Molecular breeding for resistance to biotic threats in kiwifruit (*Actinidia chinensis*)

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

Doctor of philosophy

in

Plant Biology

at Massey University, Palmerston North, Manawatu, New Zealand

Casey Flay

2023

i. Abstract

This thesis takes a genomic approach to assist kiwifruit breeding programs by identifying genomic loci that contribute to resistance to biotic threats. Three different biotic threats were investigated that represent some of the most destructive to the kiwifruit industry. These were made up of two pests including latania scale (*Hemiberlesia lataniae*) and Brown headed leafroller (*Ctenopseustis obliquana*) and the pathogen Psa (*Pseudomonas syringae* pv. *actinidiae*). Together these pests and pathogen represent the predominant feeding mechanisms of, sucking from *H. lataniae*, chewing from *C. obliquana*, and infection from the necrotrophic pathogen Psa. The inclusion of these pests and pathogen was thought to provide a well-rounded knowledge of the molecular mechanisms that influence both monogenic and polygenic resistance traits in kiwifruit. Moreover, because these experiments were conducted on multiple kiwifruit families at different sites over time with multiple interactions occurring between organisms, an appreciation was gained for the sources of error that affect experiments, and the effects of those errors on the interpretation of the resulting genomic data.

The projects that make up this thesis first analyse kiwifruit (*Actinidia chinensis* var. *chinensis*) families of different sizes to identify the molecular architecture for resistance to *H. lataniae*. Once the monogenic region for resistance to *H. lataniae* was identified, existing datasets were interrogated for individuals with the loci associated with resistance. Furthermore, due to an absence of *H. lataniae* resistance in the elite tetraploid gold fleshed *A. chinensis* var. *chinensis* breeding program, a diploid individual resistant to *H. lataniae* that produced unreduced gametes was crossed with tetraploid males and screened for tetraploid individuals that contained markers for *H. lataniae* resistance.

Unfortunately, the experiments investigating the molecular architecture for resistance of *A. chinensis* var. *chinensis* to *C. obliquana* using genotyping by sequencing (GBS) data was unsuccessful due to multiple factors discussed in chapter 4.

The last project broke away from the traditional association mapping approach to identify quantitative trait loci (QTL) for Psa resistance from several *A. chinensis* families that had lost individuals due to Psa. This fast, inexpensive, and powerful approach used a modified bulked segregant analysis (BSA) with bulking methods similar to that of selection mapping. The methodology was validated using pools created with a variable number of males and the sensitivity of detection determined by identifying QTL at the known sex locus. The full experiment then analysed whole genome sequence (WGS) data from eight pools containing multiple families to identify QTL associated with survival to Psa.

Together these experiments will advance the kiwifruit breeding program by providing the molecular architecture for two of the most prevalent biotic threats affecting kiwifruit production. Utilising the markers for resistance to *H. lataniae* and developing markers to target the QTL for resistance to Psa will enable breeding programs to use these markers as selection criteria in commercial breeding programs, increasing the rate at which cultivars resistant to these biotic threats can be developed.

ii. Acknowledgements

Thank you to Plant & Food Research for providing the scholarship and allowing this work to proceed. Where work has not been specifically attributed to others, the work was conducted by the author. I am grateful to my Massey University supervisor Dr Vaughan Symonds and co-supervisor Dr Svetla Sofkova-Bobcheva for their help and guidance. I thank the team at Plant and Food Research, Te Puke, New Zealand for guiding me through the kiwifruit breeding system, Jibran Tahir who provided the GBS sequence data for the 'Hort16A'x Male 7 population and the population itself for phenotyping, Tim Millar and the team at Plant and Food Research, Christchurch, New Zealand for providing the Capture-seq data, the team at Plant and Food Research, Auckland, New Zealand for providing the *A. chinensis* whole genome sequence data, Lena Fraser and Kelvina Barrett-Manako for running the *H. lataniae* markers, Catherine McKenzie for generating predicted means from the binomial regression analysis, I thank Kate Stannard for showing me the *H. lataniae* bioassay and sharing germplasm phenotypes. Lena Fraser for helping identify markers for fine mapping, Kate Stannard for helping with *H. lataniae* colony rearing and phenotyping kiwifruit for resistance to *H. lataniae*, Elena Hilario and the team at Plant and Food Research, Auckland, New Zealand for providing the rtGBS data that was reanalysed for this work, I also appreciate the help and guidance from Cathy McKenna, Nicola Mauchline, Shirley Dobson and the Plant and Food Research, Te Puke, New Zealand entomology team who helped phenotype the 'Hort16A' x Male 7 population for *C. obliquana* resistance. Thanks also go out to Roy Storey who helped guide my bioinformatics journey. I would like to thank my wife Sarah for her support and understanding provided during this period. Finally, thanks to my co-supervisor Paul Datson for the endless advice, guidance and assistance.

iii. List of tables

iv. List of figures

v. List of abbreviations

Bp	Base pair
cM	CentiMorgan
BSA	Bulked segregant analysis
CRISPR	Clustered regularly interspaced short palindromic repeats
DAMP	Damage-associated molecular pattern
EST	Expressed sequence tag
ETI	Effector-triggered immunity
GBS	Genotyping by sequencing
HAMP	Herbivore-associated molecular pattern
Kb	One thousand base pairs
LRR	Leucine-rich repeat
MAMP	Microbe- associated molecular pattern
Mbp	One million base pairs
NBS	Nucleotide-binding site
PAM	Protospacer adjacent motif
PAMP	Pathogen-associated molecular pattern
PCA	Principal component analysis
Psa	*Pseudomonas syringae* pv*. actinidiae*- biovar 3
PTI	Pattern-triggered immunity
QTL	Quantitative Trait Locus
rtGBS	Random tagging genotyping by sequencing
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat
WGS	Whole genome sequence

1 Introduction to kiwifruit

Kiwifruit belong to the genus *Actinidia*, a group of 66 species of perennial climbing vines in the family Actinidiaceae (Ferguson & Huang, 2007; Huang et al., 2000). These are large generally deciduous perennial vines which require free-draining soil with adequate water and a humid environment for vigorous growth (Ferguson & Huang, 2007; Morgan et al., 1985). This genus is widespread in Asia, primarily occurring around the Yangzi River and southern China (Ferguson & Huang, 2007; Zeng et al., 2009). In these ideal environments kiwifruit quickly climb up standing trees to create a canopy. When growing kiwifruit commercially, support structures are required for new shoots to grow up. The new shoots are called canes and can grow as long as three meters in a season. It is typically these one-year-old canes that bear flowers and set fruit in *A. chinensis*, while *A. arguta* will flower and fruit on two-year-old canes in addition to the one-year-old canes (Cheng et al., 2006; Ferguson & Seal, 2008). Within the *Actinidia* genus there is significant variation in fruit flavour, size, flesh colour, and storage life, as well as vine architecture, environmental adaptation, productivity and pest and pathogen resistance. Wild species are generally dioecious requiring females to have a male in close proximity to set fruit (Ferguson & Seal, 2008). The basic chromosome number for *A. chinensis* is *x*=29 with a diploid number of 58 (McNeilage & Considine, 1989). The genus has inter- and intra-taxa variation in ploidy. *A. chinensis* var. *chinensis* vines are most often found as diploid or tetraploid while *A. chinensis* var. *deliciosa* vines are often hexaploid but in the wild show ploidy variation from diploid to dodecaploid (McNeilage & Considine, 1989; Testolin et al., 2016; Yan et al., 1997).

Most current kiwifruit cultivars have been commercialized from wild seedlings or after just a few generations of crosses from the wild (Ferguson & Huang, 2007). Two species from the Actinidiaceae family have been utilized to date for commercial production worldwide. These are the hairy, green-fleshed kiwifruit produced by *A. chinensis* var. *deliciosa* (A. Chev.). and the mostly hairless, yellow-fleshed kiwifruit *A. chinensis* Planch. var. *chinensis* (Mertten et al., 2012). There has been an increasing demand for the smaller, smooth-skinned "kiwiberry" produced by *A. arguta* (Ferguson & Huang, 2007; Ferguson & Seal, 2008).

In New Zealand, commercial production of kiwifruit started in 1930 with the first green-fleshed kiwifruit orchard established with the cultivar 'Hayward' (Ferguson & Huang, 2007). This cultivar remained dominant, eventually covering 90-95 % of internationally traded kiwifruit. In

1997 in response to decreasing returns, a grower owned company, Zespri International, was established. Zespri marketed first-grade 'Hayward' fruit under the trademarked name Zespri™ Green (Martin & Luxton, 2004). In 1999 Zespri patented a yellow-fleshed diploid cultivar from *A. chinensis* var. *chinensis*. This was given the PVR name 'Hort16A' (Lowe et al., 1999) and its fruit were marketed with the name Zespri™ GOLD Kiwifruit (Martin & Luxton, 2004). 'Hort16A' was subsequently devastated by the pathogen *Pseudomonas syringae* pv. *actinidiae* – biovar 3 (Psa) leading to a major drop in production (Dwiartama, 2017). Fortunately, a tolerant gold fleshed *A. chinensis* var. *chinensis* accession had been bred before the Psa outbreak to complement 'Hort16A' as an earlier, more mature variety. Because of Psa, this accession, licenced as a cultivar with the PVR name 'Zesy002' (fruit marketed as Zespri™ SunGold), went on to replace 'Hort16A' production in New Zealand (Dwiartama, 2017; Everett et al., 2011) after the Psa outbreak. Export of Zespri kiwifruit has been hugely advantageous to New Zealand by boosting the economy. Kiwifruit production made up 59 % of export earnings from the horticultural sector in New Zealand earning NZD 2.7 billion of export revenue by March 2022 (MPI, 2022).

1.1 Biotic threats to kiwifruit

Fortunately, the first few kiwifruit seeds imported from China in 1904 did not carry the many pests and pathogens associated with kiwifruit found in China. However, the generations of kiwifruit breeding completed in New Zealand from the narrow gene pool initially imported from China, in addition to an absence of their natural pest and pathogen pressures, may have acted as a genetic bottleneck for resistance genes affecting these pests and pathogens. This may have contributed to the susceptibility of the New Zealand bred cultivar 'Hort16A' to Psa, yet there are still countless other biotic threats to kiwifruit from overseas. These range from extremely costly, destructive, export-prohibiting incursions to others that may have a minimal effect on kiwifruit. To rank the risk of biotic threat incursions, in 2016, a New Zealand organisation, Kiwifruit Vine Health (KVH), put together a "most wanted" list of biotic threats from overseas that could affect kiwifruit in New Zealand. The biotic threats on the list had been ranked from highest risk (1) to lowest risk (11), taking into account the likelihood of entry to New Zealand, the potential for establishment on entry and the likely impact on the industry as a whole (KVH, 2016).

1. Fruit Fly (Mediterranean, Oriental, and Queensland)
2. Brown marmorated stink bug (*Halyomorpha halys*)
3. White peach scale (*Pseudaulacaspis pentagona*)
4. Brazilian wilt (*Ceratocystis fimbriata*)
5. *Pseudomonas syringae* non-New Zealand biovars
6. Spotted wing drosophila (*Drosophila suzukii*)
7. *Verticillium* wilt (*Verticillium albo-atrum*)
8. Invasive *Phytophthora* species
9. Summer canker (*Pectobacterium carotovorum actinidiae*)
10. Esca disease (*Fomitiporia mediterranea*)
11. Pelargonium zonate spot virus

In New Zealand, a unique lack of pests has allowed pest migration from other crops in numbers that might otherwise have not occurred (Lo & Blank, 1989). The two groups of pests which are most damaging to kiwifruit in New Zealand include armoured scale insects and the leaf roller complex. Armoured scale insects were thought to have been accidentally imported on ornamental plant species. Armoured scales include latania scale (*Hemiberlesia lataniae*) (Signoret), greedy scale (*Hemiberlesia rapax*) and oleander scale (*Aspidiotus nerii*). These are common throughout all countries that produce kiwifruit. The second group are the leafrollers which comprise a complex of several species from the family *Tortricidae* and two species from the *Oecophoridae* family. Leafrollers are moths which can cause physical damage to the fruit, leading to lower grading or even outright rejection of fruit shipments (Steven, 1999). There are a number of lesser damaging insects such as the passion-vine hopper (*Scolypopa australis*), greenhouse thrips (*Heliothrips haemorrhoidalis*) and the chorus cicada (*Amphipsalta zelandica*) but these pose a relatively minor impact to industry (Steven, 1999).

By far the most damaging pathogen to kiwifruit is Psa (Dwiartama, 2017). This pathogen is now present in all kiwifruit-producing countries including Italy, New Zealand, Chile, Greece, and China (Dwiartama, 2017). Psa destroys susceptible genotypes, including 'Hort16A' and causes issues such as bud browning or bacterial blight on *A. chinensis* var. *chinensis*. Other species of *Pseudomonas* such as *P. viridiflava* (Burkholder) also cause bud browning like Psa on some cultivars, but to a much lesser degree. Other pathogens in New Zealand include *Armillaria* root rot (*Armillaria novae-zelandiae*) (Stevenson), which is particularly deadly to vines that are in

proximity to infected stumps of felled trees (Homer, 1991), field rot (*Sclerotinia sclerotiorum*) – a fungal infection that causes blossom blight and fruit rot (Pennycook, 1985), and storage rot which can be caused by *Botrytis cinerea*. This affects kiwifruit through tissue damage during processing or packing (Pennycook, 1985). The biotic threats that are most damaging to an orchard depend on many variables. Listed below are some of the biotic threats commonly encountered by kiwifruit growers in New Zealand:

1. *Pseudomonas syringae* pv. *actinidiae* -biovar 3 (Psa).
2. *Pseudomonas* spp. (Bacterial blight)
3. Armoured scale insects: Latania scale (*H. lataniae*) greedy scale, (*H. rapax*) and oleander scale (*A. nerii*).
4. Leafroller complex: Brown headed leafroller (*Ctenopseustis obliquana*), black-lyre leafroller (*Cnephasia jactatana*) and light brown apple moth (*Epiphyas postvittana*).
5. *Armillaria novae-zelandiae*.
6. *Sclerotinia sclerotiorum.*
7. *Botrytis cinerea.*
8. Greenhouse thrips (*Heliothrips haemorrhoidalis*), New Zealand flower thrips (*Thrips obscuratus*), and (*Nesothrips propinquus*).
9. Passion vine hoppers (*Scolypopa australis*).
10. Chorus cicada (*Amphipsalta zelandica*).
11. Fullers rose weevil (*Asynonychus cervinus*).
12. Army worm (*Spodoptera frugiperda*)
13. *Phytophthora* spp.

1.2 Armoured scale insects

Armoured scale insects (Family: *Diaspididae*) are the largest and most specialized family of the recognized superfamily Coccoidea (Beardsley Jr & Gonzalez, 1975). Among the *Diaspididae* there are three types of armoured scale that infect kiwifruit in New Zealand. Two of these, lantana scale *H. lataniae* (Signoret) and greedy scale *H. rapax* (Comstock), commonly damage kiwifruit while *A. nerii* (Bouché) is rarely present in sufficient numbers to cause damage (Blank et al., 1992; Edwards et al., 2008). In 1988 *H. rapax* scale was predominantly found on the cultivar 'Hayward' with the exception of Northland and Gisborne where *H. lataniae* made up

50 % and 82 % of scale populations, respectively (Berry et al., 1989; Edwards et al., 2008). Since *H. lataniae* was detected in New Zealand in 1951, *H. lataniae* has spread and replaced *H. rapax* as the dominant species on 'Hayward' (Edwards et al., 2008; Morales, 1988). No recent survey of armoured scale insects has been published to keep up with the proliferation of new cultivars. Of great concern to New Zealand biosecurity is an exotic species of armoured scale called white peach scale (*Pseudaulacaspis pentagona*). This is not currently in New Zealand, but is spreading throughout the world and is highly destructive in the countries where it is present (Steven, 1999; Zhuang et al., 2016).

In New Zealand *H. lataniae* is the most damaging scale on kiwifruit (Hill et al., 2011). These insects are all parthenogenic females which reach reproductive age at 6-10 weeks old and produce 50-100 crawlers over their lifetime of up to 36 weeks (Hill et al., 2011). The first instar crawlers migrate by crawling a short distance from their hatch site or being blown to new sites in the wind (Blank et al., 1987; Blank et al., 1990). Reports of their settlement time differ. Berry et al. (1989) report settlement times from 2-5 days while Edwards et al. (2008) report settlement within 24 hours. In New Zealand, *H. lataniae* crawlers can spend up to 12 hours looking for a place to settle (Hill & Holmes, 2009; Hill et al., 2015). Humidity seems to be a significant factor in settlement time with humidity greater than 80 % allowing the crawlers to be active for up to 48 hours (Casey Flay - personal observation). Once the crawlers settle, they insert their stylet into the plant and secrete a hard waxy cap over themselves (Hill & Holmes, 2009; Hill et al., 2011; Lo & Blank, 1989). Once a first instar has inserted its stylet it remains permanently fixed in this position only able to rotate around the stylet (Hill & Holmes, 2009). The second and third instar make a progressively darker cap after each moult. The third instar continues to enlarge until it produces eggs (Blank et al., 1992). The nature of juvenile production is contentious, with Edwards et al. (2008) proposing that females are ovoviviparous producing motile juveniles and Berry et al. (1989) proposing that eggs are laid in the cap and later hatch.

1.2.1 Identifying *H. lataniae* scale

Armoured scale insects can be difficult to distinguish (Morales, 1988). The species found in New Zealand that attack kiwifruit are all <5 mm in diameter, circular, convex and pale greyish to brown. One feature that distinguishes *H. rapax* from *H. lataniae* and *A. nerii* scale is that the

scale appears 'tipped over' with dark exuviae toward the edge of the scale cap, as shown in Figure 2.3 (Morales, 1988). *A. nerii* scale can be distinguished from the other two by using a microscope to view whether the dorsal region contains three pairs of pygidial lobes, as shown in Figure 2.4. Their presence positively identifies the sample as *A. nerii* scale. If a single pair of lobes is present, the insect is either *H. rapax* or *H. lataniae* (Morales, 1988).

1.2.2 Quarantine issues

New Zealand kiwifruit production is totally dependent on the export market with only low volumes of low-grade fruit going to the local market (Steven, 1999). Armoured scale insects are a significant threat to kiwifruit export due to quarantine restrictions being in place for India, Korea, Thailand, Mauritius, Chile, Uruguay and Reunion Island. Quarantine restrictions result in the rejection of shipments if any armoured scale insects are detected. Restrictions on the number of scale insects are also in place for the lucrative Japanese markets which accept no more than 3 % of fruit with *H. lataniae* (Personal communication Steven Owen). Meeting these export criteria can be costly. In packhouses, scale-infested fruit need to be positively identified to determine their quarantine requirements. However, due to the difficulty involved in positively identifying first instar scale species, they need to be sent to expert labs for identification, significantly slowing the packing process even at low levels of scale contamination. With a heavy infestation of scale on fruit it can become uneconomic to pack, resulting in the whole crop being discarded (Blank et al., 1992).

1.2.3 Scale pest management

Conventional control of armoured scale on kiwifruit generally utilises mineral oil or organophosphates (Steven et al., 1995). Mineral oil sprays have traditionally been used against scale insects in organic production systems. However, this comes at a cost since the oil is phytotoxic to plants, causing leaf and fruit damage if applied incorrectly (Steven, 1999).

1.2.4 Breeding for resistance to *H. lataniae*

Kiwifruit, like other crops such as sugar cane, citrus and passion fruit, show variation in resistance to armoured scale insects (Dutta & Devaiah, 1992; Gea, 2011; Hare et al., 1990; Neethling & De Lange, 1995). Kiwifruit resistance has shown narrow sense heritability to be H^2 0.54 when assessed in the field (M. Hill et al., 2006). Moreover the *A. chinensis* var. *chinensis*

cultivar 'Hort16A' has been shown to be highly resistant to *H. lataniae* (Hill et al., 2011). This plant has been used as a parent in numerous *A. chinensis* var. *chinensis* populations. Furthermore, after many years of 'Hort16A' production in the field, this resistance has not been overcome by *H. lataniae* (Hill et al., 2015). However, due to the high susceptibility of 'Hort16A' to Psa, this cultivar has now been almost completely replaced by 'Zesy002'. Unfortunately, 'Zesy002' is also susceptible to *H. lataniae* (personal communication: Nicola Mauchline). This highlights the need for resistance markers to be developed to reduce the time and cost associated with phenotyping to assist kiwifruit breeding programmes.

1.3 Leaf roller complex

Leafrollers are a worldwide group of pests characterised by rolling the edge of a leaf to create a silk-sealed feeding chamber (Steven, 1991). However, the species present in each country varies (Steven, 1999). All species of leafroller in New Zealand are endemic, making them quarantine pests in all export markets (McKenna et al., 2009). Leafrollers were the first pests recognised to damage kiwifruit fruit in New Zealand. Of those that damage kiwifruit, there are two species - *Stathmopoda horticola* and *S. skelloni* - from the Oecophoridae family. Several species of leafroller from the Tortricidae family also damage kiwifruit. These include the black-lyre leafroller *Cnephasia jactatana* (Walker), the endemic brown-headed leafrollers *Ctenopseustis obliquana* (Walker) and *Ctenopseustis herana* (Felder and Rofenholfer), the green-headed leafrollers *Planotortrix octo* (Dugdale) and *P. excessana* (Walker), and the Australian light-brown apple moth *Epiphyas postvittana* (Walker) (McKenna et al., 2009; Steven, 1991). Two of the leafroller species most damaging to kiwifruit are the brown-headed leafrollers *C. obliquana* and *C. herana.* These are all small, <30 mm, grey-brown moths. The females of these species are indistinguishable apart from their sex pheromone profile. Males are almost indistinguishable with the exception of the ratio of the costal fold to the length of the forewing (Steven, 1991). The green-headed leafrollers can be distinguished through their beak-like labial palpi which are 2-4 times the horizontal diameter of the compound eye (Dugdale, 1990).

These species feed mostly on dead flower parts and leaves but one species has adapted to feeding on fruit (Steven, 1999). All leafroller species are destructive in their larval (caterpillar) stage. They utilise chewing mouthparts to consume the top layers of leaf and fruit material

(Figure 1.1). The difference in feeding mechanisms between armoured scale and leafrollers indicates that different resistance mechanisms could be activated to deter each pest.

Figure 1.1, *C. obliquana* adult (right), *C. obliquana* larvae (centre) with brown damage on fruit resulting from larvae feeding (right).

Biological control agents such as *Bacillus thuringiensis* (BT) have been shown to effectively and selectively control the leafroller larval stage (Steven, 1999). Historically, leafroller species have readily developed resistance to insecticides in many crops (Lo et al., 2000; Smirle et al., 2002). Possibly due to their higher use of BT, organic apple producers are also seeing resistance to BT control (Smirle et al., 2002). This highlights the need for resistant cultivars to be bred for resistance to leafrollers before the current control pesticide control measures become ineffective.

In New Zealand, the only kiwifruit cultivar with published resistance to leafrollers is 'Hort16A' (Gea, 2011) and the current gold fleshed cultivar'Zesy002' is susceptible to *C. obliquana*. Previous research has found a lack of host preference for egg laying in *C. obliquana* between susceptible and tolerant apples (Wearing et al., 2003). This indicates that larval growth on leaves may be the best bioassay for tolerance to *C. obliquana* (Wearing et al., 2003).

1.4 *Pseudomonas syringae* pv. *actinidiae* (Psa)

Psa is a canker-causing bacterium that is devastating to certain genotypes of kiwifruit. This pathogen affects *A. chinensis* var. *deliciosa* and particularly *A. chinensis* var. *chinensis*. It appears to have spread from Asia where up to four pathovars of Psa were present. Each of these had different pathogenesis and molecular characteristics but did not cause the pathogenesis observed in the virulent Psa – biovar 3. The biovar 3 appeared to have a different

effect on kiwifruit in Europe and New Zealand (Vanneste et al., 2013). Work looking into this biovar shows it was first isolated in Japan in 1984 after which it soon spread through Korea (Koh et al., 1994), Italy (Scortichini, 1994), Portugal (Balestra et al., 2010) and made it to New Zealand in 2010 (Everett et al., 2011). Plants infected with this bacterium often start oozing a clear, brown or white liquid from cankers in spring or autumn (Figure 1.2). Canes frequently die when they extend from cankers. In highly susceptible genotypes these symptoms occur on multiple canes, leading to whole vine death. On more tolerant genotypes, symptoms can involve flower bud browning, bud drop, flower wilting, and leaf spotting (Everett et al., 2011).

Figure 1.2 Symptoms of Psa infection. Photos retrieved with permission from Kiwifruit Vine Health.

Control of Psa on moderately tolerant *A. chinensis* var. *chinensis* plants such as the cultivar 'Zesy002' involves the removal of any wood showing ooze, cankers or cane death. Chemical control measures utilise copper sprays during the infectious spring period (Onorato et al., 2010). Some chemical control agents such as Actigard® (Syngenta) function as a salicylic acid analogue (Reglinski et al., 2013). This appears to close stomata which has been hypothesised to reduce Psa entry into the plant (Reglinski et al., 2013). A combination of these control measures mitigates the risk to cultivars such as 'Zesy002' and 'Hayward', however the use of copper sprays is limited due to their phytotoxicity and the use of Actigard® may have negative impacts on fruit yield (Reglinski et al., 2013).

Since Psa was detected in New Zealand there has been a significant effort to breed for resistance to this threat. However the heritability of resistance to Psa has been shown to be moderate (De Silva et al., 2014). The kiwifruit breeding programme has led to the recently commercialised *A. chinensis* var. *chinensis* cultivar, 'Zesy002' which is tolerant to Psa. As the kiwifruit breeding programme continues, the genotypes breeders work on have constant

passive selection for Psa tolerance. Furthermore, symptoms of Psa incursion such as cane dieback are negatively selected in breeding populations leading to more tolerant cultivars being produced (personal communication Luis Gea). While passive selection for Psa tolerance is advantageous for breeding populations, the high inoculum load is not ideal for germplasm collections. The high inoculum load kills plants that are tolerant to Psa at smaller inoculum concentrations. This is not good for the future of kiwifruit breeding as the germplasm collection is required as a diverse genetic pool to reduce the effects of genetic drift (Hartl et al., 1997). The alleles included in the germplasm's diverse genes may also aid in breeding resistance to future exotic threats to kiwifruit production.

1.5 Controlling pests and pathogens on kiwifruit

Conventional pest management mitigates the risks of pests and pathogens. However, insects can be difficult to control using conventional pesticides (Steven, 1999). In the past, eight or more insecticide treatments were applied per season to control insects (Blank et al., 1992). This led to significant pesticide residues in fruit which hindered exports (Steven et al., 1994). By 1997 the entire New Zealand kiwifruit industry aimed to move to integrated pest management under a Zespri scheme called KiwiGreen® to reduce pesticide residues within fruit (Steven et al., 1994). This strategy reduces insecticide applications by controlling pests only when they have reached critical population thresholds (Steven et al., 1994). In addition, a reduction in broad spectrum insecticides such as organophosphates was achieved with the scheme by utilising targeted pesticides such as BT and insect growth regulators (Steven, 1999). This approach had the added benefit of reducing mite outbreaks, as mite predators are not disrupted by broad spectrum organophosphates (Steven, 1999). Organic growers do not utilise synthetic sprays such as organophosphates. Consequently, the efficacy of BT sprays and insect growth regulators has been well utilised by these growers while sticking to conservative applications (Fairweather, 1999).

As Psa is a bacterial pathogen it has proven to be difficult to control. The commercial cultivar 'Hort16A' was so susceptible to this pathogen that in New Zealand it has been removed from mainstream production. Currently the most productive commercial cultivar in New Zealand, 'Zesy002', is tolerant to Psa. However, management protocols, including removal of infected plant material and applying copper sprays in autumn and during bud break in spring are

required to reduce Psa inoculum load (Onorato et al., 2010). The commonly grown green cultivar 'Hayward' does not usually exhibit cane death from Psa. Instead, the main issue experienced on 'Hayward' due to Psa is that infected flower buds turn brown and do not open. Along with other management procedures, the pathogen Psa is controlled with copper sprays which mitigates the risk in tolerant cultivars. However, Psa is building resistance to copper sprays by incorporating integrative conjugative elements and resistance plasmids (Colombi et al., 2017).

