stang, J. P., Abbas, A., Leyser, O., & Nelson, D. C. (2015). SMAX1-LIKE/D53 family members enable distinct MAX2-dependent responses to strigolactones and karrikins in *Arabidopsis*. *Plant Cell, 27*(11), 3143-3159. 10.1105/tpc.15.00562

Stirnberg, P., Furner, I. J., & Ottoline Leyser, H. M. (2007). MAX2 participates in an SCF complex which acts locally at the node to suppress shoot branching. *Plant Journal, 50*(1), 80-94. 10.1111/j.1365-313X.2007.03032.x

Takeda, F., Krewer, G., Li, C. Y., MacLean, D., & Olmstead, J. W. (2013). Techniques for increasing machine harvest efficiency in highbush blueberry. *Horttechnology, 23*(4), 430-436.

Tanksley, S. D. (1993). Mapping polygenes. *Annual Review of Genetics, 27*, 205-233. 10.1146/annurev.ge.27.120193.001225

Teshima, K. M., & Przeworski, M. (2006). Directional positive selection on an allele of arbitrary dominance. *Genetics, 172*(1), 713-718. 10.1534/genetics.105.044065

Tomlinson, P. S. (1978). Some qualitative and quantitive aspects of New Zealand divaricating shrubs. *New Zealand Journal of Botany, 16*, 299-309.

Umehara, M., Hanada, A., Yoshida, S., Akiyama, K., Arite, T., Takeda-Kamiya, N., Magome, H., Kamiya, Y., Shirasu, K., Yoneyama, K., Kyozuka, J., & Yamaguchi, S. (2008). Inhibition of shoot branching by new terpenoid plant hormones. *Nature, 455*(7210), 195-U129. 10.1038/nature07272

Wang, X., Chen, Z., Li, Q., Zhang, J., Liu, S., & Duan, D. (2018). High-density SNP-based QTL mapping and candidate gene screening for yield-related blade length and width in *Saccharina japonica* (Laminariales, Phaeophyta). *Scientific Reports, 8* 10.1038/s41598-018-32015-y

Wang, Y., Sun, S., Zhu, W., Jia, K., Yang, H., & Wang, X. (2013). Strigolactone/MAX2-induced degradation of brassinosteroid transcriptional effector BES1 regulates shoot branching. *Developmental Cell, 27*(6), 681-688. 10.1016/j.devcel.2013.11.010

Waters, M. T., Gutjahr, C., Bennett, T., & Nelson, D. C. (2017). Strigolactone signaling and evolution. In S. S. Merchant (Ed.), *Annual Review of Plant Biology, Vol 68* (Vol. 68, pp. 291-322). 10.1146/annurev-arplant-042916-040925

Xu, Y., Huang, L., Ji, D., Chen, C., Zheng, H., & Xie, C. (2015). Construction of a dense genetic linkage map and mapping quantitative trait loci for economic traits of a doubled haploid population of *Pyropia haitanensis* (Bangiales, Rhodophyta). *BMC Plant Biology, 15* 10.1186/s12870-015-0604-4