Reducing pesticide and copper application with integrated pest management programmes limits the negative impacts of conventional control. But this does not completely address the issues including the safety of applied chemicals, resistance of pests to applied chemicals, phytotoxicity, or the killing of off-target beneficial insects. In addition, there are a number of costs associated with conventional control, such as communication time, planning, pest and pathogen monitoring, applying chemicals, the chemical itself and the infrastructure for storage and application of chemicals.

1.6 Types of resistance

Plants have developed a variety of defence mechanisms to limit the impact of biotic threats on their fitness (Agnew et al., 2000; Hill et al., 2016; Tahir et al., 2019). These mechanisms can either resist the multiplication of a biotic threat (Clarke, 1986; Strauss & Agrawal, 1999), or tolerate its presence by limiting the effect of the biotic threat on the host plant's fitness regardless of the level of pest/pathogen multiplication (Little et al., 2010; Råberg, 2014). A resistance mechanism will reduce the risk of infection and/or the replication rate of the pathogen on the host whereas a tolerance mechanism will not (Pagán & García-Arenal, 2018). Within a resistance mechanism there can also be variance in the strength of the resistance exerted on the target biotic threat. For example, the kiwifruit cultivar 'Hort16A' was shown by Hill et al. (2011) to exert a strong resistance on *H. lataniae*. This resistance did not necessarily kill the insect but prevented its growth and reproduction (Hill et al., 2011). Resistance to Psa is a more complicated matter with different cultivars of kiwifruit showing a continuum of resistance from weak to strong (Cheng, 2014; Tahir et al., 2020; Tahir et al., 2019).

1.7 Mechanisms for resistance to pests and pathogens

Using plant-mediated resistance to biotic threats is the most cost-effective, practical, and environmentally safe way to manage pests and pathogens in plants (Cheng, 2014; Egesi et al., 2007; Kellerhals, 2009). Understanding the mechanisms that go into eliciting a plant's response to pest attack is often dependent on the plant, the environment and threat encountered. The biological driver for this resistance can be at any level in the plant's organisation from DNA, proteins, cells, tissues, organs, plant architecture or the way plants are arranged in the orchard (Jones & Dangl, 2006). When a plant's physical barriers and pre-formed defence structures and compounds have been overcome, a change must be detected at the cellular level before the plant can initiate a response. Depending on the type of pest or pathogen, either of two main pathways are initiated. The first detects particles from slowly evolving foreign material such as flagella (Jones & Dangl, 2006). These particles are known as pathogen-associated molecular patterns (PAMPs), microbe-associated molecular patterns (MAMPs) or herbivore-associated molecular patterns (HAMPSs) (Mithöfer & Boland, 2008; Oren et al., 2016). These PAMPs, MAMPs or HAMPs are perceived by pattern recognition receptors (PRRs) which produce a response referred to as pattern-triggered immunity, PAMP-triggered immunity or pathogen-triggered immunity (PTI). The definition of the PTI acronym changes depending on the author (Bigeard et al., 2015; Mott et al., 2017; Oren et al., 2016). Generally, activation of this immunity invokes a complex set of responses which resist pathogen attack in response to a set of molecules released upon pathogen attack (Oren et al., 2016). Plants can also detect damage-associated molecular patterns (DAMPs) which sense changes in the cell environment, endogenous peptides or degradation products caused by foreign bodies (Choi & Klessig, 2016). Another defence pathway, called effector-triggered immunity (ETI), recognizes molecules synthesised by pathogens and herbivores called effectors that counteract a plant's basal defences (Jones & Dangl, 2006). ETI is mostly triggered in the cytoplasm by directly detecting either virulence factors (effectors) or by detecting proteins altered by effector activity indirectly (Oren et al., 2016). Plants which cannot detect effectors become susceptible to attack from the pest or pathogen while those that contain resistance proteins to detect effectors can activate their ETI and resist attack (Jones & Dangl, 2006).

When PTI or ETI is triggered upon PAMP or effector recognition, there is a phytohormone response. This can involve the accumulation of defence phytohormones such as salicylic acid,

jasmonic acid, abscisic acid and ethylene (Grant & Jones, 2009; Selig et al., 2016). Depending on the threat responded to, different ratios of these hormones will be released. In particular, the salicylic acid pathway will be upregulated when biotrophic organisms that require live plant tissue, such as Psa, are incurred (Cao et al., 1994; Thomma et al., 1998; Wildermuth et al., 2001). On the other hand, DAMPs recognize wounding by insects or necrotrophic microbes, causing the plant to respond by upregulating jasmonic acid (Selig et al., 2016; Thomma et al., 1998). These major hormones can be antagonistic. This can be seen with an infection of *Arabidopsis thaliana* with the hemibiotrophic pathogen *P. syringae.* pv*. tomato*. This infection increases levels of salicylic acid and causes a reduction in resistance to the necrotrophic fungus *Alternaria brassicicola* on the same leaf due to a reduction in jasmonic acid (Spoel et al., 2007). Similar increases of salicylic acid have been observed in kiwifruit affected with Psa (Cellini et al., 2014) resulting in a higher rate of damage from leafroller (Hill, 2013). In contrast, resistance of kiwifruit to the necrotrophic fungus *Sclerotinia sclerotiorum* is increased when systemic acquired response is activated by salicylic acid (Reglinski et al., 2001). This indicates that the relationships between defence signalling pathways are not always antagonistic. Studies in rice have shown that a balance between salicylic acid and jasmonic acid is required to defend against the biotrophic bacterial pathogen *Xanthomas oryzae* pv*. oryzae* (Mur et al., 2006).

1.8 Resistance genes in the kiwifruit genome

The activation of resistance pathways requires the sensing of pest or pathogen presence with pattern recognition receptors or receptor like kinases in response to PAMPs, HAMPs, MAMPs, or DAMPs (Miller et al., 2017). The most cited of these contain nucleotide-binding sites (NBS), leucine rich repeats (LRRs), and a putative amino-terminal signalling domain (Belkhadir et al., 2004). A well known example of a DAMP pattern recognition receptor pair is the LRR-RLP gene known as VulNR. This PRR confers resistance to cowpea (*Vigna unguiculata*) by triggering the accumulation of reactive oxygen species and ethylene in response to inceptin that is found in the saliva from fall army worms (*Spodoptera frugiperda)* (Reymond, 2021). The resistance proteins can be broken down into four classes depending on the function of the protein: intracellular kinases, intracellular receptors, extracellular receptors and extracellular receptors coupled to kinases (Bent, 1996). The majority of resistance genes identified have either a Toll/Interleukin-1 receptor (TIR) or a coiled-coil at the N terminal end tailed by a nucleotide binding site (NBS) (Mondragón-Palomino et al., 2002). The C terminal ends of these proteins

often contain LRRs. The LRR domain has been shown to interact directly and indirectly with pathogen molecules making these regions particularly important for facilitating resistance (Mondragón-Palomino et al., 2002).

The genes identified in kiwifruit took a major leap forward in 2013 when Huang et al. (2013) published a draft sequence of the kiwifruit *A. chinensis* cultivar 'Hongyang'. They identified 96 NBS-LRR genes which recognize specific pathogen effectors. The low number of NBS-LRR genes relative to *Arabidopsis* (166) and tomato (251) led Huang et al. (2013) to imply that these genes are not under strong selection pressure. They postulate that this is because fewer pathogens have evolved to adapt to kiwifruit. They also identified 261 putative pattern-recognition receptor genes which encode receptor-like kinases with an LRR domain. This is more than in *Arabidopsis* (220) and tomato (236), indicating that PTI has an important role in kiwifruit (Huang et al., 2013). It is important to note that a significant number of genes in the 'Hongyang' sequence have not been properly identified due to it being a draft (Pilkington et al., 2018). Furthermore NBS-LRR genes are variable in sequence, making them difficult to identify. Thus, comparing NBS-LRR genes from a draft 'Hongyang' sequence to the *Arabidopsis* and tomato genomes can be prone to underestimation of the number of these genes present (Pilkington et al., 2018).

1.9 Kiwifruit compounds suspected to be involved in resistance

Oxalate is likely to be a major compound involved in protection against insects and other organisms which chew rather than pierce and suck plant material (Rassam & Laing, 2005). Note, however, that a distinction should be drawn between oxalic acid and oxalates as they likely have different modes of action. Oxalic acid for instance is soluble, scavenges calcium ions and may deplete herbivores by restricting nutrition (Konno et al., 2014). Oxalic acid can be further transformed into oxalate salts. These calcium oxalate salts aggregate to form raphide crystals (raphides). In kiwifruit these raphide crystals work to disrupt cell membranes to allow cysteine proteases to reach their targets and degrade proteins (Konno et al., 2014). While sucking insects may not be particularly deterred by oxalate, oxalic acid can cause its own negative effects. Oxalic acid and its oxalate salts (calcium, potassium or sodium) have been shown to inhibit sucking of the brown plant hopper *Nilaparvata lugens* in rice (Sōgawa, 1982). Sōgawa (1982) tested varieties with oxalic acid concentrations between 0.18 mg/g and 0.45

mg/g fresh tissue and found that varieties with over 0.34 mg/g were resistant to attack. In kiwifruit, oxalic acid has been shown to be an elicitor of post-harvest pathogen resistance when applied as a pre-harvest spray (Zhu et al., 2016). However, this may have been due to the increases in total phenolics, flavonoids, ascorbic acid and hydroxyproline-rich glycoproteins which are involved in cell wall strength (Deepak et al., 2010). Other compounds likely to be involved with plant resistance include a protease (Actinidin), which has been shown to act together with raphide crystals extracted from kiwifruit to reduce feeding of silk worms on the castor oil plant (Konno et al., 2014). Also, terpenes have been linked with plant pathogen control and control of Psa (Donati et al., 2014).

1.10 Molecular genetic approaches to characterising and understanding resistance

When DNA sequencing technology emerged in the 1970s it was slow and very expensive. To some it was inconceivable that a whole genome could be sequenced (Koshland, 1989). The human genome project was started in 1990 (Sawicki et al., 1993) and was completed in 2003 at a cost of 3 billion USD (Gibbs, 2020). The performance of DNA sequencing technologies has been rapidly accelerating ever since with significant reductions in time to completion, cost and error rates. In late 2006, sequencing a genome cost under USD 2000 per sample (Van Nimwegen et al., 2016) with genomes in 2022 available for less than USD 600 (Preston et al., 2021). In the 1970s, manual techniques such as Maxam-Gilbert sequencing and Sanger sequencing (electrophoresis gel) were utilised. These were modified to utilise features of capillary sequencing machines established in the 1990s, but were still expensive to run (Karger, 2011). These were further improved upon and next generation sequencing technologies of cyclic array, hybridisation based, nanopore and single molecule sequencing were quickly developed (Shendure & Ji, 2008).

Next generation sequencing technologies achieve high-throughput DNA sequencing by utilising anchor oligonucleotides on a flow cell complementary to adaptors added to samples (Elshire et al., 2011; Morozova & Marra, 2008). The sample oligonucleotides are hybridised to those on the flow cell. They undergo bridge amplification to generate clusters on the flow cell. After clusters of around 1000 have been cloned on the flow cell, it is flushed with nucleotides tagged with different coloured fluorescent dyes. When these dye-tagged nucleotides bind they

flash along with others in the cloned cluster on the flow cell. These flashes are recorded by a camera and indicate the nucleotides being bound in sequence. This occurs in a massively parallel process with a million clusters being recorded at once (Morozova & Marra, 2008). Where a reference genome is available, the data produced from sequencing are mapped against it to generate sets of overlapping reads. Where reads overlap each other, they build into a consensus region of DNA sequence known as a contig. The number of these overlapping reads gives the read depth that is critical for allele identification (Baker, 2012). Depending on the sequencing method, read depth is often more informative when above 10x coverage. For diploid species below 10x coverage, the alleles are difficult to distinguish from sequencing error and data artefacts (Baker, 2012). Where the read depth is sufficient, it is from these contigs that the DNA sequence inherited from each parent can be determined for that location on the chromosome. This enables studies such as molecular mapping, QTL analysis and fine mapping (Metzker, 2010).

1.10.1 Reduced representation sampling

Analysing the whole genome of populations with WGS of individuals is a prohibitively expensive route to gaining an understanding of a population's genetic structure. By reducing the number of genomic sequences analysed and spreading them over the genome, a significant reduction in the number of base pairs read can be achieved (Hilario et al., 2015). Multiple systems attempt to achieve this including reduced representation libraries (Davey et al., 2011), restriction associated DNA (Miller et al., 2007) and genotyping by sequencing (GBS) (Elshire et al., 2011). GBS was the first widely adopted method developed due to the barcoding system developed for samples. This enabled pooling of samples which dramatically reduced the cost per sample sequenced (Elshire et al., 2011). This system classically utilises the frequently cutting restriction enzyme *ApeKI* to digest DNA samples (Elshire et al., 2011). After digestion, fragments are put through PCR to increase the fragment pool. Adaptors and barcodes are hybridised to the fragments in the pool. Barcoded fragments of 100-400 bp can be put through next generation sequencing systems that read the sequence of the barcode and the sequence of the DNA it was ligated to (Elshire et al., 2011). The data produced can be aligned to a reference genome and SNPs identified from aligned tags.

Multiple refinements have been made around restriction enzyme cut frequency and decreasing the loci bias applied by restriction enzymes with random tagging genotyping by sequencing- rtGBS (Hilario et al., 2015). This system differs from GBS by utilising genomic DNA which incorporates a random hexamer linked to the common GBS oligonucleotide. The random-tagging process has the advantage of randomly landing on the genome allowing better sampling and representation of restriction sites compared to standard GBS (Hilario et al., 2015). It also allows use of restriction enzymes with a lower frequency of restriction sites enhancing the coverage of these sites or increasing the number of individuals assessed (Hilario et al., 2015).

GBS markers are effective for covering the whole genome but the genomic positions that get amplified do not remain consistent for each site. Newer technologies, such as Capture-Seq, captures sequence data for a subset of the genome by using custom RNA sequences. These RNA sequences, termed baits, are designed to bind and subsequently amplify captured DNA fragments from gene exomes with PCR (Mercer et al., 2012). PCR fragments are then sequenced to produce data that are fully compatible with whole-genome sequence data. This method is more repeatable than genotype by sequencing (GBS) techniques since Capture-Seq focuses on the genomic areas preselected by the custom RNA fragment baits. Targeting sites in this way generates an orthologous multi-locus array that can be used to identify nucleotide polymorphisms between individuals at read positions (Andermann et al., 2020). Because the Capture-Seq data captured are a repeatable subset of WGS data, wild collections can be sequenced and compared at amplified loci to identify the allelic diversity between distantly related individuals. This is not possible with GBS because there is poor repeatability of sampled markers (Hilario et al., 2015).

1.10.2 Allelic variants

Researching the genetic basis of resistance traits relies on the identification of genetic markers that correlate with specific resistance traits. In *Actinidia* spp. a number of genetic markers have been developed including amplified fragment length polymorphisms (Novo et al., 2010), restriction fragment length polymorphisms (Crowhurst et al., 1990), random amplified polymorphic DNAs (Huang et al., 2002), single nucleotide polymorphisms (SNPs) (Zhou et al.,

2011), and simple sequence repeats (SSRs) otherwise known as microsatellite markers (Korkovelos et al., 2008).

SSRs are tandem repeats of DNA of one to six base pairs which show high levels of polymorphism. These are scattered throughout the genome with a greater frequency in noncoding regions (Li et al., 2002). In plants, AT repeats are the most common followed by AG and TC repeats (Mohan et al., 1997). The variation of microsatellite markers is based on the variation in number of SSRs rather than point mutations as in SNPs. The variation in SSRs occurs about ten times more frequently than in SNPs due to slippage during replication or unequal crossing over, making them highly informative (Hancock, 1999). SSR markers are easily amplified by PCR and assessed by allele length with electrophoresis (Sánchez-Pérez et al., 2006). Plant and Food Research has developed a library of more than 132,000 expressed sequence tags (ESTs) (Crowhurst et al., 2008). SSR markers derived from these EST libraries were used in this study. EST derived markers are preferred in breeding programmes over markers from non-coding regions as they provide greater homology and marker-trait association across species (Tsang et al., 2006).

SNP markers are the most abundant and relatively stable genetic artefacts in plants and animals (Syvänen, 2001). They are single base variations in DNA, predominantly occurring in noncoding regions. However, when they do occur in coding regions, they can change the amino acid produced by the bases, which is termed a non-synonymous change. Conversely, a synonymous SNP does not affect the codon for the amino acid produced (Edwards et al., 2007; Syvänen, 2001). In kiwifruit, 32,764 bi-allelic variants derived from ESTs were found in coding regions which overlapped between *A. chinensis* var. *chinensis, A. chinensis* var. *deliciosa* and *A. eriantha* (Wang & Gleave, 2012). These variants are valuable when looking for markers robust enough to be utilised across species.

1.10.3 Finding markers associated with resistance

Markers are advantageous in population genetics due to the genomic information they provide. Finding markers involves identifying polymorphic sites in the genome and the corresponding frequency of individuals with traits linked to these polymorphic sites in the population (Syvänen, 2001). When these variant alleles are found to be associated with a phenotype, the marker can be used to infer that phenotype for genotypes carrying the marker

(Glazier et al., 2002). For a marker to be useful, it needs to be in, or in close proximity to, the gene of interest (Beavis, 1998). As the distance between a marker and a gene decreases, the likelihood of crossover recombination occurring between the marker and the gene also decreases, thereby increasing the marker's usefulness over generations (Glazier et al., 2002). Where available, these markers are mapped onto a reference genome. There is a manually annotated reference genome available for *A. chinensis* var. *chinensis* (genotype Red 5) (Huang et al., 2013; Pilkington et al., 2018). Once a marker has been identified, further work can be done by fine mapping the marker location with family-based studies, using linkage disequilibrium mapping of a cross, high-resolution crosses, congenic strains, and isogenic lines (Glazier et al., 2002). Of these approaches, family-based studies are the most practical and cost-effective in kiwifruit due to 19 pre-existing germplasm genomes and 40+ genomes available from their progeny (Wang & Gleave, 2012). Moreover, inbreeding depression is a significant detractor to kiwifruit storage and pest and pathogen resistance. This limits the use of individuals from inbred populations for further breeding (Gea, 2011).

1.10.4 Marker analysis

Marker analysis encompasses the identification of markers and the associated variation created during sexual reproduction. The main feature creating the observed recombination is crossover recombination (Bruce Alberts, 2002; Wijnker & de Jong, 2008). This occurs during meiosis when two homologous chromosomes pair to form a tetrad during prophase 1. Two arms of the tetrad form a join that can swap the ends of paired chromosomal arms making the join (Wijnker & de Jong, 2008). Typically a single crossover occurs per chromatid but occasionally another is formed at a distance from the first, making a double crossover (Foss et al., 1993). The recombined chromosomes are separated during meiotic gamete reduction to haploid germ cells (Wijnker & de Jong, 2008). These uniquely recombined germ cells within a parent are combined with the uniquely recombined germ cell of the opposite sex (Bruce Alberts, 2002). F_1 populations with resistant and susceptible parents with at least one of the parents being heterozygous were used in the work reported here. In this case, a resistant individual from the population may have recombination occurring close to the resistance gene. The resistant nature of the individual indicates that this individual inherited the resistance haplotype from its resistant parent. If markers span this region, the point of recombination can be found between a marker landing on the resistant parent haplotype and a marker landing

on the susceptible parent haplotype in that individual. Fine mapping can be done if sufficient individuals are found with recombination closer to the resistance gene. Markers are used at higher density to identify the closest point of recombination to the gene of interest.

To assess the statistical relationship between markers, a linkage or association map is created. This map can be made in statistical programs such as JoinMap 5® (Van Ooijen, 2004) where a logarithm of the odds (LOD) score is generated based on the ratio of linkage to no linkage. A LOD score of three indicates that linkage is 1000 times more likely than no linkage with the trait (Stam, 1993). However, since non-linkages are 50 times more common than linkages, the odds are 20:1 in favour of non-linkage (Nyholt, 2000). The distance between two markers is calculated by the recombination frequency (Rf). The LOD score will decrease with an increasing recombination value (Stam, 1993). The LOD score will also increase with a greater sample size. A high LOD score with a low Rf score shows close linkage between two markers and conversely a low LOD score with a high Rf score shows insignificant linkage between markers (Stam, 1993). The markers with high linkage scores can be grouped together into linkage groups which correspond to the plants chromosomes (Collard et al., 2005). The resulting diagram of each marker in order on each chromosome is the genetic map by which further quantitative trait analysis can associate with the phenotype (Stam, 1993).

1.10.5 QTL analysis

Quantitative trait locus (QTL) analysis is a statistical method linking phenotype with genotype data to identify regions of the genome associated with quantitative traits (Falconer et al., 1996). The power of this analysis lies in its ability to link complex phenotypes to chromosome regions to identify the location of genes, the gene number and their interaction with other genes (Miles & Wayne, 2008). The process of QTL analysis works by identifying the markers genetically linked to others which influence the resistance phenotype. If these markers segregate with resistance more frequently, a significant association score will be given, while unlinked markers will show insignificant association with resistance (Miles & Wayne, 2008). Several variables influence the statistical power of QTL analysis including family size, marker density, level of significance to detect QTL, genetic distances of QTL to markers, the contribution of segregating QTL to the observed phenotypic variance, and experimental error (Miles & Wayne, 2008). The family size required to get an accurate estimate of QTL effect is

variable. Generally, a population size of 200 will have a 90 % confidence level of detecting QTL which explain 10 % of phenotypic variance. Similarly, detecting QTL that explain 5 % of phenotypic variance requires 320 individuals (Hackett, 2002). If the family size is too small, QTL, especially those with low effect on phenotypic variance, may fail to be detected (Hackett, 2002). The marker density required is also variable: one cM (centiMorgan) indicates a 1 % chance that two genes are recombined through chromosomal crossover. Less than 20 cM between markers generally gives accurate QTL estimates (Hackett, 2002). Having sufficient marker density is especially important at low population sizes with integrative composite interval mapping to detect small effect QTL (H. Li et al., 2010). This threshold is easy to meet with modern sequencing technologies such as GBS which produces thousands of markers, but the marker density has a much lower effect on detection than a larger population size (Elshire et al., 2011; Hackett, 2002).

The level of significance required to detect QTL can also affect the result. To determine the threshold of significance, permutation tests are performed. These randomly rearrange markers while keeping their values constant, breaking marker trait associations (Doerge & Churchill, 1996). QTL analyses are run on greater than 500 of these rearrangements to determine the rate of false positive marker trait associations. Just above the highest level of false positives is then given as the threshold for detection of real QTL (Churchill & Doerge, 1994; Doerge & Churchill, 1996; Hackett, 2002). Experimental error can creep in at any stage to affect these results. The two main sources of error come from marker genotyping and phenotyping error. The error contributed by genotyping and missing data can influence the distance between markers and the order of markers in linkage maps (Hackett, 2002). Of greatest concern is phenotyping error. Inaccurate phenotyping increases the experimental variance leading to a decrease in the power of detection.

1.10.6 QTL from bulked segregant analysis

In addition to using family-based association mapping to detect QTL as described above, bulked segregant analysis (BSA) can be used to detect QTL for target traits in various species in highly diverse populations. In these diverse populations, QTL have been detected for dwarfing in watermelon (Dong et al., 2018), cotyledon colour in soybean (Song et al., 2017), cold tolerance in rice (Sun et al., 2018), resistance to ascochyta blight in chickpea (Deokar et al., 2019), kernel

length-width ratio in wheat (Xin et al., 2020), and vitamin C production in kiwifruit (McCallum et al., 2019). BSA identifies regions of the genome which display patterns consistent with selection between two bulks or pools of DNA (Magwene et al., 2011; Michelmore et al., 1991). The BSA method does this by estimating genotype frequencies for each bulk and analysing the difference in allele depth between them for each SNP site. In the method applied by Mansfeld and Grumet (2018), the shift in allele depth was measured against a 1-5 Mb sliding window. This accounted for local regions of the genome, with differences in allele frequency highlighted as QTL.

A typical BSA investigates loci that differ between sample bulks segregating for a trait of interest combining ideas from linkage mapping and genome wide association studies (GWAS) (Michelmore et al. 1991). Similar to classical linkage mapping, most BSA trials are designed using two parental strains with different phenotypes. The two strains are crossed to generate an F_1 population which is back-crossed or interbred for several generations to generate sufficient recombination to break up linkage from parents. Individuals from the last generation are selected to form two bulks that segregate for the phenotype of interest. The theory is that alleles affecting the target phenotype should have a significant difference between the two bulks while unselected alleles should remain in both bulks at similar frequencies (Michelmore et al., 1991; Shen & Messer, 2022). The DNA that contributes to each of the bulks is typically quantified for each individual, and an equivalent amount of DNA added to the bulk from each individual (Dong et al., 2018; Song et al., 2017; Wang et al., 2021). While this approach ensures a precise quantity of DNA added from each individual, tracking samples and extracting DNA from individuals is costly and time-consuming.

However, the QTL peaks from BSA can be skewed by the founder effect and genetic drift. The founder effect can bias results by returning a disproportionately high frequency of alleles from less allele diverse parents (Conolly et al., 2008; Excoffier et al., 2009; James, 1970). Results can be further biased by genetic bottlenecks causing genetic drift of alleles (Wisser et al., 2008). Methods for measuring the allele shift by selection alone need to take these factors into account, establishing a baseline of population variance (Narum & Hess, 2011; Wisser et al., 2008). Once QTL for traits have been established, further work can be done to identify genes underlying the QTL.

1.10.7 Identifying candidate resistance genes

The identification of the genes responsible for resistance is ideal to enable the development of markers which target that gene as closely as possible. This reduces the chance of recombination breaking the link between the gene and its marker and increases the utility of the marker in unrelated plants which may also carry the resistance gene. However, identifying genes for resistance within a region identified by markers or QTL can be complicated if the region is large. This is because many genes may be present within the boundaries of the region. Techniques such as fine mapping with molecular markers can reduce the size of the region of interest, but fine mapping often requires large populations to have sufficient recombination events to reduce the size of the region of interest to a few genes. This is evident when calculating the spread of the 33,044 gene loci throughout *A. chinensis*'*s* 554-Mb genome (Pilkington et al., 2018). If evenly split along the genome, a gene would be present every 16.85 Kb, requiring recombination events and viable markers either side of the gene of interest and less than 33.7 kb apart. However, resistance genes are often found in clusters, so this is rarely achieved with conventional markers. Rather a set of gene candidates is retrieved. Ranking candidate genes can be done by analysing whether individual genes functional regions such as NB LRR sites are intact or analysing whether candidates have similar genes functioning in other organisms with a BLAST analysis. Genes suspected to be involved with resistance within the region of interest could be confirmed with gene knockout analysis.

1.10.8 Testing resistance gene function

Identifying the function of a particular gene with certainty often requires the targeted deactivation of the gene to determine its function. These gene knockouts can be created through a number of methods including homologous recombination, zinc-finger nucleases (Gaj et al., 2013; Zhang et al., 2013), transcription-activator like effector nucleases (Chen & Gao, 2013; Gaj et al., 2013; Zhang et al., 2013) and CRISPR Cas9 – clustered regularly interspaced short palindromic repeats (Campenhout et al., 2019; Gaj et al., 2013). Genes can also be silenced with RNA interference (Gheysen & Vanholme, 2007; Small, 2007). Currently, the cheapest and easiest method for creating gene knockout plants is by using CRISPR Cas (Bortesi & Fischer, 2015; Campenhout et al., 2019). The CRISPR Cas system's simplicity comes from its two components: a small guide RNA to bind to the complementary sequence to be altered,

and the Cas9 protein which holds the guide and cleaves target DNA. Once DNA is cleaved it is repaired, leaving a modified target site (Campenhout et al., 2019). When this target site is critical to the function of a gene, such as a start codon or a functional protein domain, it can disrupt the resistance gene function (Bortesi & Fischer, 2015). Phenotyping transformed plants to see if they have lost their resistance will tell whether the knockout out gene is responsible for resistance (Feng et al., 2013; Saint Pierre et al., 2012).

1.11 Bioinformatics in breeding

The use of bioinformatics in breeding is becoming more widespread, and the quantity of data available for analysis from genomics is surpassing that which individual computers can process in a reasonable time. Instead, analyses of large genomic datasets, that require a significant portion of computer time, are often processed in high-performance computing clusters (Powers et al., 2022). Instructions for these computers are provided using programming languages such as R, C, C++, C#, Java, Perl and Python (Fourment & Gillings, 2008; Gentleman, 2008), and parallelised where possible to utilise the full potential of server computers (Shi & Wang, 2019). Workflows are often documented in notebooks such as Rmarkdown or JupyterHub to assist the process of writing and reading code for reproducible research (Tsang & Maciocci, 2020). Organising data bioinformatically requires operations that perform certain tasks, such as reading data into R (Francois, 2018), using string operations (Wickham, 2019b), accessing SQL databases (Falcon, 2020), and organising data in an array (Matt Dowle, 2020). The R/data.table package has many code optimisations that allow code to run quickly and be efficient with computer memory (Matt Dowle, 2020).

1.12 Breeding for resistance to pests and pathogens

Utilizing plants that are resistant to pests has been shown to be the most effective and environmentally friendly route for pest and pathogen control (Brown & Rant, 2013). Host plant resistance expression is often dependent on whether the resistance is qualitative or quantitative (Willocquet et al., 2017). Qualitative resistance is often conferred by the products of a single gene directly or indirectly by recognising the products of pathogen avirulence genes (Keller et al., 2007; Marone et al., 2013). Breeding for these single genes allows resistance to be bred into a population quickly and easily, enabling its widespread utilisation. This also has the advantage of conferring strong resistance on a genotype (Willocquet et al., 2017). Single

gene traits have a drawback in that the strong selection pressure exerted on the pest or pathogen they target can lead to development of tolerance to the resistance gene (McDonald & Linde, 2002; Zhan et al., 2014). On the other hand, quantitative resistance is often driven by several minor resistance genes. The resistance expressed by the pattern of minor genes is often not as strong as with single gene traits, but it also increases their durability (Willocquet et al., 2017). Unfortunately, breeding for quantitative resistance is more costly and time-consuming than breeding for single gene traits, as quantitative resistance genes have small effects from genes distributed throughout the genome (Poland et al., 2009). Molecular techniques show the most promise for pyramiding both qualitative and quantitative resistance into cultivars. This is due to the ability of markers to identify plants with multiple resistance genes which are difficult to separate phenotypically. These markers can then be utilised in conventional breeding programmes as selection criteria to incorporate resistance markers into a single cultivar (Keller et al., 2007).

1.13 Utilising ploidy in resistance breeding

Actinidia chinensis var. *chinensis* germplasm exists cytogenetically as both diploid and tetraploid, with 58 races of diploid and 116 races of tetraploid kiwifruit (Yan et al., 1997). Many agronomically important traits exist in diploid populations. However, when Psa spread through the kiwifruit germplasm collection in New Zealand, 50 % of diploid individuals survived compared to 80 % of tetraploid individuals (Datson et al., 2013). Despite the relative susceptibility of diploids to Psa, there are still many target traits that exist in these diploid populations. Breeding between diploids and tetraploids is not easy because most of the few viable offspring will be triploids with poor vigour and fertility (Hirsch et al., 2001). But crossing a tetraploid with a diploid individual can be achieved If the diploid produces unreduced (2*n*) gametes (Wu et al., 2014). If this is crossed with the reduced (2*n*) gametes of a tetraploid, the result is (4*x*) tetraploid offspring (Yan et al., 1997). These offspring vary their allele configuration depending on the stage at which the diploid parent did not split its chromosomes during meiosis. Normal meiotic division for a diploid goes through two consecutive chromosome doubling events. After that all sets divide to make haploid gametes. Plants that produce unreduced gametes miss chromosome separation at the first or second meiotic division stages (Wu et al., 2014). Developing markers for target traits in diploids is easier than developing markers in tetraploids due to gene dosage effects. When a diploid is crossed into a

tetraploid the markers stay informative, thereby creating a less complicated system for marker development.

1.14 Conclusions

The kiwifruit industry is highly productive in New Zealand, but this production is hampered by numerous biotic threats. Overcoming these pests and pathogens can be managed conventionally to mitigate the risks they pose to orchard production, but conventional management comes with its own drawbacks. Breeding for plant resistance will give numerous on- and off-orchard benefits including cost reductions of conventional control, reductions of pesticide use and enhancement of the clean green New Zealand image.

The practical integration of resistance genes from parental lines into elite kiwifruit breeding populations is missing from the literature and needs to be addressed. This can be approached in several ways, mostly including early screens of breeding material for resistance. The traditional method to do this included phenotyping each seedling for resistance. This has been shown for many crops including kiwifruit to be prohibitively expensive and time-consuming. Modern techniques include the use of molecular markers to provide a substitute for phenotyping.

Finding markers for plant resistance requires a detailed understanding of the plant genome. This understanding often requires genome sequencing but the most informative, whole genome sequencing, is prohibitively expensive to use on a population of plants. Reduced representation sequencing technologies, such as GBS or Capture-Seq, can sequence a population more cost effectively by taking a sample from each genome. The markers generated through this method can be utilised to generate a QTL map that can identify the location of resistance genes. But the resolution of GBS markers on the genome is often not sufficient to identify the genes conferring resistance. Additional markers can be developed closer to the target region to assist with breeding for that trait.

Once candidate genes are identified, disrupting their DNA sequence and assessing whether the resistance phenotype is lost can confirm the resistance gene function. For very complex populations where numerous individuals have died, a different approach is often needed. One of these involves the sequencing of whole families together and analysing the frequency of

alleles in the population. This BSA approach can identify alleles that remain in the population with higher frequency than expected. These alleles should be involved with resistance to the pathogen that killed the surviving individual's siblings.

Genomic markers can be very informative for breeding purposes. They can be quickly implemented to allow breeding for major and minor resistance genes. This can allow the pyramiding of genes for more durable resistance to a single threat and stacking of resistance to multiple threats. Breeding can further stack the resistances of kiwifruit to biotic threats. Ploidy can be increased through crossing to give some protection from Psa. This takes advantage of the ease of marker development in diploid populations and the advantages of polyploidy for pathogen tolerance.

Three key biotic threats were identified in kiwifruit that could be addressed utilising molecular tools. Of these, both *H. lataniae* and *C. obliquana* create major quarantine problems, packhouse issues and fruit damage if numbers are high. There are also problems controlling these pests with conventional methods due to the development of resistance to pesticides and spray residues being left on fruit. Psa was also targeted due to the widespread damage it causes to kiwifruit vines. This pathogen can currently be kept in check with copper sprays, but Psa is also developing resistance to copper.

The work in this thesis describes various kiwifruit populations and analytical approaches that were used to identify genomic loci for resistance to *H. lataniae*, *C. obliquana*, and Psa:

In Chapter 2, the aim was to characterise the genetic architecture for resistance to *H. lataniae*. To achieve this an association map was made for resistance to *H. lataniae* with microsatellite markers in a small bi-parental population, tested in a related population, and a QTL map made in a larger bi-parental population for association with resistance using GBS markers.

In Chapter 3 the aim was to use markers to breed *H. lataniae* resistance into elite cultivars. This aim was achieved by using the loci associated with resistance to *H. lataniae* to identify more possibly tolerant individuals with Capture-Seq data, and breed resistance from a diploid state into a tetraploid state using elite kiwifruit breeding lines.

In Chapter 4 the aim was to characterise the genetic architecture for resistance to *C. obliquana*. In this chapter loci associated with resistance to *C. obliquana* were sought using GBS markers.

Finally, Chapter 5 aimed to identify loci associated with resistance to Psa. This was achieved by identifying QTL for resistance to Psa among an incomplete factorial population that included individuals surviving Psa exposure, using a modified bulked segregant analysis.

The kiwifruit industry has a large impact on the New Zealand economy. Understanding the genomic loci that contribute resistance to the biotic threats that impact kiwifruit production and incorporating these loci into elite cultivars will have a positive impact on kiwifruit production and the New Zealand economy as a whole.

2 Genomic architecture of resistance to latania scale (*H. lataniae*) in kiwifruit (*A. chinensis* var. *chinensis*)

2.1 Abstract

Latania scale (*Hemiberlesia lataniae* Signoret) is an armoured scale insect known to cause damage to kiwifruit plants and fruit, which ultimately reduces crop values and creates post-harvest export and quarantine issues. Resistance to *H. lataniae* does exist in some commercial cultivars of kiwifruit. However, some of the commercial cultivars bred in New Zealand have not inherited alleles for resistance to *H. lataniae* carried by their progenitors. To elucidate the architecture of resistance in the parents, these experiments analysed the inheritance of resistance to *H. lataniae* from families related to commercial cultivars. The first experiments identified a genomic region of interest for resistance to *H. lataniae* in small test and validation populations. A larger population was then used to map quantitative trait loci (QTL) for resistance to confirm the region of interest as the sole locus contributing to resistance. The larger population was also fine mapped to narrow down the region of interest to an area of low recombination of 4.1 to 12.7 Mb on chromosome 10. The number of single nucleotide polymorphisms (SNPs) and insertion or deletion polymorphisms (indels) within the whole genome sequence data of the resistant pedigree was then analysed to rank candidate genes for resistance to *H. lataniae*. The markers developed for the fine map, which target the resistance loci found in parents, will reduce the amount of costly and time-consuming phenotyping required for Latania scale resistance while breeding resistance to *H. lataniae* into new kiwifruit cultivars.

2.2 Introduction

Kiwifruit are of significant commercial value to New Zealand, with exports alone earning NZD 2.7 billion by March 2022, which was a 2.4 % increase compared to 2021 (MPI, 2022). However, despite an ever-increasing overseas demand for kiwifruit and the associated financial gain, producing fruit of a quality suitable for export can be a difficult task. One of the obstacles to producing export-grade kiwifruit is the occurrence of pests in the field, particularly the latania scale insect (*Hemiberlesia lataniae* Signoret) (M. G. Hill et al., 2006). But a greater financial impact comes from exports to the most lucrative kiwifruit markets of Japan and China which list *H. lataniae* as quarantine restriction. Fruit exported to markets with quarantine restrictions need to be inspected individually to be free of *H. lataniae* at a high labour and packing throughput cost, making it one of the most costly insects to manage in kiwifruit production in New Zealand. *H. lataniae* infests New Zealand's two main commercially produced cultivars: the green-fleshed 'Hayward' and the gold-fleshed 'Zesy002', commonly known as Gold3, with fruit marketed as Zespri™ SunGold. *H. lataniae* attacks the leaves, trunk, canes, and fruit of both cultivars (M. G. Hill et al., 2006). In New Zealand, *H. lataniae* adults are all parthenogenic females that remain permanently fixed in the location established by mobile first instar crawlers (Hill et al., 2011). The crawlers hatch under their parent's armoured scale or cap and migrate by crawling a short distance or being blown to new sites in the wind (Blank et al., 1987; Blank et al., 1990). Once the crawlers settle, they insert their stylet into the plant and secrete a hard waxy cap to cover themselves (Hill & Holmes, 2009; Hill et al., 2011; Lo & Blank, 1989). Conventional control with pest monitoring and pesticide application has proven somewhat effective, but a cultivar which is resistant to *H. lataniae* could save kiwifruit growers in New Zealand in excess of NZD 77 million per annum (personal communication Cathy McKenna).

Historically, New Zealand kiwifruit growers applied at least eight insecticide treatments per season to control *H. lataniae* on kiwifruit (Blank et al., 1992). As a result of such heavy treatment loads, export rates became hindered as the fruit being produced was found consistently to carry pesticide residues at levels undesirable in foreign markets (Steven et al., 1994). However, by 1997 the New Zealand kiwifruit industry had collectively agreed to move to integrated pest management under a scheme jointly developed by Plant and Food Research and Zespri®, named KiwiGreen®, which aimed to reduce pesticide residues within the fruit (Steven et al., 1994). The KiwiGreen® system is still successfully in use in conventional orchards

today, but some scale can remain on fruit post-harvest. This causes major delays in packing fruit as export regulations require that all insects be accurately identified before they are sold to overseas markets. The most sustainable, economically practical and environmentally friendly approach to preventing *H. lataniae* attack has been identified as prevention rather than cure; that is, to breed kiwifruit cultivars resistant to insect attack, thereby eliminating the issues associated with insect presence. This approach is particularly suited to organic growers who do not have the same treatment options available to them as conventional growers.

Like in other species, there is variation in resistance to armoured scale insects within kiwifruit species (Dutta & Devaiah, 1992; Gea, 2011; Hare et al., 1990; Neethling & De Lange, 1995). For an *Actinidia* population at Plant and Food Research, Te Puke, New Zealand, M. G. Hill et al. (2006) showed a "moderate heritability" of a hypersensitive response to *H. lataniae* in the cultivar 'Hort16A'. This hypersensitive response occurs when scale insects settle on the bark of canes of 'Hort16A' and a wound periderm is formed under each scale insect. The wound periderm develops smaller cells with thicker cell walls and increased phenols than cane bark with no *H. lataniae* present. This was proposed to physically restrict the extension of the insect's stylet into their feeding sites of parenchyma tissue and result in the resistance of 'Hort16A' to *H. lataniae* (Hill et al., 2011; Hill et al., 2015; Wu et al., 2015). However, due to the high susceptibility of 'Hort16A' to the bacterial pathogen *Pseudomonas syringae* pv. *actinidiae* (Psa), the 'Hort16A' cultivar has now been almost completely replaced by the cultivar 'Zesy002' in New Zealand, which, despite including 'Hort16A' in its pedigree, is susceptible to *H. lataniae* (personal communication: Nicola Mauchline).

Little is known about the molecular architecture governing the gene(s) responsible for the hypersensitive resistance response observed by Hill et al. (2011). But the regulation of RNA hypersensitive response has been studied in response to *H. lataniae* attack (Hill et al., 2015; Hill et al., 2016). However, the identification of hypersensitive response strongly suggests a gene-for-gene relationship between the plant's dominant resistance gene and the insect's dominant avirulence gene, as this is consistent across most species of higher plants that have been studied (Balint-Kurti, 2019). The phenotypic segregation pattern of descendants from the ancestor Female 1 (Figure 2.1) is also consistent with the hypothesis that there is a single resistance gene.

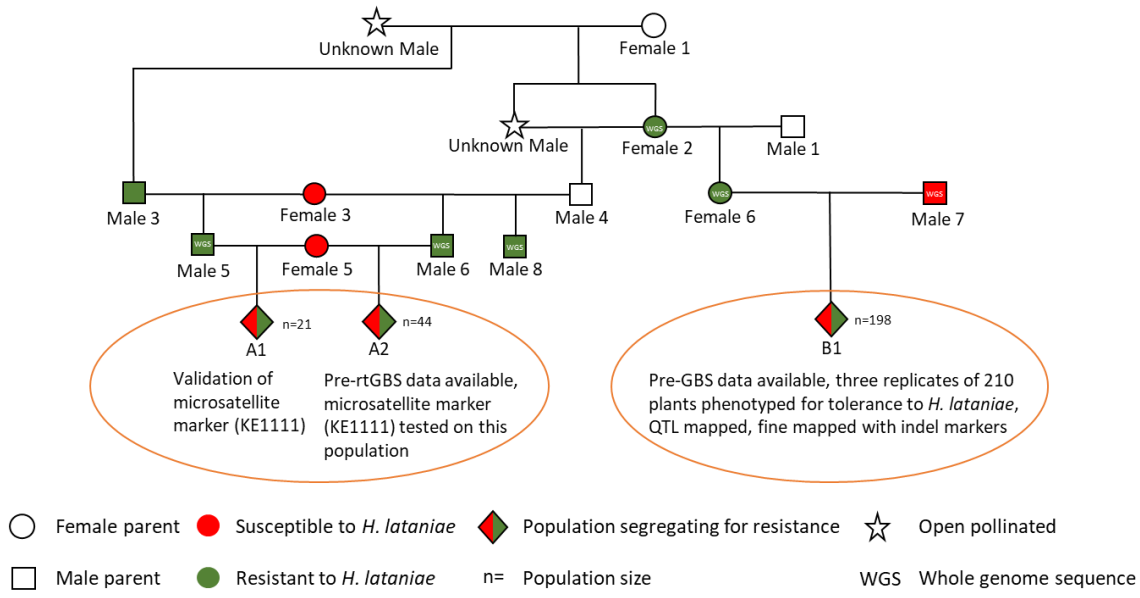


Figure 2.1. Pedigree of *A. chinensis* var. *chinensis* families A1, A2 and B1 that have a 1:1 segregation ratio of susceptible and resistant individuals. The circle nodes indicate females, with the square nodes as males. Red indicates susceptible with blue resistant to *H. lataniae*. Diamond nodes indicate families with the number of genotypes in the family. The A1, A2 and B1 families having 21, 44, and 198 individuals respectively. A star node in the pedigree tree indicates that pollen was from an open cross, thus the male parent is unknown. Inheritance of resistance to *H. lataniae* in the A1, A2 and B1 families was likely from one of the parents from the first recorded cross of an unknown Male to Female 1. From this cross, the resistance allele passed onto the resistant Female 2 and Male 3. Female 2 likely passed the resistance allele onto Female 6 and Male 4, then Male 4 passed resistance to Male 6. Male 3 also passed on its resistance to Male 5.

The two main commercial kiwifruit cultivars grown in New Zealand are susceptible to *H. lataniae*. This is due to *H. lataniae* resistance being viewed as a secondary trait for breeding purposes behind traits such as fruit flavour, yield, and storage. As the costs associated with phenotyping are high, the parents of early populations were not phenotyped for resistance to *H. lataniae*. One of the ways to reduce the time and labour involved with identifying resistant kiwifruit individuals is to develop molecular markers associated with resistance. Markers for

resistance to *H. lataniae* could be implemented in breeding programmes alongside sex markers which are already routinely used in marker-assisted selection to increase the number of fruit-bearing females in the field. This will enable the cost-effective integration of resistance to *H. lataniae* in commercial kiwifruit breeding programmes.

The work presented here aimed to characterize the genetic architecture of resistance to *H. lataniae* and develop insertion-deletion (InDel) markers associated with that resistance. This was achieved by first using a small F_1 family of kiwifruit segregating for resistance to identify regions of interest and markers associated with that region. A larger F_1 segregating family was then used for QTL mapping using GBS data and fine mapped to further narrow the region of interest. Populations and data sources are summarised in Table 2.1.

Table 2.1. *A. chinensis* var. *chinensis* family population sizes of the A1, A2 and B1 families. The results of phenotyping and genotyping for resistance to *H. lataniae* show an equal segregation of resistant and susceptible F_1 seedlings from all three families. Phenotype and genotype data sources are identified for each family.

Number of plants	Individuals resistant to *H. lataniae*	Individuals susceptible to *H. lataniae*	Percentage of resistant individuals	Replicates	Phenotype data source	Source of genotyping data
21	12	9	57	1	Hill G.	Microsatelite - Flay, C.
42	24	18	57	1	Hill G.	rtGBS - PFR + Microsatelite - Flay, C.
202	96	106	48	3	C. Flay	GBS - Tahir, J. (PFR)

2.3 Methods

2.3.1 Identifying a genomic region of interest for *H. lataniae* resistance in the A2 family

The identification of loci for scale resistance began by analysing the available genotypic and phenotypic data of 20 diploid *A. chinensis* F_1 individuals from the A2 family (Figure 2.1) provided from a study done by Cheng et al. (2015). This population was utilised as it had been phenotyped for *H. lataniae* scale resistance based on a cut cane bioassay (described later in this chapter) available from Garry Hill (Plant and Food Research, Te Puke, New Zealand, unpublished data). There were also rtGBS genotype data available for the same 20 individuals, as the A2 population had been previously used for a trial of genomic selection (Cheng et al., 2015). The A2 family also showed an even 1:1 segregation of resistant and susceptible phenotypes in the field grown F_1 individuals. The A2 family was derived from a cross between a resistant parent, Male 6, and a susceptible parent, Female 5 (Figure 2.1). The A2 family was especially useful to identify the genomic region of interest for resistance to *H. lataniae* because a related but independent validation family with phenotypes and extracted DNA also was available. The validation family was an important check to see whether markers trained on the A2 family associated with resistance to *H. lataniae* in another family inheriting the same alleles. The related A1 validation family shared the susceptible mother (Female 5) with the A2 family, but was crossed with a different resistant father, Male 5. Male 5 is related to Male 6 by sharing a mother (Female 3) and having a related father (Figure 2.1). The A1 family also had an even 1:1 segregation of resistant and susceptible phenotypes, indicating that a single locus underlies resistance.

The A2 family had been rtGBS genotyped following the method of Hilario et al. (2015) and data consisting of 39,000 SNPs for each of the 10 resistant and 10 susceptible plants were made available for this study. Data in the form of International Union of Pure and Applied Chemistry (IUPAC) degenerate base symbols for each SNP were imported into R statistical software for analysis. SNP calls were recoded to numeric for data checking, filtered to remove homozygous SNPs, and filtered to remove SNPs with bases the same as those of the resistant male parent (Cornish-Bowden, 1985). A Kruskal-Wallis test (Kruskal & Wallis, 1952) was run for each SNP site against each individual's phenotype call in the remaining 3,800 SNPs heterozygous in the resistant parent. The Kruskal-Wallis test evaluated the null hypothesis for association between

phenotype and genotype at each SNP site. The resulting p-values were corrected for multiple tests using "fdr" (Benjamini & Hochberg, 1995) to reduce the chance of including false positive results. SNPs with significant phenotypic association ($p \leq 0.05$) were also checked manually to confirm they segregated as expected, i.e., with a SNP inherited from the resistant male parent being present in all the resistant individuals and absent in the susceptible individuals. It was expected that the Kruskal-Wallis test would identify a genomic region of interest for scale resistance as a series of linked markers from one region in the genome. Markers for this region could then be identified and tested in a validation population to see if they continued to show association with scale resistance.

2.3.2 Association of marker KE1111 with phenotype in the related A2 and A1 families

To validate the association of rtGBS markers on chromosome 10 with resistance to *H. lataniae* in the A2 family, molecular markers were sought. Microsatellite markers were used for this association because of their homology between related families (Huang et al., 1998), the ability to be run on poor quality DNA (Vieira et al., 2016), and because a set of EST-derived microsatellite markers was available at Plant and Food Research (Fraser et al., 2004). The Plant and Food Research database was searched for a GA repeat microsatellite marker for the A2 family that landed within the 2.83 – 13.89 Mb region on chromosome 10 identified from the rtGBS data (Table 2.1). A marker named Ke1111 was found to land at 9.75 Mb on chromosome 10 in the Red5 *A. chinensis* reference genome (Pilkington et al., 2018). To assess whether KE1111 was informative, whole genome sequence data of the susceptible mother (Female 5) and resistant fathers of the A2 and A1 families (Male 6 and Male 5 respectively) were loaded as separate tracks into the integrative genome viewer- IGV (Robinson et al., 2011). The KE1111 primer binding sites were added as tracks in IGV and the sequence between primers was checked for variants affecting the amplified fragment length within the SSR. Primer sites were also checked for SNPs to see whether primers would bind unimpeded. The parental genome sequence indicated that Ke1111 was fully informative in the A2 family and in the related A1 validation family with an AB x AC pattern of inheritance. The A allele from both fathers was thought to be associated with resistance.

Initially, 43 individuals from the A2 family and their parents were genotyped with Ke1111 to test the marker for association with resistance. The 21 related individuals from the A1

validation family were then genotyped with the same marker to verify the association of the A allele from the resistant male parent (Male 5) with the resistance phenotype.

Genotyping was performed by PCR for the A2 and A1 families in 96-well plates. Plates contained 15 ul of reagents with 1.2 mM dNTPs, 10x PCR buffer, 50 mM $MgCl_2$, Invitrogen platinum *Taq* DNA polymerase, and 0.34 pmol of forward (TGGTCTGCCGGGTTATCTGTTA), and reverse (TCCCTGACTACATGAACTTC) primers for Ke1111. PCR conditions included heating the lid to 105 °C, denaturation at 94 °C for 3 min, 35 cycles at 30 s at 94 °C, then 30 s at 58 °C, and 1 min at 72 °C. Genotyping with marker Ke1111 was performed on an Agilent 5300 Fragment Analyzer™ (Agilent Technologies Inc. Santa Clara, California, USA) using a "dsDNA 910 Reagent Kit, 35 bp - 500 bp".

To test for association between genotype and phenotype in the A2 and A1 families, the resistant phenotype was used as a marker in JoinMap®5 and a regression analysis completed to get a LOD score for the association. The *H. lataniae* phenotype can be used as a marker in this family due to its binomial expression and because the fathers of both families are heterozygous for resistance, while the mother in common with both families was homozygous susceptible. These data were imported into JoinMap®5 (Van Ooijen, 2004) using the p allele in a nn x np pattern of resistance inheritance. The data were checked with a Kruskal-Wallis test in R with p-values corrected using a false discovery rate algorythm.

2.3.3 QTL mapping the larger related B1 family

To build on the results from the A1 and A2 families and to further validate the association between the 2.83 – 13.89 Mb region on chromosome 10 in the A2 family, a larger family related to the resistant parents of the A1 and A2 families was sought to map the QTL associated with resistance. A family of 202 previously GBS-genotyped individuals with three clonal replicates for each individual was made available from a population developed by Tahir et al. (2017). This F_1 cross was phenotyped for *H. lataniae* resistance or susceptibility using the live plant bioassay described later in this chapter. All experimental replicates were combined after removing individuals that had inconsistent phenotype calls across replicates. This population was from the *H. lataniae* resistant individual Female 6 as mother and the susceptible Male 7 as father (Figure 2.1). The plants used in this experiment were up to 1.5 m high and grown in 7.5-L pots in a glasshouse. The family was GBS-genotyped and cloned by tissue culture of seedling leaves

to make three replicates per genotype at Plant and Food Research by Tahir et al. (2017). Each clonal replicate was grown in a different environment. One was grown in an open-walled glasshouse at Plant and Food Research, Palmerston North New Zealand with an average temperature over the 10 weeks assessment of 22.5 ± 4 °C. The other two replicates were managed by the author at Plant and Food Research, Te Puke, New Zealand. One of these replicates was grown in a fan-vented glasshouse at 20 ± 7 °C, while the other was in a shaded passively vented plastic house at 19 ± 11 °C. These locations were selected to give a measure of the environmental influence on the resistant phenotype.

The GBS genotype data were collected by Tahir et al. (2017) and sequenced at the Australian Genome Research Facility (AGRF) on an Illumina HiSeq with 100 bp paired end reads. Data were initially made available from a linkage map in JoinMap format and phased with allele codes <llxlm>. However, the bioinformatics pipeline used by Tahir et al. (2017) to make the linkage map used an older bioinformatics pipeline and an old reference genome described in detail in Chapter 4. To ensure the accuracy of the linkage map and resulting QTL, the raw FASTA files from AGRF were re-aligned by replacing the older pipeline which used TASSELL5 for alignment with Red5:PS1.68.5 with a BWA-MEM/Samtools pipeline aligned to the more recent Red5:PS1.69.0 reference genome available from NCBI (Pilkington et al., 2018) by the author. The newly aligned genotype data from the B1 family were used to re-make linkage maps in JoinMap® 4. Three plants were removed from the dataset while making the linkage map because about a third of the markers coded as null. Around a quarter of the markers were identical to others and were removed before calculating the linkage map. The linkage map data were exported as .map and LOC files for QTL mapping in MapQTL 5 (Van Ooijen, 2004). A cross pollinators (CP) population type was chosen to match the heterogeneously heterozygous and homozygous diploid parents of the studied population. The interval mapping method was used to detect QTL between markers and phenotype (Van Ooijen, 2004). The genome-wide significance level thresholds for interval mapping were estimated to be 57.9 when using a permutation test with 10000 replicates. Any LOD scores under 57.9 were considered insignificant. Because the phenotype data for this trait were binomial, the Kruskal-Wallis method was also applied in MapQTL 5 (Van Ooijen, 2004).

2.3.4 Fine mapping the B1 family

To lower the risk of the association between resistance gene/s and marker/s being broken through recombination, selected markers should be as close to the resistance genes as possible. To achieve this, fine mapping was completed with markers covering the region of interest. To start the fine map, eight microsatellite markers were designed by Lena Fraser (Plant and Food Research) to target the area under the QTL peak found in the B1 family. These marker sites were checked, and additional marker sites identified, by loading the *A. chinensis* reference genome 'Red 5' (Pilkington et al., 2018) in an integrated genome viewer - IGV (Robinson et al., 2011). Along with the reference genome, binary alignment map (BAM) files of the B1 family parents and individuals resistant to *H. lataniae*, i.e., Female 2, Male 5, and Male 6, (Figure 2.1), were loaded into IGV as tracks to visually identify informative alleles and check marker quality. These resistant individuals were used as checks because of their likely inheritance of the resistance haplotype from a common ancestor as determined from their pedigree, covered further in Chapter 3, and to ensure the resistance alleles and null alleles selected were informative.

InDels over 4 bp in length that were in the resistant parent (Female 6), but not in the susceptible parent (Male 7) were checked to see if they were consistent with reads from the other resistant parents. Twenty-two InDel markers were selected to span the QTL region with a higher density of 25 markers under the QTL peak at 8-13 Mb, compared to eight markers outside this region. Primers were then generated to target informative InDels in Geneious with Tm ± 1.5 °C, and product lengths between 136 and 461 bp. Primers were generated using the *A. chinensis* reference genome scaffold 'Red5' (Pilkington et al., 2018). Primer binding sites were checked in IGV to see if they had any SNPs or InDels in Female 6, Male 7, Female 2, Male 5, or Male 6. Twenty-four markers were selected after filtering on the above criteria, landing at 4.1, 4.8, 5.1, 5.6, 6.4, 7.3, 8.1, 8.6, 8.8, 9.1, 9.9, 10.3, 10.9, 11.1, 11.5, 11.6, 11.7, 11.8, 11.9, 12.2, 12.4, 12.7, 12.9, and 13.8 Mb along chromosome 10. For fine mapping, PCRs were set up as per Schuelke (2000). This involved tagging the forward primers with a dye labelled-M13(-21) primer tail at the 5' end. PCR reagents and conditions were as presented in Schuelke (2000). For a final 10-µl reaction volume in a Perkin-Elmer Standard PCR buffer, eight pmol of reverse primer and dye labeled-M13(-21) primer, two pmol of the forward primer, 0.2 mM dNTPs, 50–100 ng template DNA, and 1 U Platinum Taq DNA polymerase were used. Conditions of the

PCR amplification: 94 °C (5 min), then 30 cycles at 94 °C (30 s) / 56 °C (45 s) / 72 °C (45 s), 8 cycles 94 °C (30 s) / 53 °C (45 s) / 72 °C (45 s), final extension at 72 °C for 10 min. To assess fragment lengths, 1 µl of the PCR product was added to 22 µl formamide and 0.5 µl ROX standard (PerkinElmer) and run on an Applied Biosystems 3130 Genetic Analyzer.

DNA for fine mapping of the B1 family was extracted from glasshouse-grown plants. Young actively growing leaves <40 mm in length were collected and ground with a pestle and mortar in liquid nitrogen. DNA was extracted using a Qiagen DNeasy® Plant Mini Kit, following the manufacturer's protocol (QIAGEN, 2016). DNA quality was analysed on a NanoDrop® 1000 spectrophotometer (Thermo Fisher Scientific) for the 260/280 absorbance ratio and DNA quantity. Samples with A260/A280 ratios outside the 1.6 - 2.2 range were re-extracted (El Bali et al., 2014). As the DNA yield from glasshouse-grown plants was low, a urea solution was applied to the growing media to increase DNA yield. The solution consisted of 1 g/L of urea pellets with about 100 ml of solution applied to each plant one week before DNA extraction.

The exon sequence of annotated genes within the 5.17 and 10.91 Mb boundaries identified by the fine map was pasted into BioView (Crowhurst et al., 2006) to BLAST the sequence against the "Genbank Non-Human genomic" database containing non-human genomic sequences. Genes with sequences like those known to be associated with pest and pathogen resistance mechanisms in other plant species were recorded as candidate genes for resistance to *H. lataniae*.

2.3.5 Assessing the similarity of genomic SNPs between resistant ancestors of Female 2

The database of kiwifruit whole genome sequence data was checked for individuals which share their pedigree with Female 1, the suspected ancestor of resistance to *H. lataniae* (Figure 2.1). The R packages RLinuxModules (Grønvold, 2021), Data.table (Matt Dowle, 2020), Glue (Hester, 2020), Here (Müller, 2017), Tidyverse (Wickham et al., 2019), Stringr (Wickham, 2019b), Magrittr (Wickham, 2019a), RSQLite (Falcon, 2020), Readr (Francois, 2018), and Knitr (Xie, 2020) were used to organise the data and assess the coefficient of relationship between individuals. Five individuals (Female 2, Female 6, Male 5, Male 6, and Male 8) were identified that were resistant with whole genome sequences aligned to the reference genome Red5:PS1.69.0 (Pilkington et al., 2018). Variant call files were made from binary alignment files using bcftools mpileup (Li et al., 2009) for chromosome 10. Input options for the variant caller

included disabling probabilistic realignment for the computation of base alignment quality to reduce false SNP alignments, and the minimum base quality was set to 10. Mpileup format tags included total depth and the depth for each allele and information tags included the depth for each allele. Once the variant call files were made, bcftools norm was used to retrieve variant sites. Data were piped to bcftools query to pull sample names, genomic position, reference allele and depth, each alternate allele and their depths, and InDel sites. Data were merged by position into a single data table with each sample's data in a new column. Data were cleaned up by removing genomic positions with no reads from any of the five contributors. Alternative alleles with a single read at any site were replaced with a depth of 0 and allele nucleotide call removed. A new column for each sample was made with a "1" for each site containing a SNP. Another column was added to count the number of alleles in each sample which were the same as those in the other five samples, with a maximum of 5 indicating SNPs were present at each site. A new column was made and a "1" placed in each allele site that had all five individuals with the same nucleotide SNP or inDel as Female 1, and a "0" if they had dissimilar nucleotides. Another new column was added, with a "1" at each site, for SNPs that were not in Female 1. These sites were considered mutations which may negatively affect the likelihood of a particular gene being involved with resistance to *H. lataniae*. The average percentage of identical and dissimilar SNP and InDel sites was calculated as a global average for chromosome 10.

To get an estimate of whether the frequency of identical and dissimilar SNPs and InDels would vary across chromosome 10, a .bed file was made to include nucleotides from the start to the end of chromosome 10 in 0.5 Mb bins. IRanges (Lawrence et al., 2013) was used to align the 5' and 3' boundaries of bins from the .bed file to the data table and assign the bin names to a new column. The number of nucleotides available for each bin, i.e., those which were present among contigs from the five resistant individuals, and the number of sites which had SNPs and InDels in all five individuals identical to Female 1, were assessed as a percentage of analysed nucleotides. The number of sites which had SNPs and inDel sites dissimilar to those of Female 2 was also assessed as a percentage of analysed sites and presented with GGplot2. To analyse and rank candidate genes for resistance to *H. lataniae*, the same database as above was used, but the .bin" file was altered. The ".bin" file for gene regions included gene names with gene regions, and 1,000 bases were added to either end of the gene region to include promotor

regions. To further analyse candidate genes for resistance to *H. lataniae* within the 5.74 Mb region of interest, the same database as above was used, but the genomic regions in the .bed file contained only gene exons with no 1,000 bp boundaries.

To rank candidate genes, the percentage of all polymorphisms was averaged within genes. This may artificially inflate the percentage of polymorphisms in the genes selected for exon analysis, but it adds a better ranking of within-gene polymorphisms, with less emphasis on the surrounding 1,000 bp gene region.

2.3.6 Phenotyping for *H. lataniae* resistance in the A1 and A2 families using the cut cane method

Two methods were used to establish the *H. lataniae* resistance phenotype of kiwifruit in this study: a cut cane method used in families A1 and A2 and a live plant method used for the B1 family. The cut cane method was developed by M. G. Hill et al. (2006) to phenotype the A1 and A2 families genotyped for this study. This method used two 1.2-m sections of one-year-old kiwifruit cane, at least 15 mm thick. Cane wood was taken from field grown plants in the Te Puke Plant and Food Research Centre orchard. These were placed into long plastic bags, sealed and moved to cool storage until the bioassay could be completed. The order of genotypes was randomised. Three 400-mm lengths of the healthiest canes greater than 15 mm in diameter were selected for the bioassay, with the aim of making three replicates for each genotype. Lateral buds were removed as close to the cane as possible. Wax was applied to cut areas, excluding the basal end of the cut section, and wool was wound around a 150-200-mm section 20 mm down from the apical end of the cane. The canes were then labelled and placed horizontally side by side on a tray. *H. lataniae* crawlers were brushed onto the canes from a population of adult scale (Figure 2.2). Canes were carefully lifted, and the basal end placed in containers with 20 mm of water. Canes were kept upright standing them through wire mesh over the top of the bin holding the containers (Figure 2.2). After ten days the wool was removed, and the scale insects were left to develop for ten weeks in a sealed room at 20.5 ± 1.3 °C and 51.5 ± 7 % relative humidity. After the scale development period they were measured using the same sizing and calling parameters as in the potted plant bioassay.

Figure 2.2. Woody cane bioassay to phenotype *A. chinensis* var. *chinensis* for *H. lataniae* resistance. *H. lataniae* crawlers are applied from a butternut squash (*Cucurbita moschata*) (left). A 400-mm length of winter-dormant kiwifruit cane is placed upright in water with wool wrapped around each cane. A section of cane resistant to *H. lataniae* is shown (centre), with no scale caps above 0.4 mm in diameter, and a cane susceptible to *H. lataniae* (right) with scale caps greater than 1.4 mm in diameter.

2.3.7 Phenotyping for *H. lataniae* resistance in the larger B1 family using the live plant method

The cut cane method of phenotyping was effective for small numbers of individuals when sufficient *H. lataniae* scale crawlers were available. However, the large size of the B2 family and the small canes of some plants made them unsuitable for the cut cane bioassay. A higher throughput method was required to complete the phenotyping of all replicates before winter, when the growth of *H. lataniae* slows. Therefore, the live plant method described in Hill et al. (2015) was modified to increase the bioassay throughput for the three replicates of 198 individuals in the B1 family. The adapted assay method used live plants with crawlers applied to plants in the glasshouse instead of using the cut-cane method in the lab. For the bioassay, plants were prepared by wrapping wool around a 20-cm section of kiwifruit cane around 20 cm above the soil level. Scale crawlers were brushed off the rearing colony into a sample container coated with Fluon® paint to prevent crawler escape. The sample container was transported to the glasshouse and around 150 crawlers transferred onto plant canes with a paintbrush between the times of 9 am and 11 am. A damp tissue was wrapped around sites where crawlers were applied to increase the humidity of the settlement site. Tissue was kept wet for three days by sprinkling water onto tissues twice a day. Crawler establishment was

checked three days after application by inspecting the cane for the small white caps made by the sessile crawlers. Canes with fewer than 10 visible caps had crawlers reapplied. At the completion of the 10-week period from scale crawler application, the size of each scale cap was recorded by removing each scale cap and assessing whether the scale was alive. Six size classes were created after measuring the diameter of live and dead scale: <0.4 mm, 0.4-0.6 mm, 0.6-0.8 mm, 0.8-1.0 mm, 1.0-1.2 mm and >1.2 mm. Canes with greater than five live scale in the largest three categories were recorded as 'susceptible'. Canes with greater than ten scale in the dead category with no scale in the three largest live scale categories were recorded as 'resistant'. This corresponded to the resistance present in the mother Female 6, restricting the growth of the scale to around 0.2 mm.

Three locations were selected to bioassay each clonal replicate of 198 plants. Two of the sites were at Plant and Food Research, Te Puke, and the third was a glasshouse in Palmerston North. Two resistant Female 6 plants and two susceptible plants from the cultivar 'Hayward' were included as positive and susceptible measures of resistance, respectively. The first replicate had *H. lataniae* applied to plants in Te Puke in a fan-vented and mist-humidified glasshouse at 20.8 °C plus/minus standard deviation (±) 3.9 °C, and relative humidity (RH) of 72.7 % ± 12.2 % RH, over six days from 27 August 2018. The second replicate also had crawlers applied to plants in Te Puke in a separate passively vented plastic house with environmental conditions of 17.7 °C ± 4.1 °C and 74 ± 13.2 % RH over three days from 13 September 2018. The third replicate was phenotyped in a passively vented glasshouse at Massey University Palmerston North, under conditions of 22.5 °C ± 3.8 °C and 61.8 % ± 14.6 % RH.

2.3.8 B1 family plant maintenance

The large population of 202 individuals from the B1 family came from the *H. lataniae*-resistant female parent (Female 6) and the susceptible male parent (Male 7) (Figure 2.1). The plants making up the population were up to 1.5 m high and grown in 2.8-L planter bags in a plastic house in Palmerston North, New Zealand. The population was cloned to make three replicates, each transported to a different location for different environmental effects. All plant replicates were re-potted into 4.5-L pots to enhance plant health, with potting mix and Osmocote-Exact-Protect slow-release fertiliser tablets both provided by Daltons (Mount Manganui, New Zealand). Yates Thrive all-purpose liquid plant food, supplied by Mitre 10 (Te Puke, New

Zealand), was applied at 3-month intervals at the recommended dose. Two replicates were relocated to the Plant and Food Research, Te Puke site. Both replicates at Te Puke were maintained in the environments in which they were phenotyped (described above). The third replicate remaining in Palmerston North was maintained in a passively vented plastic house at Plant and Food Research, Palmerston North, New Zealand.

2.3.9 *H. lataniae* scale colony rearing

To allow cut cane and live plant bioassays to be implemented when required, a colony of *H. lataniae* was maintained in a sealed room at 20 ± 2 °C and 55 ± 5 % RH at the Te Puke lab at Plant and Food Research, New Zealand. In New Zealand, *H. lataniae* are all parthenogenic females which reach a reproductive age at 6-10 weeks and produce 50-100 crawlers over their lifetime of up to 36 weeks (Hill et al., 2011). Crawlers search for a feeding site for up to 48 h (personal observation) before inserting their proboscis and becoming sessile for the remainder of their lives (Hill et al., 2015). The colony of *H. lataniae* was grown on fruit of butternut squash (*Cucurbita moschata*). Crawlers from squash infested with adult *H. lataniae* were brushed onto clean undamaged squash every second day. Population brushing was done between 8 am and 11.00 am to coincide with the peak in crawler emergence from under adult scale caps (personal observation, Kate Stannard). It was critical to clean the squash thoroughly, as population contaminants of *Aspidiotus nerii* or *Hemiberlesia rapax* can be introduced which are difficult to distinguish from *H. lataniae*, except when mounted under a microscope, as seen in Figure 2.3 and Figure 2.4 (Morales, 1988). Net pantyhose was applied around uninfested squash before brushing to assist scale establishment. The pantyhose was removed once squash had been populated with >1,000 juvenile scale or after 14 days. Squash were kept in bins with a mesh bottom and mesh vented lid. The lid was sealed closed with double-sided tape between the bin's lid and base to prevent predation from earwigs (*Forficula auricularia*) and parasitoid wasps of *Encarsia* spp. (Figure 2.5).

Figure 2.3. Adult armoured scale on kiwifruit with differing cap characteristics (Henderson, 2011). Visual identification of adult species uses exuviae placement in the cap. *H. lataniae* caps have a characteristic brown colour and centralised exuviae in their cap. *H. rapax* have dark exuviae in the side of the cap and *A. nerii* have a clearer soft cap in comparison to both *H. lataniae* and *H. rapax*.

Figure 2.4. Ventral view of the dorsal region of adult armoured scale (Henderson, 2011). The number of pygidial lobes in light brown at the bottom of the picture are indicative of whether the insect in question is *H. lataniae* or *A. nerii*. Individuals with three pairs of pygidial lobes are *A. nerii*.

Figure 2.5. A *H. lataniae* scale colony was maintained in fourteen fish bins with mesh-covered vents in the lid and double-sided tape around the lid to seal them. In response to an outbreak of the parasitoid *Encarsia citrina* (imaged right) in the scale population, two air-supplied, gasket-sealed, mesh-vented bins with raised floors (left) were added with sticky traps above the vents to catch *E. citrina* present in the room.

2.4 Results

2.4.1 Identifying a genomic region of interest for *H. lataniae* resistance

To determine the genomic architecture of kiwifruit resistance to *H. lataniae*, pre-available rtGBS genotypes and *H. lataniae* resistance phenotype data for the A2 family were interrogated. The 32,352 SNPs from the re-analysed rtGBS data included 3,800 SNPs heterozygous in parent samples which were sequenced with the family. These SNPs were analysed using a Kruskal-Wallis test which found 20 SNPs associated with the phenotype ($p = 0.001$) (Table 2.2). All significant associations had a heterozygous male parent, which confirms the male as the contributor of resistance to *H. lataniae* in this family.

Table 2.2. Alleles from the A2 family of *A. chinensis* var. *chinensis* with significant association (p = 0.001) between genotype and phenotype. The A2 family was an F_1 family from a cross between resistant Male 6 and susceptible Female 5 which was sequenced with rtGBS. Individuals resistant to *H. lataniae* are highlighted with blue fill, and susceptible individuals are highlighted by salmon fill. Nucleotides not aligned with other individuals of their phenotype are highlighted in grey. Nucleotides are named according to their IPUAC codes. Alignment position of variants by chromosome is shown for both the earlier PS1.68.5 reference genome and the more recent PS1.69.0 (Red 5) reference genome. Significance of association was analysed with a Kruskal-Wallis test and an false discovery rate test used to account for multiple tests at each position. Variant sites likely due to error are shown in italic text.

Chromosome	PS1.69.0 Position (bp)	Male 6 (Parent)	T1_A2	T2_A2	T3_A2	T4_A2	T5_A2	T6_A2	T7_A2	T8_A2	T9_A2	Female 5 (Parent)	S1_A2	S2_A2	S3_A2	S4_A2	S5_A2	S6_A2	S7_A2	S8_A2	S9_A2	KWT (p < 0.001)	FDR
chr9	*16,471,354*	*R*	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*R*	*R*	*R*	*R*	*R*	*R*	*R*	*R*	*R*	*0.00004*	*0.0204*
chr9	*16,471,380*	*K*	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*K*	*K*	*K*	*K*	*K*	*K*	*K*	*K*	*K*	*0.00004*	*0.0204*
chr9	*16,471,381*	*Y*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*	*0.00004*	*0.0204*
chr9	*16,471,385*	*R*	*R*	*R*	*R*	*R*	*R*	*R*	*R*	*R*	*R*	*S*	*S*	*S*	*S*	*S*	*S*	*S*	*S*	*S*	*S*	*0.00004*	*0.0204*
chr10	2,979,307	Y	C	Y	C	C	C	C	C	C	C	C	Y	Y	Y	Y	Y	Y	Y	Y	Y	0.00023	0.0204
chr10	3,230,961	R	A	R	A	A	A	A	A	A	A	A	R	R	R	R	R	R	R	R	R	0.00023	0.0739
chr10	6,250,277	Y	Y	Y	Y	C	Y	N	Y	Y	Y	C	C	C	C	C	C	C	C	C	C	0.00030	0.0923
chr10	6,714,699	R	R	R	R	G	R	R	R	R	R	G	G	G	G	G	G	G	G	G	G	0.00023	0.0204
chr10	9,308,847	R	R	R	R	R	R	R	R	R	R	G	G	G	G	G	G	G	G	G	G	0.00004	0.0204
chr10	9,308,870	S	S	S	S	S	S	S	S	S	S	C	C	C	C	C	C	C	C	C	C	0.00004	0.0204
chr10	11,453,261	R	A	A	A	A	A	A	A	A	A	A	R	R	R	R	R	R	R	R	R	0.00004	0.0204
chr10	11,453,291	Y	C	C	C	C	C	C	C	C	C	C	Y	Y	Y	Y	Y	Y	Y	Y	Y	0.00004	0.0204
chr10	11,746,281	K	K	K	K	K	K	K	K	K	K	G	G	G	G	G	G	G	G	G	G	0.00004	0.0204
chr10	12,453,780	Y	T	T	T	T	T	T	T	T	T	T	Y	Y	Y	Y	Y	Y	Y	Y	Y	0.00004	0.0204
chr10	15,611,339	W	T	W	T	T	T	T	T	T	T	T	W	W	W	W	W	W	W	W	W	0.00023	0.0739
chr10	15,611,362	R	R	G	R	R	R	R	R	R	R	G	G	G	G	G	G	G	G	G	G	0.00023	0.0739
chr10	19,026,423	M	C	C	C	C	C	C	C	C	C	C	M	M	M	M	C	M	M	M	M	0.00023	0.0739
chr10	19,202,224	Y	C	C	C	C	C	C	C	C	C	C	Y	Y	Y	Y	Y	Y	Y	Y	Y	0.00004	0.0204
chr12	*10,871,123*	*S*	*S*	*C*	*S*	*S*	*S*	*S*	*S*	*S*	*S*	*C*	*C*	*C*	*C*	*C*	*C*	*C*	*C*	*C*	*C*	*0.00023*	*0.0739*
chr12	*10,871,140*	*Y*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*	*Y*	*0.00004*	*0.0204*
	phenotype	Tol.	Tol.	Tol.	Tol.	Tol.	Tol.	Tol.	Tol.	Tol.	Tol.	Sus.	Sus.	Sus.	Sus.	Sus.	Sus.	Sus.	Sus.	Sus.	Sus.		

The greatest number of markers was centred on chromosome 10, with fourteen markers showing a significant association with the phenotype (Table 2.2). Six markers were aligned to chromosomes 9 and 12, but as the marker positions on chromosomes 9 and 12 were less than 31 bp apart, it was assumed that each came from a single sequence read and were misaligned in reference genomes PS1.68.5 and PS1.69.0. Marker positions on chromosome 10 were found to segregate with resistance in an 11.1-Mb region between 2.8 and 13.9 Mb when aligned to PS1.68.5. There was some difference in results from reference genome alignment with the newer genome version PS1.69.0, showing markers segregating over a 16.6-Mb region between 2.9 Mb and 19.2 Mb on chromosome 10 (Table 2.2). The significance of markers adjusted using the FDR method for repeated sampling at p =0.05 was not considered for the region of interest as it provided a statistically insignificant result for the PS1.68.5 reference genome markers between 2,831,848 and 4,139,753, which was likely caused by a recombination event in the T2_A2 individual between the markers at 3,230,961 and 6,250,277 bp. T2_A2 was thought to be recombinant as a single arm of the chromosome was affected and the markers matched those of the susceptible individual markers from the Female 5 parent instead of those of other resistant individuals.

2.4.2 Genotyping the A2 and A1 families

To confirm the association of the 11.1-Mb region identified in the A2 family, a microsatellite marker was identified within the 2.8 to 13.9 Mb region identified from the rtGBS data above. This marker was tested on pre-extracted DNA from the A2 family. The identified marker, Ke1111, amplified a 451-462 bp region at 10.7 Mb on chromosome 10. Parental WGS data, viewed in IGV, predicted KE1111 to have an AB x AC pattern of inheritance, with A*C coming from the father (Male 6) and AB from the mother (Female 5). The PCR fragments were run on a genetic analyser to assess fragment length. Three peaks were found at 451, 462, and 447 bp. The peak at 451 amplified both the father's A* and the mother's A allele, and the peaks at 462 and 447 corresponded to the mother's B and the father's C alleles, respectively. The allele segregation ratio of individuals from the A2 family was approximately ¼ A*A, ¼ A*B, ¼ AC, and ¼ BC. Despite the inheritance of resistance being linked to the father's A* allele being confounded with the female's A allele, individuals with the C allele from Male 5 are unlikely to also carry the father's A* allele. Therefore, resistance to *H. lataniae* inherited from Male 5's A* allele can be determined by inference. Forty-two individuals from the A2 family were

genotyped with the Ke1111 marker. Peaks from the fragment run were observed in 24 genotypes, while two genotypes did not associate with the phenotype (Table 2.3).

The remaining 18 individuals did not have reliable fragment peaks, possibly due to the age of DNA, DNA quality, or a bad run through the fragment analyser. To assess whether an association of genotype with phenotype was present, genotypes were loaded into JoinMap®. Phenotypes were loaded as a marker with a <nnxnp> pattern for susceptible and resistant plants respectively (Table 2.3).

The resulting recombination frequency (RF) of 0.0833 and a LOD score of 3.78 indicates that the odds are greater than 1,000:1 that these loci are linked to *H. lataniae* resistance. These data were also analysed with a Kruskal-Wallis test in R resulting in a chi-squared value of 17.051 and a p-value of <0.001. The significant linkage with resistance indicates that the distance between the marker and the gene region responsible for resistance in this family is small enough to have a low frequency of disruption by recombination.

Table 2.3. KE1111 marker genotypes in the A2 and A1 populations of *A. chinensis* var. *chinensis*. Analysis of phenotype-genotype association was run in JoinMap as a cross pollinator (CP) population type. To run the expected phenotype as a marker for association in JoinMap®, genotype codes of loci segregation for the AB x AC pattern of inheritance were recoded as AA = ee, AB = ef, AC = eg, and BC = fg. Phenotypes observed as resistant were entered as np, and susceptible as nn. The observed phenotype highlighted in blue indicate the two samples that did not match the expected phenotype.

Family	Number of samples	Observed Phenotype	Expected Phenotype	Alleles (ABxAC)	Expected phenotype <efxeg>	Observed phenotype <nnxnp>
A2	8	Resistant	Resistant	AA	ee	np
A2	5	Resistant	Resistant	AB	ef	np
A2	7	Susceptible	Susceptible	AC	eg	nn
A2	2	Susceptible	Susceptible	BC	fg	nn
A2	1	Resistant	Susceptible	BC	fg	np
A2	1	Susceptible	Resistant	AB	ef	nn
A1	9	Resistant	Resistant	AA	ee	np
A1	1	Resistant	Resistant	AB	ef	np
A1	2	Susceptible	Susceptible	AC	eg	nn
A1	5	Susceptible	Susceptible	BC	fg	nn

To validate the Ke1111 marker trained on the A2 family, it was further tested on the A1 family. The A1 family had 21 plants from which DNA could be sourced. The marker site Ke1111 was amplified in the same manner as in the A2 family, producing observable allele peaks for 17 of the 21 plants genotyped. It was found that this family showed no differences between the observed phenotype and the marker predicted phenotype i.e., every plant carrying the resistant allele was resistant. The allele and phenotype data were loaded into JoinMap®, and a regression analysis was run, which resulted in an RF of 0.0 and a LOD score of 3.7. These data were also analysed with a Kruskal-Wallis test in R resulting in a chi-squared value of 9.0264 and a p-value of 0.002661. This indicated that in this family the marker was near the locus responsible for resistance. Both the marker training family (A2) and the validation family (A1) showed consensus with the scale resistance phenotype. The general agreement between alleles and the phenotype indicates that the marker is close to the gene of interest.

2.4.3 Phenotyping the B1 family

To develop better markers, the larger B1 family, an F_1 family from a cross between resistant mother Female 6 and susceptible father (Male 7), was phenotyped by the author, resulting in the identification of 96 resistant and 106 susceptible plants. This was consistent with the 50/50 segregation pattern seen in the A2 and A1 families, indicating a single dominant gene for resistance. The resistance phenotype observed in the resistant parent Female 6 and the resistant individuals from the B1 family was the same as that observed by Hill et al. (2011). In their study, Hill et al. (2011) observed that resistant individuals allow crawlers to spin a white cap over themselves in the first few days after settling, but they do not develop a cap over 0.8 mm in diameter and do not develop into mature adults. Susceptible plants allowed unrestricted growth of *H. lataniae* to adulthood over the assay duration, with scale caps over 0.8 mm in diameter. This was consistent with the phenotype observed in the susceptible control 'Hayward'. Plant genotypes were removed from analysis if *H. lataniae* growth was inconsistent between the three replicates of the study. Plants were rechecked for live scale before this was done to eliminate recording error. Despite the measures to reduce error, an incorrect phenotype call was detected by an unusual pattern of fine mapping markers. The correction of this error enhanced the QTL map and fine map.

2.4.4 QTL mapping the B1 family

To develop more precise markers for resistance, the distance between the gene and the marker needs to be as small as possible to avoid recombination between the gene and the marker. QTL analysis of the larger B1 family was proposed to provide more recombinant individuals to narrow down this association. QTL analysis of the provided GBS data showed association with the resistance phenotype on a 16-Mb section of chromosome 10 with a LOD score of over 6.4 at the tails of the peak, and 82.9 at the peak (Figure 2.6). This peak explained 87.2 % of the variance with phenotype. The QTL location aligns with previous SNP data obtained from Plant and Food Research. The QTL analysis was re-run using chromosome 10 as a co-factor. No other peaks were observed that were above the threshold of significance at a LOD of 6.4. This indicated that one or multiple loci are located under the QTL peak, with no other sites contributing to resistance. However, due to the noise associated with the GBS data, it was difficult to assess which GBS markers were on either side of a recombination event in

recombinant individuals. An indicator of error is the difference between markers on the same read in the new map. As the two markers S10_7718533 and S10_7718620 are 87 bp apart, they are likely to be from the same read, but they were mapped to be 2.8 cM apart from each other. Recombination between markers so close together is unlikely.

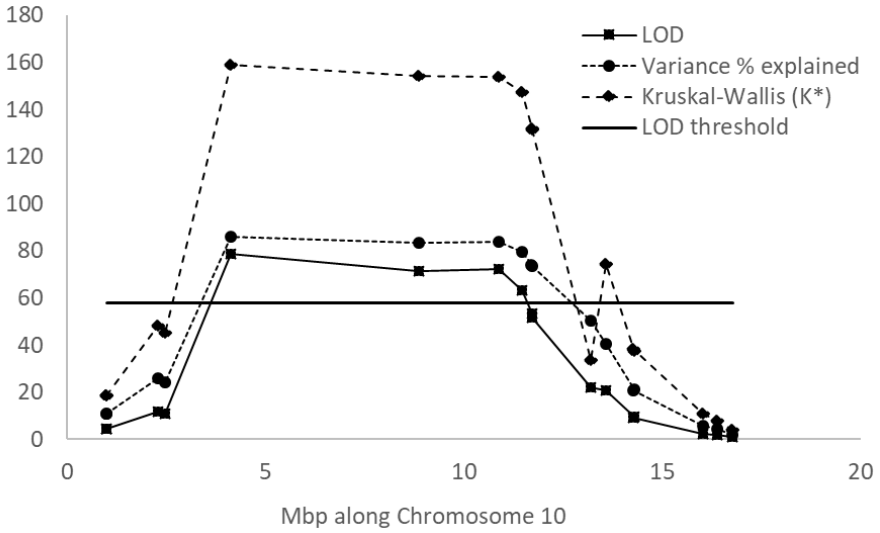


Figure 2.6. QTL map of the 20-Mb length of chromosome 10 from the B1 family of *A. chinensis* var. *chinensis*. A wide peak is shown with logarithm of the odds (LOD) scores of 78.67 and 63.48 at 4.12 Mb and 11.45 Mb along chromosome 10. The marker with the highest LOD score accounted for 86.2 % of variance. Kruskal-Wallis Test values agreed with the LOD score (p=0.001).

To clean up the noise, the raw GBS sequencing files were re-mapped with a newer bioinformatics pipeline (BWA-MEM) by the author to a newer reference genome known as PS1.69.0 (Pilkington et al., 2018). This method of mapping produced 1391 unique markers on chromosome 10 compared with the 497 from the tassel pipeline with the same filters applied. QTL analysis of the GBS data showed five markers with LOD scores over the permutation test interval of 57.9. These markers landed in similar positions to those on the previous genome on chromosome 10, but as the permutation test interval was much higher, there was a smaller region between significant markers of 7.59 Mb between 4.12 and 11.71 Mb (Figure 2.6). The marker with the highest LOD score of 78.67 explained 86.2 % of the variance in phenotype.

The increased marker density reduced some of the error, producing a sharper peak on the QTL map, but the error from the GBS data remained, restricting the ordering of markers at a fine level. To visualise the QTL region, LOD maps were made to look at the phenotype quantitatively with the 1-3 scale, but this did not fit the binomial nature of resistance conferred by Female 6. To test whether this had an effect, a Kruskal-Wallis test was run on the GBS data in Map QTL. The results showed significant association of markers on chromosome 10 with the phenotype, but there was no distinction between markers from 1 Mb to 14.27 Mb.

Aligning with the BWA-MEM pipeline and mapping with interval mapping generated a better QTL map, but to get more accurate genotyping results, another approach was required. The decision was made to fine map the B1 family to narrow the region associated with resistance, using InDel markers for the resistant female alleles under the interval map QTL peak.

2.4.5 Fine map of the B1 family

To fine map the B1 family in the region under the QTL peak on chromosome 10, 22 microsatellite markers were designed, which spanned a region of 4.1 to 12.7 Mb on chromosome 10 (Table 2.4). Phenotype data and DNA of sufficient quality were obtained from 180 plants from the B1 population for this mapping. Fifteen recombinant plants were found, which narrowed the region of interest to a 5.74-Mb region between 5.17 and 10.91 Mb. At this resolution, there were only three recombinant plants with phenotypes suitable for narrowing the region further. Since with few recombinants there is a high chance of including phenotyping or genotyping errors, other populations were bred using Female 6 as a parent to find further recombinants (outlined in Chapter 4).

Table 2.4. Markers used to fine map recombinant individuals from the B1 family of *A. chinensis* var. *chinensis*. All markers were designed to have a size between 136 and 451 bp between 4187500 bp and 13845500 bp along chromosome 10. All primers used a melting temperature of 58.5 – 60.5 °C. All forward primers had a TGTAAAACGACGGCCAGT M13 tag added onto the 5' end. Variants present in the parent Female 6 are shown with both alleles separated by a forward slash. 'Ins.' refer to insertions, and 'del.' deletions present in the amplified fragment, while 'cons.' indicates consensus with reference.

Marker name	Chromosome 10 location (Mb)	Variant/s in parent "Female 6"	Expected fragment size	Forward Primer	Reverse Primer
Lat-4.1	4.187500	8 bp del. / cons.	136	5'-AGTTTCACTGCTGTATCAGGCA	5'-CGGGCCGTTCATTATTACGC
Lat-4.41	4.412500	9 bp del. / 21 bp del.	337	5'-CAGTCCGAGACTCACGTTCC	5'-TCGACCGAGCACAAAGGATC
Lat-4.8	4.856300	6 bp del / 2 bp ins.	392	5'-CGTTGCTCAGTGCTCTACCA	5'-GTCCACCTCTATCATGGAACCC
Lat-5.1	5.168820	8 bp del. / cons.	148	5'-CTACCGCATCCTAGCAGACC	5'-TTGGTCTATTTCACTATTGTATGCTTT
Lat-5.6	5.661800	1 bp del. / cons.	310	5'-CGGTTGAACCACGGTTGAAC	5'-CACCCAGCCAATCCAAATGC
Lat-6.4	6.456400	3 bp ins. / cons.	376	5'-GGAGTATGAATTTACAAACACTAGTGA	5'-AAAATTAAGTTAAACTGGGTCAAACAA
Lat-7.3	7.339100	21 bp del. / 1 bp ins.	435	5'-TCTTGAGGTTCCGCTGTAGC	5'-TGTCAAGACATTTGCAGACGA
Lat-8.6	8.630980	5 bp del. / cons.	200	5'-TGAGTTGTAGAAGCACACGTTC	5'-GGTAATCGAAGGAGGAGTCTGC
Lat-9.1	9.187300	11 bp ins. / cons.	399	5'-AACACAACAATGGCTTCACC	5'-GCTCGAGAACTAGCAGTTGGT
KellII	9.757660	16 bp del. / 3 bp ins.	461	5'-TGGTCTGCCGGGTTATCTGTTA	5'-TCCCTGACTACATGAACTTC
Lat-9.9	9.916400	4 bp ins. / 1 bp ins.	403	5'-AGCTTAATGGTTGCCCGTCA	5'-CCGGTTAGGGTGCAGATACC
Lat-10.3	10.363300	9 bp del. / cons.	233	5'-AAAAGCAGGAGTTGTGTCAAAA	5'-CTCAACCTACCACCCAGAACA
Lat-10.9	10.903200	4 bp ins. / 1 bp del.	160	5'-GTTTGCAAGAACCCGAAAGCT	5'-ATATGCTCTCAATGCAATAAATAAGCA
Lat-11.5	11.592140	2 bp del. / cons.	300	5'-ACAGATGCCCCCATGAAGTT	5'-CTTCCACTTATCCCCGCTCC
Lat-11.6	11.646800	3 bp ins / 5 bp ins.	150	5'-TGCCTCTCCTCTATCAGCCT	5'-CCAAGCCCAGTGGTTAAGCT
Lat-11.7	11.788400	3 bp ins. / cons.	272	5'-GCCACATTTCTACTTTTTCTCTTTCA	5'-CCGGTCATTAAGAGGCGAGG
Lat-11.8	11.863700	5 bp del. / 3 bp ins.	326	5'-AGGTGCGAAATGATAAAGACCA	5'-AAACTACAGGGACCATTAGTGAA
Lat-11.9	11.923800	2 bp del. / cons.	246	5 '-GGGTGGTGGTGGTTGTGTAA	5'-GCAATGACGTGCACGGTATG
Lat-12.2	12.234100	10 bp del. / 9 bp del.	136	5'-TCGAACGTCAGACCAATAGCT	5'-TCACCTGGATATATCTCGATTTTGGA
Lat-12.4	12.405200	2 bp del. / cons.	336	5'-TCGGGATAGATACCCCCATTCA	5'-ACATGCTATTCCCCAATCCCT
Lat-12.7	12.752200	6 bp del / 2 bp ins.	451	5'-ACCCTCTTTATCAAAGCGATTAAGA	5'-TGTGTTGGTTATGACTGATGGT
Lat-12.9	12.937600	9 bp del. / 10 bp ins.	216	5'-TCCATCACACCACAACAGTGA	5'-TGAGGCTAGCTCACCATATTACA
Lat-13.8	13.845500	3 bp del. / 1 bp ins.	326	5'-TCACAAGGCAAATCGGGGAA	5'-TGGTTTCTCCTTTGTGGATGA

2.4.6 Genes within the region of interest for resistance to *H. lataniae*

The 5.74-Mb region found by fine mapping between 5.17 and 10.91 Mb of chromosome 10 aligned to the manually annotated Red 5 PS1.69.0 kiwifruit reference genome (Pilkington et al., 2018) was viewed in JBrowse (Buels et al., 2016). This region was found to contain 139 manually annotated genes with a low recombination rate, likely caused by the presence of the centromere in this region (Ding et al., 2019). Twenty-seven of the 139 genes were annotated to have structures involved with pathogen resistance in other plants (Table 2.5).

Table 2.5. Genes of *A. chinensis* var. *chinensis* annotated to have structures involved with pathogen resistance within the 5.17 and 10.91 Mb region on chromosome 10 associated with resistance to *H. lataniae*.

Name	Chromosome 10 position (Mb)	Homology	Function	Reference
Acc11247.1	4.9700	WRKY, transcription	Disease transcription	Pandey and Somssich, 2009
Acc11270.1	5.4125	F-box protein	Disease resistance	Piisilä et al., 2015
Acc11271.1	5.4250	Metacapase	Programmed cell death	Fagundes et al., 2015
Acc11277.1	5.9875	Polyadenylate-binding protein	Lesion promoting	Li et al., 2000
Acc11278.1	6.1875	NBS-LRR Toll/interleukin-1 receptor	Resistance protein	McHale et al., 2006
Acc11286.1	6.3625	Thioredoxin domain-containing protein	Hypersensitive response triggering	Rivas et al., 2004
Acc11302.1	6.8125	Ankyrin	Protein binding	Fowler et al., 2009
Acc11303.1	6.8250	Callose synthase	Cell wall building callose synthase	Jacobs et al., 2003
Acc11308.1	6.9320	Phytosulfokine-beta protein	Damage associated molecular pattern detecting	Zhang et al., 2018
Acc11309.1	6.9550	F-box protein	Disease resistance	Piisilä et al., 2015
Acc11312.1	7.0250	MYB like gene	Cuticle wax depositing	Zhang et al., 2019
Acc11354.1	8.7325	Hsp70-Hsp90	Auxin and gibberellic signalling protein components	Zhang et al., 2015
Acc06636.1	9.4700	Polyadenylate-binding protein	Lesion promoting	Li et al., 2000
Acc11367.1	9.6125	F-box protein	Disease resistance	Piisilä et al., 2015
Acc11369.1	9.6375	Polyadenylate-binding protein	Lesion promoting	Li et al., 2000
Acc11375.1	9.7250	F-box protein	Disease resistance	Piisilä et al., 2015
Acc11383.1	9.9000	Ankyrin	Protein binding	Fowler et al., 2009
Acc11384.1	9.9100	Ankyrin	Protein binding	Fowler et al., 2009
Acc11389.1	9.9600	WRKY disease transcription factors	Disease transcription	Pandey and Somssich, 2009
Acc11390.1	9.9650	leucine aminopeptidases	Protein binding	Vo et al., 2015
Acc11391.1	9.9800	(S)-coclaurine	Isoquinoline alkaloid precursor protein	Wang et al., 2018
Acc11440.1	10.3650	F-box protein	Disease resistance	Piisilä et al., 2015
Acc11417.1	10.8050	Vinorine synthase	Flavonoid biosynthesis	Gu et al., 2021
Acc11418.1	10.8175	BAHD acyltransferase	Pathogen induced accumulation of salicylic acid	Zheng et al., 2009

While many of these genes could be involved in resistance processes, genes with nucleotide binding sites (NBS) and a leucine rich repeat (LRR) region are most commonly associated with resistance genes in other plants (McHale et al., 2006). Within the fine map region of interest, a single gene "Acc11278.1" had NBS-LRR regions and a resistance protein Toll/interleukin-1 receptor homology (TLR) site. This gene was annotated as being like *Tobacco mosaic virus* (TMV) resistance proteins. This 3,208 bp gene contains an N-terminal NBS, a TLR domain and a c-terminal LRR domain, like a winged helix DNA-binding protein. The first exon contained a 264-bp long N-terminal NBS site, which was conserved among resistant individuals with a single heterozygous SNP site. The second and third exons of 1,102 bp and 438 bp respectively were less conserved, with 63 and 16 SNPs, respectively, among resistant individuals. The 140-bp C-terminal LRR at exon four was fully conserved between resistant individuals. The sequence of this gene was blasted against available sequences in the BioView database at Plant and Food Research for similarity to genes in other species. The TMV resistance-like protein presented between 58 and 59 % nucleotide identity to TMV resistance-like proteins in prune (*Prunus mume*), apple (*Malus domestica*) and pear (*Pyrus* x), 52 % nucleotide identity to a TMV

resistance protein in *Solanum lycopersicum* and 30 % nucleotide identity to the TLR-NBS-LRR class pathogen resistance proteins RLM1A and RPS6 in *Arabidopsis thaliana*, which trigger a hypersensitive response to *Leptosphaeria maculans* and *Pseudomonas syringae*, respectively.

2.4.7 Ranking candidate genes found within the region of interest for resistance to *H. lataniae*

To rank candidate genes for resistance to *H. lataniae* within the region of interest, the average percentage of polymorphic sites was analysed for chromosome 10 to provide a baseline for each polymorphism type. The average percentage of identical polymorphisms that were dissimilar to those of Female 2 over chromosome 10 were calculated to be 1.52 % for identical SNPs, and 0.07 % for identical InDels (Figure 2.7, and Figure 2.8). Analysis of the 24 candidate genes for resistance to *H. lataniae* identified five of the candidates with a percentage of identical SNPs above the average for all SNPs on chromosome 10 i.e., Acc11417.1, Acc11278.1, Acc11354.1, Acc11367.1, and Acc11369.1 (Figure 2.7). The NBS-LRR gene Acc11278.1 also had an above-average percentage of identical InDels along with Acc11367.1 and Acc11417.1. Detailed examination of exons from the candidate genes with an above-average percentage of identical SNPs showed Acc11278.1 exons 1, 2 and 3, had a higher percentage of identical SNPs and InDels than all the other candidate genes combined (Figure 2.8). The exon regions from Acc11278.1 also showed the highest percentage of SNPs dissimilar to those of Female 2 (Figure 2.8).

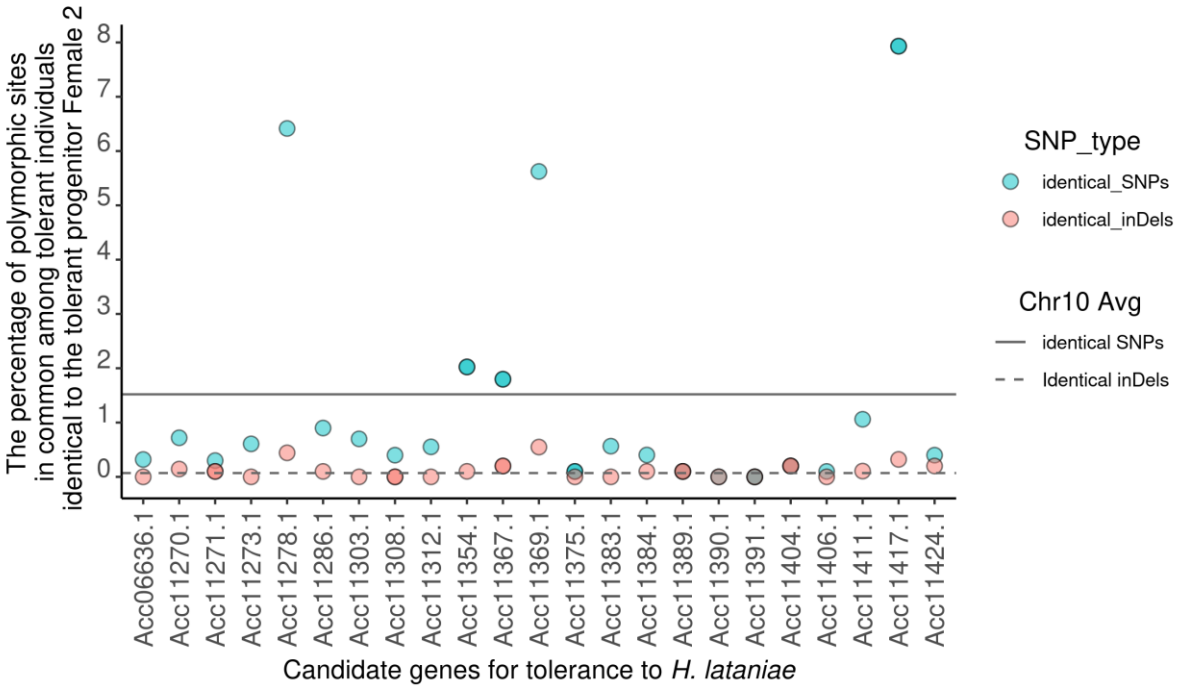


Figure 2.7. The percentage of polymorphisms within genes in *A. chinensis* var. *chinensis* , found among six individuals resistant to *H. lataniae,* which are identical to those of their resistant ancestor Female 2.

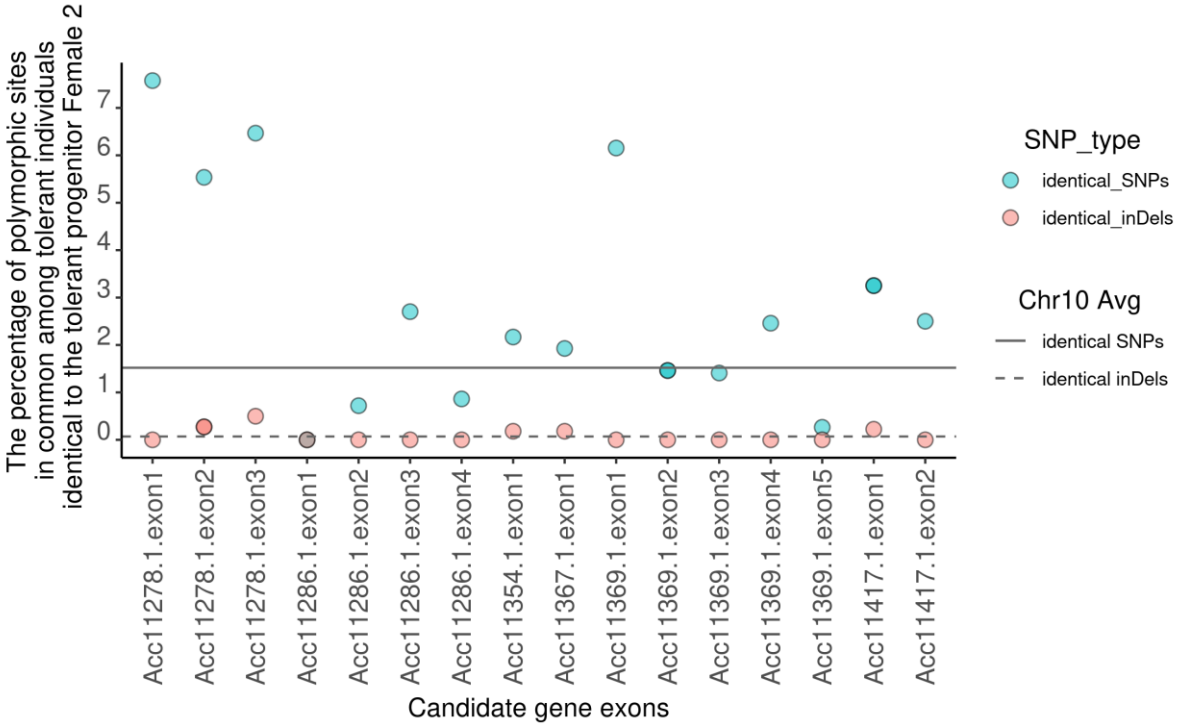


Figure 2.8. The percentage of polymorphisms within *H. lataniae* resistance candidate gene exons of *A. chinensis* var. *chinensis* which are identical or dissimilar to those of their resistant ancestor Female 2.

From these results, the candidate genes controlling resistance to *H. lataniae* can be tentatively ranked from most likely to least likely to be the gene controlling resistance to *H. lataniae* (Table 2.6). The results showing the genes most likely to be controlling resistance to *H. lataniae* are ranked with the most likely NBS-LRR gene "Acc11278.1" ranked highest, the F-box/kelch-repeat protein "Acc11270.1" ranked second, and the Metacapsase-1 like gene "Acc11271.1" ranked third. The NBS-LRR (Acc11278.1) gene had a higher percentage of polymorphisms in common with Female 2 in gene exon regions than in "Acc11417.1", hence the higher ranking of "Acc11270.1". However, all these genes are within one megabase of each other in an area of the genome in which low recombination was detected based on the data analysed for this study. Therefore, it is likely that the variation in the percentage of polymorphisms described within genes the same as those of Female 2 is likely to be due to the higher rate of polymorphism within those genes. This will not give an accurate ranking of genes, unless resistant plants that are recombinant are added to the set.

Gene Name	Rank	Position on chromosome 10 (Mb)	Percentage of SNPs identical among five resistant individuals within the gene region ± 1 Kb	Percentage of SNPs identical among five resistant individuals within the gene exons
Acc11417.1	1	10.801201	4.13	4.33
Acc11278.1	2	6.186232	3.43	3.46
Acc11369.1	3	9.636592	3.09	1.18
Acc11354.1	4	8.730230	1.06	1.07
Acc11367.1	5	9.608805	1.00	1.69
Acc11286.1	6	6.354021	0.50	1.14
Acc11270.1	7	5.407778	0.43	not assessed
Acc11303.1	8	6.820981	0.35	not assessed
Acc11273.1	9	5.796838	0.30	not assessed
Acc11424.1	10	10.918674	0.30	not assessed
Acc11383.1	11	9.896326	0.28	not assessed
Acc11312.1	12	7.021588	0.28	not assessed
Acc11384.1	13	9.904729	0.25	not assessed
Acc11404.1	14	10.363738	0.20	not assessed
Acc11308.1	15	6.931403	0.20	not assessed
Acc11271.1	16	5.424807	0.20	not assessed
Acc06636.1	17	9.464648	0.16	not assessed
Acc11389.1	18	9.958227	0.10	not assessed
Acc11375.1	19	9.710936	0.05	not assessed
Acc11406.1	20	10.519983	0.05	not assessed
Acc11390.1	21	9.959688	0	not assessed
Acc11391.1	22	9.978148	0	not assessed

Table 2.6. Rank of candidate resistance genes of *A. chinensis* var. *chinensis* in the 5.74-Mb region of interest for *H. lataniae* resistance on chromosome 10 (Mb). The rank was based on the percentage of polymorphisms which were in common among six resistant individuals descended from Female 2. The nucleotide positions included for each gene extended 1 Kb beyond the 3' and 5' ends of the annotated gene region. The percentage of polymorphisms in gene exons was based only on bases analysed within gene exon regions. There was a single re-ranking of the order of genes most likely to confer resistance to *H. lataniae* in Acc11278.1. The NBS-LRR (Acc11278.1) gene had a higher percentage of polymorphisms in common with Female 2 in gene exon regions, but a lower percentage than that of Acc11417.1 when the whole gene region was analysed.

To gauge the percentage of nucleotide sites with polymorphisms, bins of 0.5 Mb were made for all polymorphisms along chromosome 10 and the polymorphism number calculated as a percentage of bases in each bin (Figure 2.9). The resulting plot showed an above-average percentage of SNPs identical to that of Female 2 below 2.0 – 12.0 Mb. This indicates that within

the individuals tested it is likely that the 2-12 Mb area is conserved among the six resistant individuals, with possible recombination occurring outside this region.

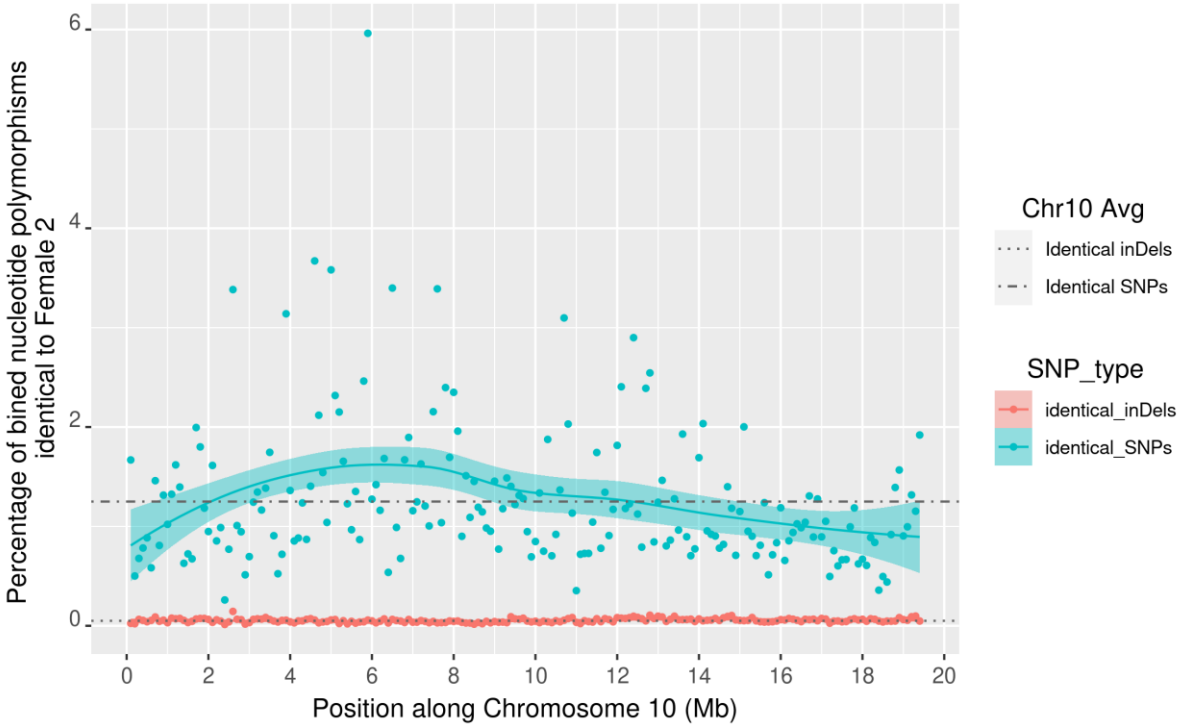


Figure 2.9. The percentage of polymorphisms in 0.1 Mb bins on chromosome 10 of *A. chinensis* var. *chinensis* which are identical to those of Female 2. An above-average region of identical single nucleotide polymorphisms is shown from 2 Mb to 10 Mb along chromosome 10, indicating recombination has occurred to incorporate different chromosome arms at both ends of the chromosome from 0 to 2 Mb and from 12 Mb in at least one of the five individuals tested.

2.5 Discussion

The main commercial kiwifruit cultivars grown in New Zealand are susceptible to the armoured scale insect Latania scale - *H. lataniae* - (personal communication Cathy McKenna). This is primarily due to the high cost and slow turnaround of breeding kiwifruit from seed to a vine that can be reliably phenotyped for resistance to *H. lataniae*. However, these restrictions on conventional phenotyping for kiwifruit breeding can potentially be overcome using molecular markers developed to differentiate resistant and susceptible individuals at the seedling stage for the selection of parents resistant to *H. lataniae*. But before markers can be developed, the genetic architecture of resistance or gene/s markers associated with the trait needs to be identified. Markers developed with close association to resistance loci can provide a cost-effective solution for commercial breeding programmes to incorporate resistance to *H. lataniae* into new kiwifruit cultivars. Ideally, cultivars that incorporate resistance to *H. lataniae* should not need control of this insect.

To identify the genetic architecture of resistance to *H. lataniae* in kiwifruit (*A. chinensis* var. *chinensis*), three families descended from a common ancestor (Figure 2.1) were studied. From that work, a single locus spanning a 5.74-Mb region on chromosome 10 between 5.17 and 10.91 Mb was identified. This 5.74-Mb region accounted for all resistance associations with genomic data in the families tested. Further investigation of the genomic polymorphisms on chromosome 10 that were in common among resistant parents and Female 2, which is directly descended from the common ancestor Female 1, highlighted a likely NBS-LRR candidate gene for resistance to *H. lataniae*. These data and the population segregation ratios from susceptible by resistant crosses led to the hypothesis that resistance to *H. lataniae* is governed by a single NBS-LRR gene (11278.1). The microsatellite and InDel markers designed to target the region associated with resistance are the first published molecular markers for resistance to *H. lataniae* in kiwifruit and, to our knowledge, the first molecular markers for resistance to armoured scale insects in plants.

The identification of a single locus which triggers a hypersensitive response to *H. lataniae* (Hill et al., 2011) is consistent with the gene-for-gene interaction typically observed in plants that have a prolonged intimate relationship with pests such as Hessian fly larvae (Aljbory et al., 2020), nematodes (Rossi et al., 1998), or aphids (Balint-Kurti, 2019; Kamphuis et al., 2019;

Klingler et al., 2009; Pointeau et al., 2012; Zhang et al., 2017). Since *H. lataniae* is immobile once feeding, has a lifelong relationship with its host, and is a parthenogenic insect in New Zealand, it fits with the hypothesis that Female 6 has a dominant resistance gene which responds to the dominant avirulence gene in *H. lataniae* by stimulating the salicylic acid signalling cascade found by (Hill et al., 2015).

As well as identifying a genetic basis for resistance to *H. lataniae* in kiwifruit, this study also identified several InDel markers associated with resistance that can be used for selective breeding in kiwifruit. Molecular markers have been employed in kiwifruit science to develop linkage maps (Fraser et al., 2004; Fraser et al., 2009; Pilkington et al., 2018), characterise kiwifruit diversity (Huang & Ferguson, 2006; Kamphuis et al., 2019), and to make associations with Psa resistance (Tahir et al., 2019). However, no pest or pathogen resistance markers are used in the green, gold and red kiwifruit breeding programmes (personal communication Alan Seal) (Hale et al., 2018). This is likely because developing useful markers as close as possible to resistance genes requires numerous individuals to be phenotyped and genotyped to have sufficient recombination events around resistance genes. As there is a high cost of establishing large families, and the commercial breeding cycle of kiwifruit is greater than three years, the size of these populations is often restricted. Fortunately, for the experiments on *H. lataniae* resistance described here, Plant and Food Research had multiple kiwifruit families available using resistant and susceptible parents (Cheng et al., 2015; Tahir et al., 2017). These families were used to develop molecular markers. The first InDel marker identified, Ke1111, accurately predicted phenotypes in its training family (A2) and the related validation population (A1) (Table 2.3).

As these populations had a low number of plants, the larger, replicated, GBS sequenced family B1 was QTL mapped. This identified sixteen recombinant individuals in the region of interest first identified in the A2 and A1 populations. However, there was too much sampling error or DNA contamination in the GBS markers, and possibly phenotyping error, to enable accurate ordering of those markers.

The fine map created for the B1 family (Table 2.3) using InDel markers had fewer errors than the GBS data, which enabled a reduction in the size of the region of interest to a 5.74-Mb locus on chromosome 10 at 5.17 to 10.91 Mb. However, the ability of the fine mapping approach to

narrow down the region of interest further was limited by the lack of individuals with recombination in the region. This was likely due to a centromere being present in the region of interest reducing the rate of recombination around its binding site (Peñalba & Wolf, 2020; Zelkowski et al., 2019).

Despite the lack of recombination detected in the 5.74-Mb region, further clues about the genes contributing to *H. lataniae* resistance were extracted from mutations found within the region of interest. It was hypothesised that genes more similar to their resistant progenitor, among descendants which retain resistance, will have a greater chance of their functionality being retained. This is due to the accumulation of mutations over time (Schoen & Schultz, 2019) in long-lived species such as kiwifruit. The percentage of polymorphisms retained between genes was sufficient to rank genes within the 5.74-Mb region of interest in order of likelihood to remain functional and thus confer resistance (Figure 2.7, Figure 2.8, and Table 2.6). When this approach was applied to all of chromosome 10, a signal of recombination was detected at both ends of chromosome 10 from the start of the chromosome to 2 Mb, and from 10 Mb to the end of the chromosome (Figure 2.9). The loss of alleles in common with the resistant ancestor in the recombinant ends of chromosome 10 would have further decreased the number of polymorphisms in common with the resistant ancestor and lowered the resistance rank for the affected genes.

The bioassay method developed for this study reduced handling error by leaving plants in the glasshouse when *H. lataniae* insects were applied to canes. This was possible due to the higher humidity generated by the wet tissue paper applied around the site of application. It is thought that this allowed greater numbers of crawlers to settle by preventing the thin flat crawlers from becoming dehydrated before finding a suitable site to settle. The modified bioassay method reduced the workload and time taken for the assay setup compared to the method described by Hill et al. (2015) for live plant assays, which involved seeding plants in the lab and moving plants to the glasshouse after settlement. This is likely to have reduced environmental error within treatments which could have been introduced through environmental variation between the lab and the glasshouse, as well as reducing damage to insects and plants due to handling. The use of three clonal replicates was a great advantage, allowing individuals with incorrect phenotyping results in one of the three plant clones to be identified and removed from the analysis. Despite the measures to reduce phenotyping error using the clonal replicate

data, errors were detected through investigation of markers in recombinant individuals in the fine map of the B1 family. The erroneous susceptible phenotype calls had markers for resistance which did not fit with other recombinant susceptible individuals or other recombinant resistant individuals. If there were insufficient recombinant individuals to check whether the region of interest of the incorrect call lined up with the rest, an incorrect region of interest position could have been established. This was a particularly important issue when narrowing down the region associated with resistance as the low number of recombinant individuals reduced the number of consistent boundary calls to increase confidence and correct errors such as these. Yet with a less distinctive and a poorly characterised phenotype, the reassessment of phenotyping calls to confirm phenotyping error may not have been possible.

Using the markers developed in these experiments for prediction of *H. lataniae* resistance should assist plant breeding, but it is important to note that the resistance observed in the A2, A1, and B1 families was inherited from the common resistant ancestors of Female 2 and Male 3 (Figure 2.1). Testing these markers on unrelated *Actinidia* populations that are resistant to *H. lataniae* may establish whether resistance to *H. lataniae* is conferred by a single gene across *Actinidia* germplasm or whether there are many variants of resistance to *H. lataniae* in *Actinidia* species. However, because the 5.74-Mb region is large, the *H. lataniae* resistance loci can dissociate from the markers designed to target it. This is particularly a problem when testing markers on unrelated populations which share the resistance gene, as there is a greater chance that linkage between the marker and resistance gene has been broken. The large region associated with resistance is also an issue when using these markers in future breeding programmes as the parents of populations will still need to be phenotyped to make sure that the association between the marker and resistance gene has not been broken. If the marker association is intact, these markers could be used to screen seedlings for individuals resistant to *H. lataniae*.

Using the markers developed here for prediction of *H. lataniae* resistance should significantly reduce the cost, time, workload and facilities required for *H. lataniae* phenotyping. The markers developed here are used by Plant and Food Research to further develop a universal marker for *H. lataniae* resistance that will work between distant families of kiwifruit. These markers would be particularly useful for developing male polleniser cultivars with resistance

to *H. lataniae*, since there are fewer competing traits in pollinisers other than pollen quantity and performance. Because pollinisers are often planted in blocks with fruiting plants, future pollinisers with resistance to *H. lataniae* would prevent these individuals from spreading *H. lataniae* crawlers to fruiting plants. A commercial polleniser could be crossed to an individual descending from the resistant ancestor of Female 2 and Male 3. Then, the resistance markers developed here could be used to select resistant seedlings.

While a lack of recombination in the 5.74-Mb region was an impediment for gene identification, it is an advantage for using markers in plant breeding programmes. This is because markers within regions of low recombination have a lower risk of being disassociated from their target gene by recombination and thus, will provide fewer false positives and false negatives to marker calls. Moreover, the non-recombinant region is a great target for breeding resistance to *H. lataniae* into kiwifruit as the parthenogenic nature of *H. lataniae* in New Zealand may not allow the adaptation of avirulence genes in *H. lataniae* as rapidly as in those species which undergo sexual recombination. Further, the selection pressure to overcome the resistance gene in a polyphagous species such as *H. lataniae* will be lower than that in a monophagous species dependent for its survival on a specific cultivar. Plants which carry this resistance could be a great advantage to future kiwifruit cultivars. Future work should investigate whether other genes for resistance exist in families unrelated to Female 1. If a second gene for resistance could be identified, the two resistance genes could be stacked into breeding parents to make commercial cultivars better able to prevent resistance to resistance genes forming in *H. lataniae.*

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the student and the student's main supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the student's contribution as indicated below in the Statement of Originality.

Student name:	Casey Flay
Name and title of main supervisor:	Vaughan Symonds
In which chapter is the manuscript/published work?	Chapter 2
What percentage of the manuscript/published work was contributed by the student?	100% unless otherwise stated in text

Describe the contribution that the student has made to the manuscript/published work:

Some resources were existing at Plant and Food Research which the student used to start their work.

Please select one of the following three options:

- ○ The manuscript/published work is published or in press
  Please provide the full reference of the research output:
- ○ The manuscript is currently under review for publication
  Please provide the name of the journal:
- ✓ It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Student's signature:	Flay	Main supervisor's signature:	Vaughan Symonds

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

Doctoral Research Committee
July 2022

3 Breeding kiwifruit (*A. chinensis* var. *chinensis*) resistant to latania scale (*Hemiberlesia lataniae*)

3.1 Abstract

Hemiberlesia lataniae is a problematic pest of kiwifruit (*Actinidia chinensis*) production and export, with more than NZD 77 million per annum spent on controlling *H. lataniae* in New Zealand alone. Cultivars that are resistant to *H. lataniae* could mitigate that cost and reduce the reliance on pest control. Individuals resistant to *H. lataniae* have been identified in several related genotypes of diploid *A. chinensis* var. *chinensis*. Resistance to *H. lataniae* in these individuals has been associated with a region of chromosome 10 (see Chapter 2 for details). *Actinidia chinensis* var. *chinensis* exists as different ploidy races, and both diploid and tetraploid cultivars are grown in New Zealand. To help facilitate the breeding of new diploid and tetraploid cultivars resistant to *H. lataniae* three experimental approaches were undertaken. The first approach identified variant sites in whole genome sequence data that were associated with resistance to *H. lataniae*. The variant sites were then filtered based on Capture-Seq data from 403 individuals. Two resistance-associated variants were identified in 29 of the 403 individuals tested. The second approach was designed to characterize the architecture of resistance from resistant plants that are unrelated to those previously identified with resistance to *H. lataniae*. F_1 seedlings from a cross between resistant and susceptible parents were phenotyped, but the families' lack of segregation for resistance and susceptibility prevented genotyping. In the final experiment, a resistant diploid female was crossed with four tetraploid males to integrate resistance to *H. lataniae* from diploids into elite tetraploid lineages. Tetraploid offspring were selected, and molecular markers used to screen seedlings for resistance to *H. lataniae*. As elite breeding males were selected based on their prevalence in elite breeding populations, it is expected that the progeny of the F_1 cross should combine elite fruit characteristics with resistance to *H. lataniae*.

3.2 Introduction

Kiwifruit (*Actinidia chinensis*) is a recently developed crop, bred from wild populations within the last 120 years (Datson & Ferguson, 2011; McNeilage, 2014). The wild populations possess variation for several characteristics including fruit size, colour, and shape, ploidy, and pest and pathogen resistance (Gea, 2011; McNeilage, 2014; Yan et al., 1997). Among the many pests of kiwifruit, the armoured scale insect (*Hemiberlesia lataniae*) can cause significant damage to both vines and fruit. The damage caused, with the associated packing, processing, and export quarantine restrictions, costs the kiwifruit industry more than NZD 77 million per annum (personal communication Cathy McKenna).

Resistance to *H. lataniae* is difficult to phenotype because it may be controlled by molecular mechanisms activated on the reception of elicitors likely from *H. lataniae* saliva (Hill et al., 2011; Hill et al., 2016). These elicitors, and the signalling cascades that respond to them to promote resistance, take time to function and are not visible to the eye. The lack of an immediately apparent phenotype, combined with the high cost and difficulty involved with phenotyping for resistance, has resulted in most genotypes being poorly characterised for *H. lataniae* resistance. A substantial part of the cost of the *H. lataniae* resistance bioassay is the time and resources required to complete it. The bioassay takes more than eight weeks to run, and the work involved in phenotyping one genotype is around 1 hour from setting up the canes to data entry. Access to a temperature- and humidity-controlled room can also restrict phenotyping throughput, with a 9.2 m^2 room enabling the phenotyping of around 300 plants in one year. Yet, plants can only be phenotyped for six months of the year. The short time frame is dictated by the requirement for winter-dormant canes to complete the bioassay on plants from the field. Collected canes must be phenotyped within six months before cane decay affects the resistance phenotype. Despite the challenges, some individuals and populations have been phenotyped for resistance to *H. lataniae* using a bioassay (M. G. Hill et al., 2006). The work completed in Chapter 2 built upon that prior study by tracking the phenotype from a single kiwifruit lineage into three related families. Genotyping those families and using standard association of genotype with phenotype led to the identification of an 8.6 Mb region between 4.1 and 12.7 Mb on chromosome 10 associated with resistance to *H. lataniae*.

Identifying a region associated with resistance is useful when breeding resistant populations because the resistance trait can be inferred based on the inheritance of specific genotype variants. Bioinformatics can then be used to match variants from existing genomic datasets to identify individuals which carry the same trait. Yet, the type of genomic data can influence how much can be inferred from these data. On one end of the scale, data from microsatellite markers and single nucleotide polymorphisms (SNPs) can identify individual alleles at a low cost for a few markers, but the techniques do not scale well and there is a high cost per marker compared to next generation sequencing methods. Conversely, the process of getting whole genome sequence (WGS) data is costly but it covers the entire genome. Newer techniques, such as Capture-Seq, can bridge the gap by obtaining a subset of WGS data that can be captured at a lower cost. Capture-Seq obtains sequence data for a subset of the genome in a repeatable, cost-effective way. Capture-Seq works by using custom RNA sequences, termed baits, which are designed to bind and subsequently amplify captured DNA fragments from gene exomes with PCR (Mercer et al., 2012). PCR fragments are then sequenced to produce data that are fully compatible with whole-genome sequence data. This method is more repeatable than genotype by sequencing (GBS) techniques since Capture-Seq focuses on the genomic areas preselected by the custom RNA fragment baits. Targeting sites in this way generates an orthologous multi-locus array that can be used to identify nucleotide polymorphisms between individuals (Andermann et al., 2020). Because the Capture-Seq data are a subset of WGS data, variants found to be associated with resistance that are also in the Capture-Seq reads can help identify individuals carrying those variants. Because generating Capture-Seq data comes at a low cost, it is feasible to genotype entire populations to capture allelic diversity among individuals.

Understanding the breadth of resistance loci present in wild populations is critical for breeding durable resistance. But identifying unique resistance mechanisms requires an approach that is not based on previously identified loci. This is because different mechanisms that promote resistance cannot be found at the same loci. Having multiple mechanisms for resistance in a cultivar is desirable because it reduces the chances of the pest developing tolerance to an individual resistance mechanism (Quenouille et al., 2013; Udagawa et al., 2020).

The process of characterising the resistance described in Chapter 2 benefited from the comparative ease of developing markers in diploids compared to tetraploids (Bartkiewicz et

al., 2018). However, breeding *H. lataniae* resistance into parents that can be used in commercial breeding populations is complicated by the diploid nature of resistant plants and the tetraploid nature of many commercial populations. This is because most breeding of gold-fleshed *A. chinensis* var. *chinensis* is done in tetraploid families (personal communication Paul Datson).

Utilising the resistance locus from a diploid in tetraploid populations requires the transfer of alleles from the diploids into tetraploids. While breeding between diploids and tetraploids typically produces mostly weak triploid plants (Hirsch et al., 2001), tetraploid offspring can be produced if the diploid parent produces unreduced (2*n*) gametes (Wu et al., 2014; Yan et al., 1997). When the parent producing unreduced (2*n*) gametes is crossed with the reduced (2*n*) gametes of a tetraploid, the result is (4*x*) tetraploid offspring (Yan et al., 1997). However, the ploidy of kiwifruit plants resulting from a diploid mother crossed with a tetraploid father can include diploid, triploid, and tetraploid seedlings. This is because reduced (1*n*) gametes are produced along with unreduced 2*n* gametes (Yan et al., 1997). Some males can produce 93 % tetraploids with 7 % diploids, while the same female with a different male can produce 52 % tetraploid, 44 % triploid, 2 % diploid and 2 % octoploid offspring (Yan et al., 1997). The tetraploid offspring also vary their allele configuration depending on the stage that the diploid parent failed to split its chromosomes during meiosis. Normal meiotic division for a diploid goes through two consecutive chromosome doubling events (Figure 3.1). After the doubling events, all sets divide to make haploid gametes. The plants that produce unreduced gametes miss chromosome separation at the first or second meiotic division stages (Wu et al., 2014; Yan et al., 1997). The result can be a doubling of some chromosomes and the complete loss of other chromosomes in second division restitution (Figure 3.1).

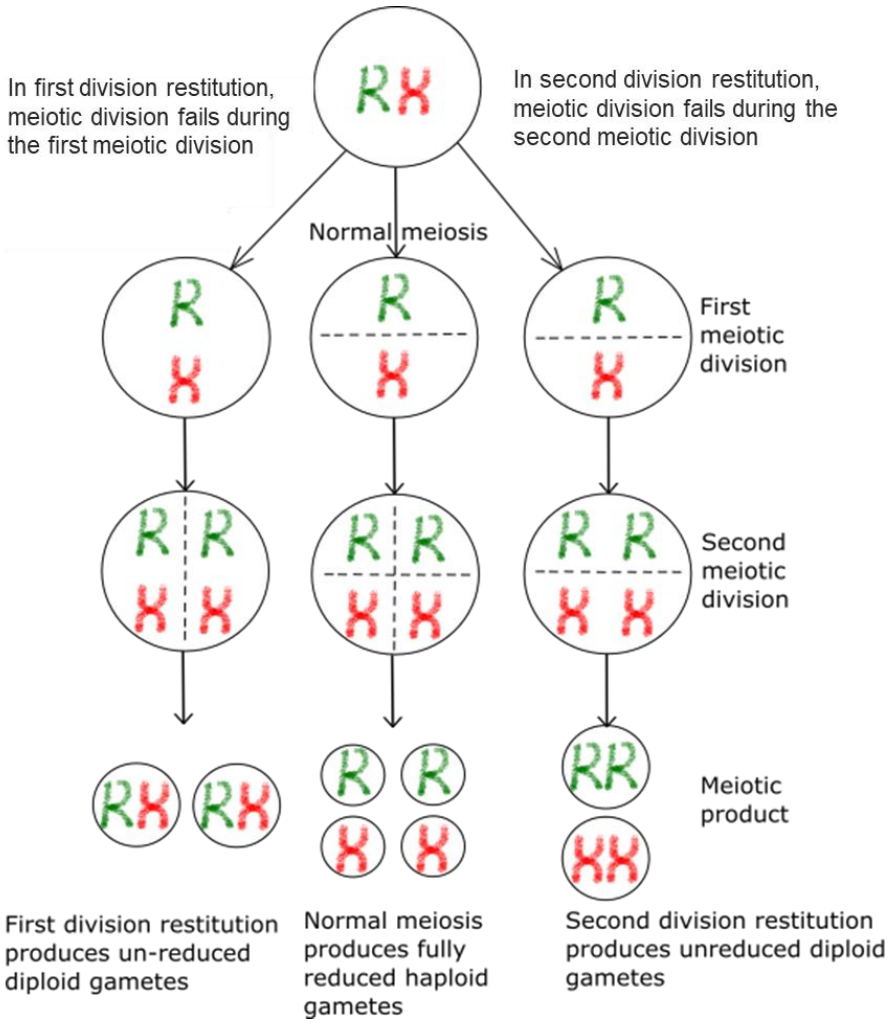


Figure 3.1. Diploids can produce 2*n* gametes if meiotic division fails at the first or second meiotic division stages. The chromosome coloured in green indicates the homologue that contains resistance. The chromosomes coloured in red have no resistance homologue. Figure reproduced from (Yan et al., 1997).

Another rare situation can occur in post-meiotic restitution when normal meiosis proceeds, but the fully reduced haploid gametes undergo a doubling stage prior to gamete fusion (Bastiaanssen, 1998). Whether the individuals that have two sets of chromosomes with resistance genes are produced by second division restitution or post meiotic restitution, they would both be advantageous for breeding. This is because parents carrying two copies of the resistance gene on different chromosomes are more likely to pass one of the chromosomes carrying dominant resistance genes onto the next generation than those carrying a single resistance gene.

The three approaches described in this chapter aim to (1) integrate the information available throughout WGS and Capture-Seq data sources to identify individuals resistant to *H. lataniae* in diploid and tetraploid *A. chinensis* var. *chinensis*, (2) identify different loci for resistance to *H. lataniae* from families of wild *A. chinensis* var. *chinensis,* and (3) breed the alleles for resistance to *H. lataniae* from diploids into elite tetraploid populations.

3.3 Methods

3.3.1 Using identity by state to identify individuals resistant to *H. lataniae* in higher ploidies

The first of three experiments described in this chapter analysed whole-genome sequence (WGS) data in the 8.6 Mb region of chromosome 10 associated with *H. lataniae* resistance identified in Chapter 2. This experiment was conducted to understand whether genomic variants could be identified that were consistently present in the five resistant individuals with WGS data that descended from the parent Female 1, but were not present in the susceptible individual, Female 7, also descended from Female 1 (Figure 3.2). Finding these variants in other individuals with available sequence data but without phenotype data could be used to predict the resistant phenotype and allow targeted phenotyping to be applied to a smaller set of plants. Using bioinformatics to associate phenotype with genotype can reduce the number of individuals that have to undergo the time-consuming and expensive process of phenotyping. The available phenotype data for *H. lataniae* resistance available for this work was from prior work done at Plant and Food Research using a cut cane bioassay described in Chapter 2. The WGS data available for this analysis was in the form of FASTQ files from individuals containing *H. lataniae* phenotype data, and Capture-Seq data were from Plant and Food Research's database. The Capture-Seq data used baits that were designed to target genes throughout the kiwifruit genome. Capture-Seq data from 403 individuals were processed by Tim Millar into variant call files aligned to the Red5:PS1.69.0 reference genome available from NCBI (Pilkington et al., 2018). Variant call files were imported into the array-based format Data.table in R for analysis by the author.

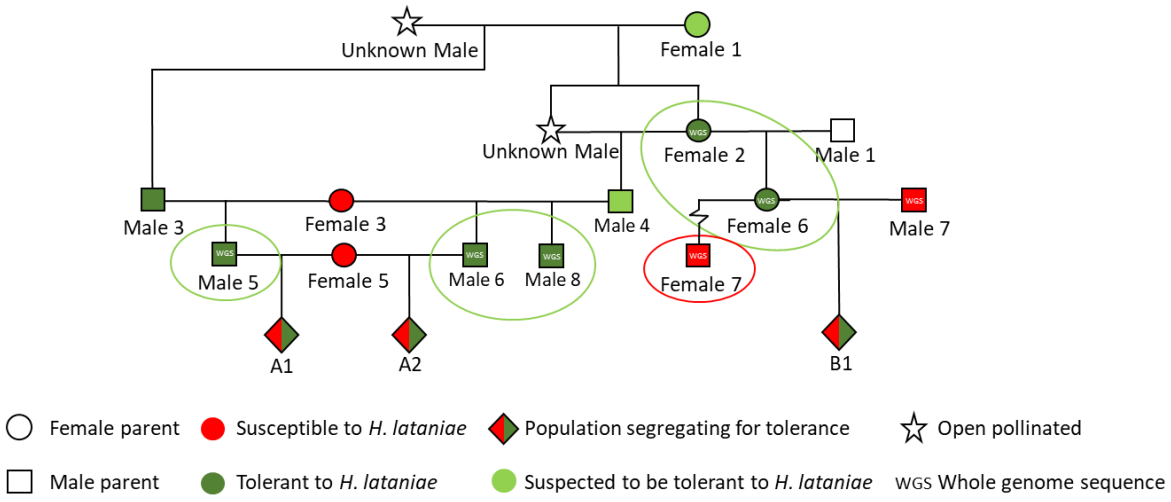


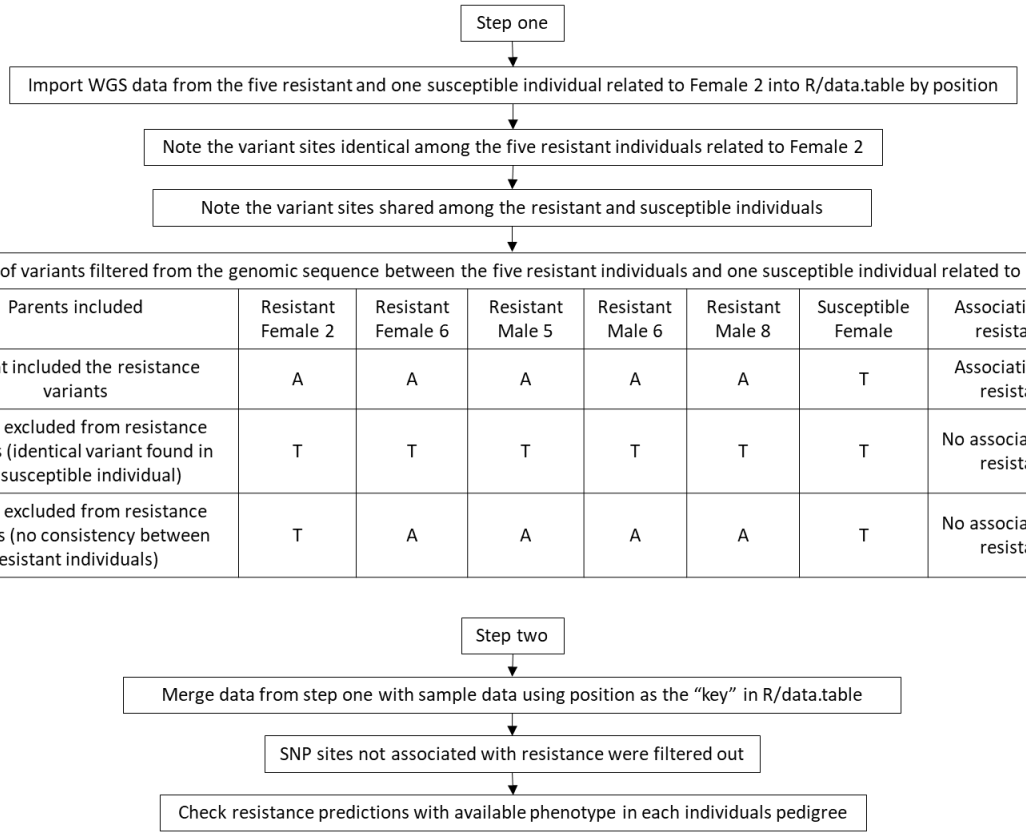
Figure 3.2. The pedigree of *A. chinensis var. chinensis* individuals included in the samples used to identify the SNPs in common among resistant individuals. The nodes circled in green represent the resistant individuals identical by descent from Female 1. The red circled individual "Female 7" was the susceptible individual, related to Female 1, used to reduce the number of homozygous alleles included with the resistance SNPs. Diamond nodes indicate families segregating for resistance to *H. lataniae* analysed in Chapter 2. A star indicates that crossing was by open pollination. Solid lines indicate direct descent from parents. The broken line leading from Female 6 to Female 7 indicates the use of Female 6 in Female 7's pedigree.

To eliminate variation from different bioinformatics pipelines, WGS data from five resistant individuals (Male 5, Male 6, Male 8, Female 2, and Female 6; Figure 3.2), which were descended from Female 1, and related to Female 2, were processed from raw FASTQ sequence files housed in the Plant and Food Research database. The same process was completed for the only susceptible individual, "Female 7", with WGS data that were related to Female 1. These data were aligned to the Red5 NCBI reference genome in R with BWA-MEM, indexed using Samtools, and SNP variants called using BCFtools "Mpileup" (Li et al., 2009; Samtools, 2020). To capture a bit more than the 6.5 Mb region of interest, the region from 3-13 Mb on chromosome 10 was extracted from the variant call files to a data table using the BCFtools query command. These data included sample, position, reference allele calls, alternate allele calls, and the sampling depth for each different reference and alternate allele. Alternate alleles with a sequencing depth of less than two were recoded as missing to reduce potential sequencing errors. To identify SNPs that were identical among the five related resistant

individuals and different from the alternate SNPs in the susceptible individual, a “resistant” note was made in a new column. A “susceptible” note was placed in an adjacent column where SNPs from the susceptible individual matched those of the resistant individuals.

To predict phenotypes based on genotypes, Capture-Seq data from 271 diploid and 132 tetraploid parental individuals at Plant and Food Research were compared with the data from step one (Figure 3.3). The Capture-Seq data were generated from baits designed to target genes across the *A. chinensis* var. *chinensis* genome and amplifying 10000 loci. Capture-Seq data were provided as VCF files that were generated using the same bioinformatics tools as the genomic data from step one. Ninety-two of the Capture-Seq bait sites were found to land in the 3-13 Mb region on chromosome 10. The genomic data were then filtered to a 300-bp region around each Capture-Seq bait site using IRanges. The 300-bp region around each bait landing site was selected based on each bait amplifying an average of 150 bp in the 3' and 5' directions of the bait landing site. Data tables from step one and step two were merged using allele position as "key" to combine the data table from step one with the Capture-Seq data. Merging the tables added Capture-Seq data as new columns to the data table from step one while repeating the SNP position for each individual from the Capture-Seq data. Data were filtered for each position where the alternate SNPs from the Capture-Seq data matched the five resistant individuals and did not match the susceptible individual. The remaining SNP sites and the individuals containing them were predicted to be associated with resistance to *H. lataniae*.

To investigate the likelihood that the predicted phenotypes were accurate, the pedigree of individuals predicted to be resistant was plotted with the available phenotype of parents to see if they were related to other resistant genotypes.



Example of variants filtered from the genomic sequence between the five resistant individuals and one susceptible individual related to Female 2

Parents included	Resistant Female 2	Resistant Female 6	Resistant Male 5	Resistant Male 6	Resistant Male 8	Susceptible Female	Association with resistance?
variant included the resistance variants	A	A	A	A	A	T	Association with resistance
variant excluded from resistance variants (identical variant found in the susceptible individual)	T	T	T	T	T	T	No association with resistance
variant excluded from resistance variants (no consistency between resistant individuals)	T	A	A	A	A	T	No association with resistance

Figure 3.3. Flow chart for analysis of resistant *A. chinensis* var. *chinensis* samples with single nucleotide polymorphisms from Capture-Seq data identical to those of Female 2.

3.3.2 Resistance to *H. lataniae* in families unrelated to Female 2

To identify different genetic sources of resistance to *H. lataniae*, resistant individuals unrelated to Female 2 were sought from available germplasm. Resistant phenotypes were found in two individuals, namely, Female 15, and Male 18, which originated from spatially and geographically isolated areas of China (Figure 3.4). The two resistant individuals were also identified as parents in crosses to susceptible individuals established at Plant and Food Research, Kerikeri, New Zealand site. Each resistant parent had been crossed to two susceptible individuals by Paul Datson in 2012 to make four small families, C1, D1, E1, and F_1, with 27, 16, 31, and 29 individuals, respectively (Figure 3.4). F_1 individuals were phenotyped by the author using the cut cane bioassay, described in Chapter 2. Segregation of resistance and susceptibility was desired as this would allow genotyping and subsequent identification of the resistance loci. However, issues with the phenotypic data of parents caused a lack of segregation for resistance and susceptibility in families. Contamination of the *H. lataniae*

colony with *A. nerii* detected during the phenotyping of the C1, D1, E1 and F1 families also resulted in a low number of individuals phenotyped. These issues prevented further marker work from proceeding.

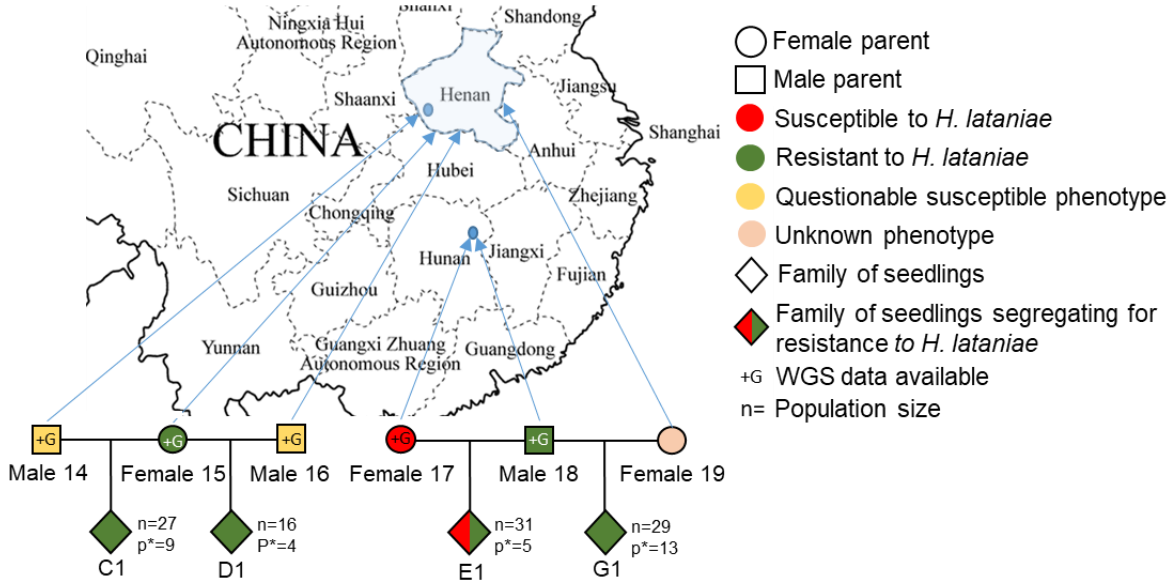


Figure 3.4. Pedigree tree for a wide cross of *A. chinensis* var. *chinensis* parental individuals originating from different parts of China. The blue arrows indicate each parent's approximate area of origin. The number of genotypes assayed (n=) is shown next to families with the number of genotypes with reliable phenotyping data (p*=). Yellow nodes indicate an individual where phenotyping data could have been inaccurate.

3.3.3 Crossing resistance to *H. lataniae* from a diploid into tetraploids

Within the populations phenotyped at Plant and Food Research, most individuals resistant to *H. lataniae* have been identified in diploid lineages related to Female 2. But significant effort has been put into breeding tetraploid plants due to their excellent fruit characteristics and resistance to the leading pathogen of kiwifruit, *Pseudomonas syringae* pv. *actinidiae* (Psa). To integrate *H. lataniae* resistance into higher ploidies of *A. chinensis* var. *chinensis*, a diploid female (Female 6) descended from Female 2 was selected as the source of resistance. Female 2 was selected because it had previously been shown to produce tetraploid offspring in crosses with tetraploid *A. chinensis* var. *chinensis.* It contains the *H. lataniae* resistance locus identified

in Chapter two, and it has excellent fruit characteristics. Four tetraploid male parents, Male 20, Male 21, Male 22, and Male 23, were selected to cross to the diploid Female 6 based on their prevalence in elite breeding populations and the good tolerance to Psa in their progeny families (Russell Lowe, unpublished data). To have flowers ready for crossing, healthy canes from Female 6 were selected from vines growing under a plastic cover in the field at the Plant and Food Research, Te Puke, New Zealand site. Because a lower germination rate was expected than that of a cross between parents of the same ploidy, and around 2 % of individuals were expected to be tetraploids (McNeilage & Considine, 1989), the number of flowers pollinated was selected to produce the equivalent of 1000 seeds from a diploid x diploid cross. This equated to three shoots each containing 3-6 flowers for each of the four male pollinators. The author placed paper bags over each shoot to be pollinated and stapled them closed around the cane to prevent pollen contamination before the flowers opened in September 2018. In early October, the flowers of Female 6 had opened and were ready for pollination. The paper bags were briefly removed, and all flowers from a shoot were manually pollinated. Pollination was done by transferring stored pollen collected from one of the four male pollinators with a clean paintbrush. Once pollination had occurred, the paper bag was replaced over the shoot and stapled closed. After two weeks, the paper bag was replaced with a plastic mesh bag to protect the developing fruit.

3.3.4 Growing seedlings from the interploid cross

The fruit size on the pollinated plant was monitored every four weeks to collect mature seeds from the cross-pollinated fruit. Once the fruit had reached full size, the average sugar content of other fruit on the vine was monitored each week. When the sugar content had reached a level above 8 %, the fruit were mature enough for harvest. The fruit from pollination was harvested in April 2019 and stored at 2 °C for eight weeks to allow seed development. Once seeds were developed, all the fruit was ripened in a chamber with ethylene for seven days. To extract the seed, the ripe fruit from each shoot was gently mashed with a manual potato masher and placed into a polygalacturonase (LAFAZYM® EXTRACT, Laffort, France) solution at a concentration of 5 g/L of water for two days to degrade the remaining fruit flesh. The seeds were then washed in water, cleaned with a paper towel, and surface sterilised in a 20 % sodium hypochlorite solution. These were kept dry at 4 °C for storage. Seeds to be germinated were sterilised in a 20 % sodium hypochlorite solution for 30 minutes on a shaker, then rinsed with

reverse osmosis water and plated in a sealed Petri dish on filter paper with a solution of 1.25 g/L Captan fungicide (Captan 80W, Nufarm Ltd, NZ). Petri dishes containing seeds spent six weeks at 4 °C for stratification. After stratification, they were moved to a variable temperature and light chamber with cycles of 16 hours light at 24 °C and 8 hours dark at 12 °C. After germination occurred, germinated seeds were counted and individually placed into rockwool cells on an ebb and flow hydroponics table in a glasshouse. Seedlings were grown until they were over 150 mm high. A Previcur® dip was used to prevent fungal damping-off disease.

After eight weeks of seedling growth, samples were taken for ploidy analysis (see next section for methods) on a flow cytometer (described below). Plants found to be diploid (possibly caused by pollen contamination) and triploid were disposed of, leaving only tetraploid individuals. Tetraploid plants were grown on 1.5 m bamboo standards in 8-litre pots in a plastic house. DNA was extracted from these individuals by Slip-stream Automation and tested with *H. lataniae* scale markers from Chapter 2 (described below). All tetraploid plants were planted in the field in 2021. The resistance phenotype will be confirmed in future workstreams at the kiwifruit breeding centre KBC (Te Puke) to verify whether seedlings contained resistance to *H. lataniae*.

3.3.5 Flow cytometry for ploidy analysis of interploid crosses

Measuring the ploidy of seedlings from the interploid cross required analysis of leaf samples with a flow cytometer (Sysmex Ploidy Analyser). Before extraction and analysis occurred, extraction buffer was prepared fresh daily by mixing 0.5 ml of Tween™ 20 in 100 ml of reverse osmosis water and adding 0.5 g of citric acid to 25 ml of Tween solution. The staining solution, which was stable for up to two weeks, was made using 2 mg of powdered DAPI (4',6-diamidino-2-phenylindole) in 1 ml of reverse osmosis water. DAPI solution (250 µl) was added to 200 ml of saturated disodium orthophosphate solution in a light-proof bottle to make the staining solution. The day before making up the staining solution, a 500-ml solution of saturated disodium orthophosphate was prepared by dissolving disodium orthophosphate in warm reverse osmosis water until the saturated disodium orthophosphate would no longer dissolve.

Fresh leaf samples for analysis were collected from the seedlings when they were eight weeks old. As seedlings were arranged in trays in a 96-well plate format, Eppendorf sample tubes were labelled and racked to reflect the 12 x 8 format of the seedling tray. Around 1.5 cm^2 of

leaf material was taken below the third leaf from the growing tip and placed directly into a sample tube. Samples were immediately transferred to the lab for analysis. Analysis of the 1748 samples required a leaf cell extraction method that would maximise sample processing speed and accuracy. For this method, a tool was made from a 6-mm wide stainless steel measuring spatula by shaping and sharpening the flat end to fit the shape of the bottom of an Eppendorf tube. Releasing cell nuclei from leaf samples was achieved by inserting the sharpened spatula into an Eppendorf tube containing the leaf sample and extraction buffer solution. The spatula was pressed to the bottom of the Eppendorf tube around seven times to cut leaves and release cell nuclei. Samples containing buffer and cell nuclei were filtered through 30-µm nylon filters into sample tubes, and the staining solution was added to the filtrate. Samples were run immediately on the flow cytometer according to the operation manual. Samples from parents, run before seedling samples, presented the expected peaks for diploid and tetraploid seedlings. Tetraploid seedlings had a peak in a similar location to their tetraploid male parent, diploid seedlings had a peak similar to their female diploid parent, and peaks for triploids presented in between the tetraploid and diploid peak locations.

3.3.6 DNA collection and marker testing for resistance to *H. lataniae*

To test whether the tetraploid seedlings from the interploid cross contained alleles for resistance to *H. lataniae*, leaves of tetraploid seedlings were collected, and DNA extracted by the company Slipstream Automation. Markers for resistance to *H. lataniae* were selected from the Indel markers described in Chapter 2. Three markers were selected for the DNA marker testing. The three primer pairs were designed to land at 4.412500, 5.661800, and 10.903200 Mb along chromosome 10. The primer sequences were the same as those used in Chapter 2 for primers at that location without the dye labelled M13 tag. The first primer pairs were labelled directly with FAM, NED, or ROX. Dye labelled PCR fragments were amplified by Lena Fraser using the following PCR conditions: 94 °C (5 min), then 30 cycles at 94 °C (30 s) / 56 °C (45 s) / 72 °C (45 s), 8 cycles 94 °C (30 s) / 53 °C (45 s) / 72 °C (45 s), final extension at 72 °C for 10 min. The length of amplified PCR fragments was measured by Kelvina Barrett-Manako using capillary electrophoresis on an ABI Prism® 3100 Genetic Analyzer (Filter Set D, ROX™ GS300HD size standard). Peak data were analysed using GeneMapper™ Software Version 3.0 (Applied Biosystems).

3.3.7 Statistics

The analysis of the difference between means for the percentage of germinated seeds and the percentage of tetraploid individuals produced from the interploid cross was performed using binomial regression analysis in R by the author. Predicted means were generated with the assistance of Catherine McKenzie. The least significant differences in predictions among predicted means were used to test for significance between predicted means.

3.4 Results

3.4.1 The analysis of *H. lataniae* resistance SNPs found in Capture-Seq data

Predicting whether kiwifruit plants have resistance to *H. lataniae* using sequencing data may be much cheaper and faster than phenotyping. This experiment predicted 29 diploid plants, and no tetraploid plants, were likely to be resistant to *H. lataniae* from among the 271 diploid and 132 tetraploid plants that had available Capture-Seq data. The predictions were made in a two-step process. The first step analysed genotype variants in common among five individuals resistant to *H. lataniae*, which were not present in an individual susceptible to *H. lataniae*, Female 7. The analysis found 14597 variant positions in common among the five resistant individuals that were not present in Female 7.

The resistance associated variants (RAVs) from chromosome 10 were filtered based on the 334 variant positions that were present in the Capture-Seq data from 403 individuals. The subset of data followed the pattern of Capture-Seq data read positions with a low number of reads between 4.19 Mb and 9.52 Mb, and a wide gap between 7.26 and 9.52 Mb along chromosome 10 (Figure 3.5). The area with few Capture-Seq reads is likely due to the lower number of genes within the centromere region decreasing the number of Capture-Seq bait targets for exome capture.

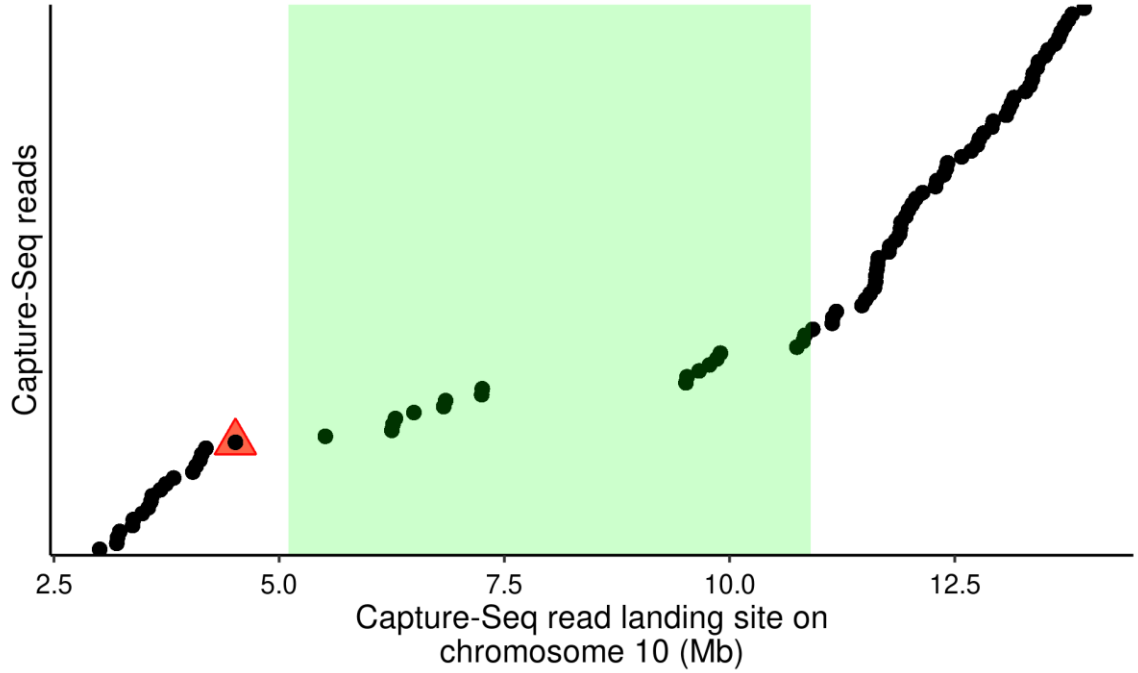


Figure 3.5. Positions of Capture-Seq reads ordered along chromosome 10 *A. chinensis* var. *chinensis*. The red triangle identifies the location of the resistance associated variants at 4.515211 Mb. The vertical green band indicates the fine-mapped region associated with *H. lataniae* resistance from 5.17 Mb to 10.91 Mb.

The number of RAVs was filtered further, finding six RAVs in the Capture-Seq data that had the same nucleotide and position as those from the five resistant individuals which were not in the susceptible Female 7 (Figure 3.5). The six RAVs were at positions 3.196810, 4.515211, 4.515213, and 4.515251, 10.747329 and 13.503752 Mb with variants (C/A, C/T, T/C, G/A, A/T and G/T respectively). The C/A variant at 3.196810 Mb amplified 19 tetraploid individuals. The three variants of C/T, T/C, and G/A at 4.515211, 4.515213, and 4.515251 Mb were found only in diploid individuals from one Capture-Seq read region in 29, 27, and one individual, respectively. The A/T and G/T variants at 10.747329 and 13.503752 Mb were found in 126 and 48 individuals, respectively.

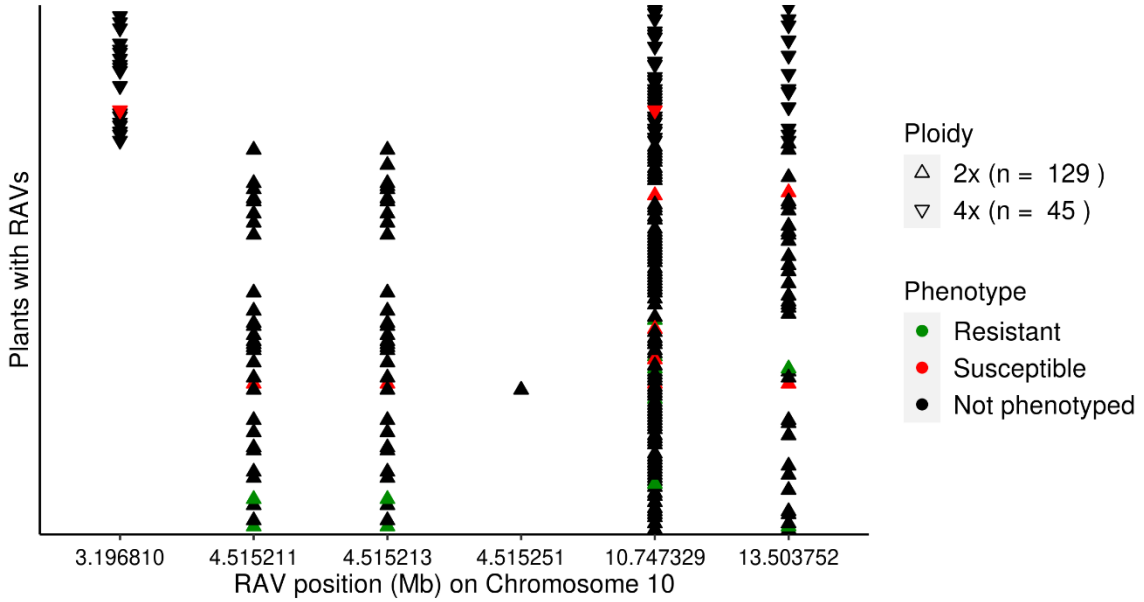


Figure 3.6. Capture-Seq read positions on chromosome 10 for variant loci that were identical among the five resistant *A. chinensis* var. *chinensis* individuals with available whole genome sequence data.

To check that the BWA-Mem variant caller and RAV filtering approach had retrieved data from the BAM files appropriately, the six RAV positions were cross-checked in IGV. BAM file data from the five resistant and one susceptible individual were loaded into IGV to view the raw BAM files of each RAV site (Figure 3.7). As expected, the variants at 4.515211 and 4.515213 Mb were found in all five resistant individuals but not in the susceptible individuals. However, the SNP sites at 3.196810, 10.747329, and 13.503752 showed an unexpected and inconsistent representation in the resistant individuals loaded into IGV. The variant at 3.196810 Mb was found in two of the five resistant individuals. The variants at 10.747329 and 13.503752 Mb were found in four resistant individuals but were absent from the resistant Male 5.

The SNPs at 4.515211 and 4.515213 Mb were just outside the 6.49 Mb region associated with resistance on chromosome 10 from 5.17 to 10.91 Mb identified in Chapter 2 (Figure 3.5).

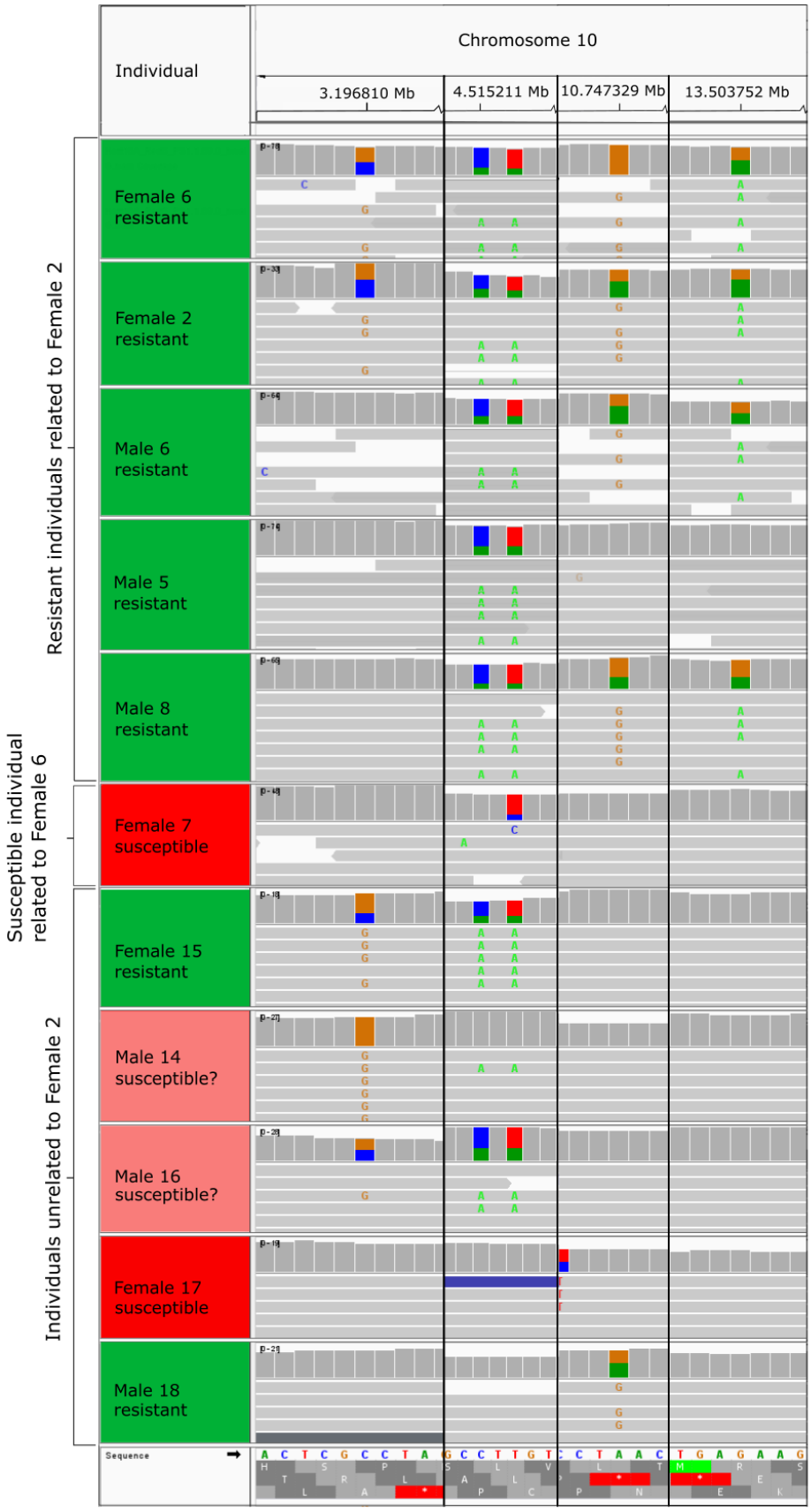


Figure 3.7. Alignment of binary matrix alignment file whole genome sequence (WGS) data in the integrated genome viewer at four variant sites (columns) which were identical among five

individuals of *A. chinensis* var. *chinensis* resistant to *H. lataniae* and not in the susceptible individual related to Female 6. Individuals unrelated to the parent Female 2 with WGS data from the parental population cross are also shown at the same sites. Resistant individuals, with data shown in rows, are coloured green and susceptible individuals are coloured red. The nucleotides identified in the reference genome are shown at the bottom of the figure to the right of the black sequence arrow. For each read, identified by the grey horizontal bands, variants from the reference are shown with the variant's nucleotide letter. The number of reads with the variant are shown as a proportion of total reads in that position at the top of each individual's row.

To test whether the 29 individuals out of the original 403 accurately predicted the *H. lataniae* resistance phenotype, the RAV analysis provided a set of predictions about resistance to *H. lataniae*. To investigate whether this was supported by known pedigree and phenotypes, the available pedigree and phenotype information from individuals contributing to the pedigree was plotted. The pedigree tree supported the predicted phenotype association for 20 of the 29 individuals that had a phenotyped parent which had also been phenotyped as resistant to *H. lataniae* (Figure 3.8). Because the *H. lataniae* resistance phenotype inherited from Female 2 is a dominant trait, each individual with a heterozygous resistant parent would have a 50 % chance of inheriting resistance. Moreover, two of the individuals - 2283 and 2271 - had two parents phenotyped as resistant, which would result in ~75 % chance of an individual inheriting resistance. However, the phenotyping data obtained from Plant and Food Research showed some relationships between phenotyped parents and F_1 plants that would not occur with the dominant resistance trait with which the RAVs were associated (Figure 3.8). This anomaly may have been due to the resistant male parent of individual 2031 carrying a recessive resistance allele from Female 2. This is possible because the parents of 2031 are unrelated to Female 1 and Female 2. The other more likely explanation is phenotyping error. Phenotyping error is likely because individual 2031 inherited the same RAVs as found in Female 2. An interesting result was that individual 2053 is a commercial male polleniser. Confirming the phenotype of this individual would be desirable as it could be used as a parent for breeding pollinators with resistance to *H. lataniae* (Figure 3.8).

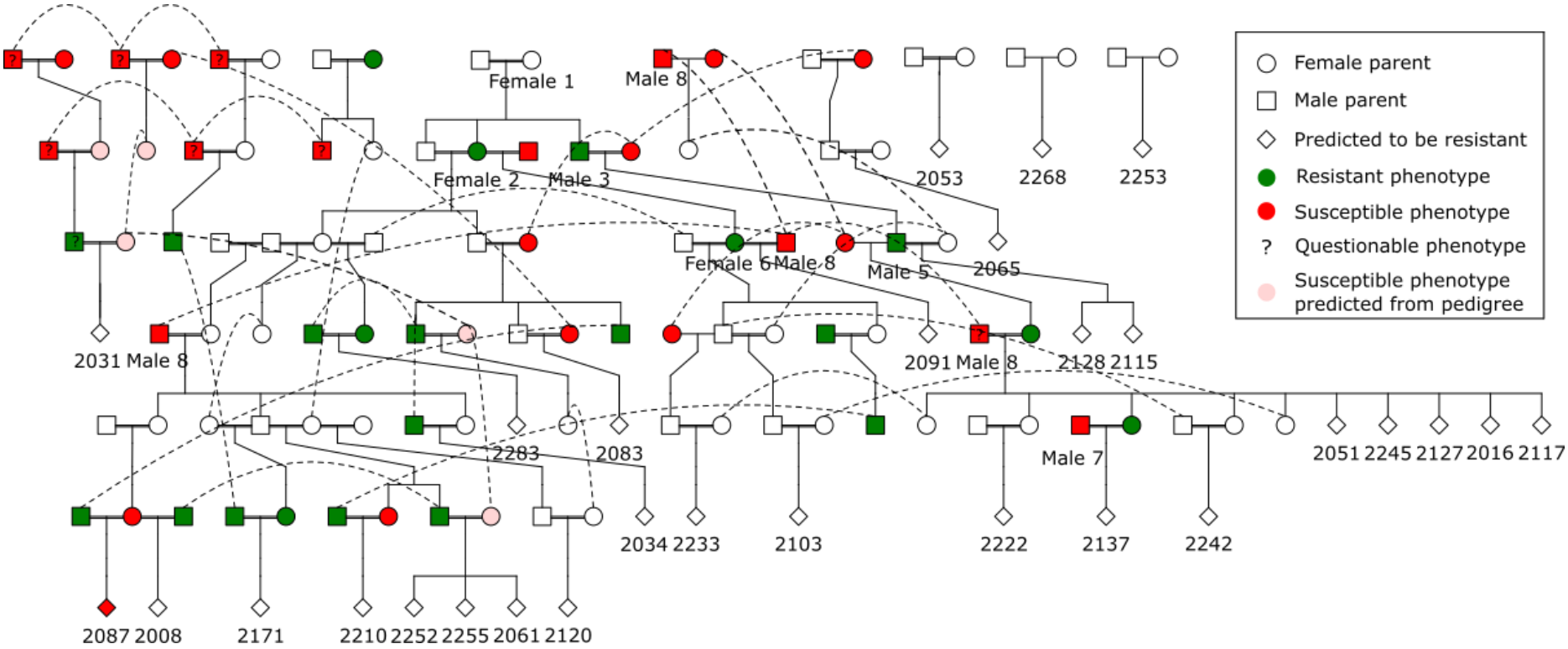


Figure 3.8. The pedigree tree of 29 *A. chinensis* var. *chinensis* individuals that were predicted to be resistant to *H. lataniae* (diamonds). Names of individuals predicted to be resistant are below their diamond-shaped node. Double lines between female and male parents identify reciprocal crossing between the two parents. Lines descending between parents, and occasionally branching out, are the seedlings of the crosses from parents whose lines they intersect. Dashed lines indicate the same individual represented more than once.

3.4.2 Identifying the architecture for resistance to *H. lataniae* in individuals unrelated to the five resistant individuals

A cross of distantly related individuals was phenotyped for resistance using the cut cane method to try to identify resistant individuals with no known relationship to Female 1 or Female 2 (Figure 3.4). However, during the application of scale crawlers to canes to phenotype these four families, it was discovered that the *H. lataniae* scale colony was contaminated with oleander scale (*Aspidiotus nerii*). This led to the loss of phenotype data for 18 of the 27 C1 progeny, 12 of the 16 D1 progeny, 26 of the 31 E1 progeny, and 16 of the 29 F1 progeny. Despite the setback, a subset of these families was phenotyped. Within the C1 family, nine individuals were all phenotyped as resistant, the D1 family had four individuals phenotyped as resistant, the E1 family had five individuals phenotyped with two resistant and three susceptible, and the F1 family had 13 individuals that were all phenotyped as resistant (Figure 3.4). Unfortunately, the fully resistant families meant that using molecular genetics to distinguish resistance alleles based on a population segregating for resistance was impossible. Further, these families were removed from the field following the initial phenotyping, so they could not be re-phenotyped.

The small number of plants phenotyped in the E1 family segregating for resistance/susceptibility to *H. lataniae* meant that the development of markers to target resistance in this family would be difficult. Testing whether there was an error in the original parental phenotyping was done by re-phenotyping the parents, Male 14, Female 15, and Male 16. Re-phenotyping results confirmed the resistance phenotype in Female 15, but Male 16 was found to be resistant, contrary to the original susceptible phenotype call. Unfortunately, phenotype data could not be obtained for Male 16 because the canes used in the cut cane bioassay died before it was completed.

Parents of the C1, D1, E1, and F1 families with available WGS data were loaded into IGV to determine whether the RAVs at sites 4.515211 and 4.515213 Mb, identified in the first experiment, would be able to provide more clues to the lack of phenotypic segregation in these families (Figure 3.4). It was found that Female 15, Male 14, and Male 16 were heterozygous for the RAVs at 4.515211 and 4.515213 Mb. Sharing the same RAVs indicates that the phenotype of Male 14 and Male 16 may have been resistant. Yet, the segregation ratio of 1:3

was not observed, as would be expected if the parents were both heterozygous resistant. One possibility is that the low number of individuals phenotyped missed the susceptible F_1 seedlings. However, it is also likely that the resistant Female 17 was homozygous for resistance. The missing RAVs on the other homologue may be from an old recombination event that broke the association of the resistance gene with the RAVs. If Female 15 was homozygous dominant, it would explain the lack of segregation observed in families C1 and D1 (Figure 3.4).

Female 17 and Male 18 did not contain the RAVs in common with the five resistant individuals. The lack of RAVs indicates that the Female 17 susceptible phenotype is correct with a different allele for resistance in Male 18, or there may have been a break in the association between the RAVs and the gene for resistance to *H. lataniae*. Since there is segregation in the E1 family, it is possible that Female 19 was a phenotyping error and should have been classed as resistant. A homozygous resistant or dominant resistant allele in Female 19 would lead to the G1 family that did not segregate for resistance to *H. lataniae*. Unfortunately, these families were no longer available for re-phenotyping. Otherwise, individuals from the C1 and G1 families could have been crossed with a confirmed susceptible parent such as Male 7 to assess the phenotypic segregation in progeny. Making a second cross may confirm the homozygous dominant nature of Female 15 and get families segregating for resistance to identify the architecture for resistance to *H. lataniae* inherited from Male 18.

As the majority of resistance alleles identified to date are related to those from Female 2 and Female 6, and since Female 6 also has elite fruit characteristics and produces unreduced gametes, Female 6 was used for further breeding to integrate scale resistance alleles into elite parents.

3.4.3 Crossing resistance from diploids to tetraploids

Incorporating resistance to *H. lataniae* from a diploid to elite tetraploid families was achieved by crossing the diploid Female 6 with the tetraploid males Male 20, Male 21, Male 22 and Male 23 (Table 3.1).

Table 3.1. The ploidy of *A. chinensis* var. *chinensis* families from crosses of diploid Female 6 to tetraploid males.

Family	Coefficient of relationship	Germination rate ± SE	Number of seedlings	Percentage of tetraploid seedlings ± SE
Female 6 x Male 20	0.00%	44.11±6.45a	116	81.03 ± 5.71a
Female 6 x Male 21	0.19%	63.54±6.26ab	313	49.74±4.02b
Female 6 x Male 22	0.08%	75.35±6.28bc	355	5.63±1.92c
Female 6 x Male 23	0.06%	79.45±3.42c	964	17.95±1.94d

An additional cross was made with pollen from the diploid red male, Male 24, to incorporate resistance to *H. lataniae* into breeding populations with red-fleshed fruit. The diploid, Female 6, could be crossed with tetraploids because it produces a mix of reduced and unreduced gametes (Yan et al., 1997). The unreduced gametes from Female 6 combine with the normally reduced gametes of the tetraploid males to make tetraploid offspring. This experiment showed that all pollinated fruit developed normally. But the germination rate of seed from tetraploid crosses ranged between 44 % and 79 %, depending on the male polleniser used (Table 3.1). For comparison, fruit set by pollinating the diploid Female 2 with the diploid Male 24 resulted in a germination rate of 98 %.

Seedlings from each family were analysed by flow cytometry to identify the ploidy of seedlings. Results showed a mix of triploid (3x) and tetraploid (4x) individuals (Table 3.2). The triploid seedlings were discarded because their unbalanced chromosome number makes crossing to tetraploid families difficult; therefore, they are unsuitable for commercial breeding pipelines. The percentage of tetraploid individuals from each cross varied widely from 5 % to 81 % for Female 6 x Male 22 and Female 6 x Male 20, respectively (Table 3.2).

Table 3.2. Genotyping results for resistance to *H. lataniae* in diploid and tetraploid families from Female 6.

Families	**Number of seedlings grown**	**Number of genotyped individuals**	**Individuals genotyped as resistant**	**Individuals genotyped as susceptible**	**Percent of resistant individuals**
Female 6 x Male 20	348	39	17	22	44%
Female 6 x Male 21	178	33	19	14	58%
Female 6 x Male 22	321	24	12	12	50%
Female 6 x Male 23	304	135	53	82	39%
Female 6 x diploid Male 24	262	217	91	126	42%

3.4.4 Genotyping *A. chinensis* var. *chinensis* seedlings with markers associated with resistance to *H. lataniae*

To check whether the cross of tetraploid and diploid individuals to Female 6 produced seedlings resistant to *H. lataniae*, three microsatellite markers associated with resistance from Chapter 2 were used. Of the markers designed to land at 4.412500, 5.661800, and 10.903200 Mb along chromosome 10, the only marker to consistently amplify alleles was the one that landed at 10.903200 Mb. The 10.9-Mb marker produced four 151, 157, 158, and 163 bp peaks. The peak of 158 bp was the same as the Red5:PS1.69.0 reference genome and amplified in all samples, while the peak of 151 was also amplified in all samples. The peaks of 163 and 157 were from Female 6 with even segregation between individuals. The peak at 163 was associated with resistance as it matches the allele present in the five resistant individuals with WGS data, while the peak at 157 bp was associated with susceptibility as it is not present in the five resistant individuals with WGS data. Unfortunately, there was a low genotyping success, possibly attributable to poor DNA quality. Of the F1 seedlings for which this marker amplified, the genotyping results showed that ~50 % from each cross contained the 163-bp peak associated with resistance to *H. lataniae* (Table 3.2)

.

3.5 Discussion

H. lataniae is a pest that impacts the production and export of kiwifruit, both in New Zealand and worldwide (M. G. Hill et al., 2006; Hill et al., 2011). The cost of this insect's presence to the kiwifruit industry in New Zealand alone is estimated to be over NZD 77 million per annum (personal communication Cathy McKenna). The damage caused by *H. lataniae,* and a significant portion of the associated costs could be reduced by the development of cultivars resistant to *H. lataniae* attack. Despite several studies investigating the effects of *H. lataniae* feeding on kiwifruit (Blank et al., 1992; Hill & Holmes, 2009; Hill et al., 2005; M. G. Hill et al., 2006; Hill et al., 2009; Hill et al., 2011; Hill et al., 2015; Hill et al., 2016; Morales, 1988), there has been no research that attempts to use this knowledge for the breeding of *H. lataniae* resistance into elite kiwifruit breeding lines. The work described in this chapter focused on advancing the breeding of elite parents with resistance to *H. lataniae* by utilising marker-assisted selection to avoid the expensive and time-consuming process of phenotyping for resistance to *H. lataniae,* seeking out novel alleles for resistance to *H. lataniae*, and breeding resistance into elite parents for integration into commercial breeding programmes.

Gathering phenotype data is widely recognised as a bottleneck in association mapping studies and plant breeding alike (Furbank & Tester, 2011; Minervini et al., 2015; Mir et al., 2019; Song et al., 2021). To identify individuals likely to be carrying resistance alleles without phenotyping, the first experiment conducted here associated phenotype with genotype data from six related individuals, five of which were resistant and one susceptible to *H. lataniae*. Variants of GWAS experiments have been used in the past to associate thousands of loci with hundreds of plant traits (Tian et al., 2020), but these studies typically analysed data within families. This limits the possible number of recombination events within individuals and increases linkage disequilibrium between the target loci and nearby alleles (Flint-Garcia et al., 2003; Gaut & Long, 2003). In contrast, the number of recombination events among related individuals between families and between different generations is greater than that among individuals within a single family since they have more generations to accumulate recombination events between the target loci and neighbouring alleles, thus decreasing linkage disequilibrium (Flint-Garcia et al., 2003; Gaut & Long, 2003). Despite having only six individuals to associate with phenotype in this experiment, these individuals had an increased number of recombination events that occurred between related individuals descending from their common ancestor.

Also, because the region associated with resistance to *H. lataniae* was known from Chapter 2, this was used to narrow the region of the genome used for the association map, thus decreasing the rate of false positive associations with resistance. The two RAVs at 4.515211 and 4.515213 Mb along chromosome 10 were found in 29 of the 403 individuals with Capture-Seq data. These 29 individuals are likely to contain resistance to *H. lataniae*. However, because only six individuals were available to associate the variants with resistance, further support for the *H. lataniae* resistance predictions was sought through the pedigree of individuals predicted to be resistant (Figure 3.8). The likelihood of inheritance could be assessed because resistance to *H. lataniae* was found to be dominant (Chapter 2). However, there were some inconsistencies in the predicted phenotype, or the phenotyping data, since an individual predicted to be resistant had a susceptible phenotype (Figure 3.8). An incorrect phenotype prediction is not unexpected from individuals carrying the RAVs in this location as the RAVs land outside the region associated with resistance found in Chapter 2 (Figure 3.6). The decay in LD between the RAVs' location and the resistance gene may lead to recombination and loss of association between the RAVs and the resistance gene (Flint-Garcia et al., 2003; Gaut & Long, 2003). The breakdown in LD was observed for one plant known to be resistant, which was descended from Female 2 and included in the Capture-Seq data but was missing the RAVs. It is also possible that the phenotype was miscalled. The analysis of SNPs done in experiment one could also have been affected by a high rate of false negatives or a lack of resistance detection. A lack of resistance detection can happen if a Capture-Seq bait fails to land and amplify that position in a sample (Hale et al., 2020). Also, because the read length of Capture-Seq baits is variable, the sample read may not be long enough to capture the desired alleles (Hale et al., 2020).

The utility of the Capture-Seq data was limited by the patchy landing of Capture-Seq baits within the 4.4 and 10.9 Mb resistance locus boundaries found on chromosome 10 in Chapter 2 (Figure 3.6). The low number of bait landing sites within the region of interest would have reduced the chance of identifying RAVs inside the region of interest for resistance to *H. lataniae* because less of the genome is amplified in the region associated with resistance. Conversely, more baits landing in the region of interest would increase the chances that RAVs could be identified with a lower chance of having the association with resistance broken through

recombination. Future work may alleviate the concerns over the prediction accuracy by phenotyping the individuals predicted to be resistant.

The second experiment of this chapter sought to identify alleles for resistance differing from those in Female 2 and Female 6. Diversifying alleles for resistance is a critical goal when breeding for resistance to enable durable resistance to the target pest (McDonald & Linde, 2002; Mundt, 2014). However, before this experiment was run, there was contamination of the *H. lataniae* colony with *A. nerii*. Such contamination can cause false susceptible calls due to the ability of *A. nerii* to infest plants that are resistant to *H. lataniae* without detection (Morales, 1988). *A. nerii* can often be distinguished from *H. lataniae* based on their waxy cap characteristics. But, this becomes increasingly difficult, often requiring the microscopic examination of the posterior end of each insect in parental plants, see Chapter 2 - Figure 2.3, (Morales, 1988). The contamination prevented the complete phenotyping of the four families from this experiment, but a subset of phenotypes from each family was rescued. The families used in this experiment were expected to segregate for resistance from the cross of resistant and susceptible individuals. However, there was no segregation of resistance from the families from Female 15 as expected when a resistant individual is crossed with a susceptible individual. The families from Male 18 showed one segregating family in E1, but members of the G1 family were also all resistant. Several factors could cause the lack of segregation observed in families from Female 15 and Male 18. But because the E1 family was segregating, there is likely a difference in resistance gene copy number between Female 15 and Male 18. The simplest explanation for the lack of susceptible individuals in families from Female 15 would be if Female 15 was homozygous dominant for resistance to *H. lataniae*. A parent homozygous dominant for resistance would convey one copy of the resistance allele to every seedling, resulting in all seedlings being resistant.

An apparent lack of segregation in the C1 and D1 families could also be caused by incorrect phenotype data from the Male 14, Male 16 and Female 19 parents, which may have been resistant despite their original susceptible phenotype call. The incorrect phenotyping hypothesis was supported by the presence of the RAVs found in the first experiment (Figure 3.6) in the susceptible male parents, Male 14 and Male 16. The theory that incorrect phenotyping calls may be affecting the unexpected results was tested by re-phenotyping Female 15 and Male 16 using the cut cane bioassay and an *H. lataniae* scale colony free from

A. nerii. The re-phenotyping showed a resistant phenotype in Male 16, which conflicted with the original phenotype. The cross of a dominant heterozygous resistant individual with another heterozygous resistant individual with a single resistance locus should produce a 3:1 segregation ratio of resistant and susceptible seedlings. But no susceptible seedlings were phenotyped. Two problems were noted when measuring the phenotype on contaminated canes that could have contributed to the lack of susceptible phenotype identification where it should have been apparent. First, the low number of individuals phenotyped may not have included any susceptible F_1 seedlings. Second, the lack of a susceptible phenotype could have also been due to the heavy infestation of *A. nerii* on the bioassay canes. The faster growth rate of *A. nerii* compared to *H. lataniae* (personal observation) may have allowed *A. nerii* to outcompete *H. lataniae* crawlers for settlement sites, resulting in a false resistant phenotype.

The segregation pattern in the E1 and G1 families from the cross of the resistant Male 18 to two susceptible individuals indicated that Male 18 was heterozygous for resistance to *H. lataniae*. The heterozygosity was inferred because the cross of Male 18 to the susceptible Female 17 produced a 1:1 segregation ratio of resistant and susceptible phenotypes. Unfortunately, only four individuals could be confidently phenotyped because of the contamination with *A. nerii*. The cross of Male 18 to Female 19 may have encountered similar issues to the families from Female 15 with an incorrect susceptible phenotype from Female 19. For this cross, Female 19 may be homozygous resistant to *H. lataniae,* as evidenced by the 13 resistant individuals phenotyped. Yet, phenotyping would need to be done accurately, and other crosses made to be sure.

If the families from experiment two were still alive, the families would have been re-phenotyped using a clean colony of *H. lataniae* using the cut cane bioassay. If no segregation was seen among individuals from each family after accurate phenotyping, the next step would have been to pollinate Female 15 with a known susceptible individual. If individuals in the resulting family were all resistant, then Female 15 must be homozygous resistant. Characterising the genomic architecture for resistance alleles from this individual could be done with a back-cross from several F_1 seedlings to the susceptible G1 parent. The resulting F_2B_1 families should allow segregation of the homozygous resistance allele from Female 15 and allow the genotyping of families to identify the architecture for resistance. If any F_2B_1 families

did not have even segregation in resistance and susceptibility, more than one gene must be responsible for resistance from Female 15.

A substantial volume of research has been conducted on the physiology and fruit characteristics of kiwifruit (Cheng et al., 2015; Costa et al., 2017; Datson et al., 2013; D. Li et al., 2010; Martin & Luxton, 2004; Tahir et al., 2019). However, despite resistance to *H. lataniae* being a desirable trait, it is not well integrated into commercial breeding programmes, with a notable absence of resistance to *H. lataniae* in the elite tetraploid *A. chinensis* var. *chinensis* breeding populations. This chapter's third and final experiment focused on resolving this issue by breeding resistance to *H. lataniae* from the diploid Female 6 into four elite tetraploid kiwifruit families and one diploid red family. Crossing a diploid with another diploid is simple because the number of gametes matches to prevent abnormal embryo and endosperm development. However, making a cross between a diploid such as Female 6 and a tetraploid typically results in mismatched gametes and sterile triploid offspring (Seal et al., 2013). Fortunately, prior work by Yan et al. (1997) found unreduced 2*n* gametes in Female 6. The unreduced gametes from Female 6, potentially carrying resistance to *H. lataniae*, were used in this experiment to complement the reduced 2*n* gametes from tetraploid males. This strategy successfully produced four families of tetraploid seedlings (Table 3.1).

Tetraploid seedlings from each cross to tetraploid males, and all the seedlings from the cross to the diploid male were genotyped to identify individuals with resistance to *H. lataniae*. The marker landing at 10.9 Mb amplified both alleles of Female 2 to identify tetraploid individuals that share the allele associated with resistance to *H. lataniae*. The approximate 1:1 segregation ratio of individuals carrying the marker for resistance versus no marker for resistance was expected with the dominant monogenic trait inherited from the heterozygous Female 6. Because this marker was close to the non-recombinant region surrounding the centromere, these markers should be close to 100 % accurate.

Because the marker landing at 10.9 Mb was fully informative for the haplotype from Female 6, and no seedlings had both markers, we can conclude that the 2*n* gametes in Female 6 were not produced by first division restitution. Instead, second division restitution or post-meiotic doubling must have formed the 2*n* gametes in Female 6 (Bastiaanssen, 1998; Yan et al., 1997). Unfortunately, Covid-19 restrictions interfered with amplifying other markers spanning

chromosome 10, which may have identified the mechanism producing 2*n* gametes in Female 6. If enough of these families were genotyped with markers spanning chromosome 10 to detect recombination between markers, the data could identify whether unreduced gametes were produced by second division restitution or post-meiotic restitution. If post-meiotic restitution were producing the unreduced gametes, a recombination event at the same locus in both chromosomes of chromosome 10 would confirm this mechanism over that of second division restitution, which would present a recombination event on one of the chromosomes (Bastiaanssen, 1998).

Integrating resistance alleles for a qualitative trait into elite breeding lines can be enhanced by using resistant parents with resistance on more than one chromosome. This would increase the chances of integrating resistance alleles into the resulting population. The utility of individuals from this experiment to be used as breeding parents will partly depend upon the number of resistance alleles inherited from Female 6. Because either second division restitution or mitotic doubling of gametes produced the unreduced gametes in Female 6, two resistance alleles would have been conferred on two different homologous chromosomes to half of the seedlings, with no resistance alleles in the other half of the seedlings (Brownfield & Köhler, 2011; Yan et al., 1997). The individuals with resistance on two homologous chromosomes offer a significant advantage to breeding programmes because it is more likely that at least one copy of the resistance gene will be passed on to seedlings when using these individuals as parents.

The number of seedlings produced from these crosses is likely dependent on the interaction between the maternal and paternal reproduction structures and the molecular processes pre-germination and post-germination (Seal et al., 2013). Therefore, ensuring the successful integration of resistance alleles from Female 6 into tetraploid families required enough seedlings to retain enough alleles to be representative of paternal contributors to the cross. However, the influence of molecular processes on crosses was reflected in the variable germination rate and the number of tetraploid seedlings produced from each male (Table 3.1). Variation in germination rate was also found by Harvey et al. (1991) in crosses between diploid *A. chinensis* var. *chinensis* females and hexaploid *A. chinensis* var. *deliciosa* males. The differences seen in the diploid tetraploid crosses could be caused by the larger 2*n* gametes restricting normal maternal parental processes (Harvey et al., 1991). However, the pollen used

in the crosses was sourced from frozen samples that were one or two years old. The age of the frozen pollen may have led to variation in quality within the pollen samples that affected germination rates. The variation in germination may also reflect the male's coefficient of relationship with Female 6 (Table 3.1), where males Male 21, Male 22, and Male 23 had some relationship to Female 6 in their pedigree due to an early cross of Female 6 to Male 20 used in the breeding programme. However, the relationship coefficient was based on crosses done with known pedigree at Plant and Food Research New Zealand and does not take into account any genetic relationships of the original progenitors.

The kiwifruit families developed for this study were also designed with resistance to Psa in mind. Psa is a pervasive threat to kiwifruit breeding in New Zealand which can lead to the death of susceptible plants (Dwiartama, 2017). Crossing families with resistance to *H. lataniae* with those carrying Psa resistance was done to prevent the families resistant to *H. lataniae* from dying when planted in the field. The mechanisms for resistance to Psa currently remain unknown, but a QTL for Psa resistance has been found (Chapter 5) by Tahir et al. (2019). However, conventional breeding methods such as family-based selection can identify parents that contribute the greatest resistance to their progeny. For this study, tetraploid males with the best performance against Psa in the commercial gold breeding programme were selected with assistance from commercial kiwifruit breeder Russell Lowe. The tolerance to Psa gained from their alleles is expected to help the survival of families crossed to the Psa susceptible Female 6. Moreover, there is increased field survival of autotetraploids (personal communication Canhong Chen) derived by induced chromosome doubling of the diploid cultivar 'Hort16A' (Wu et al., 2014). The tetraploid families created in this experiment, which may stack resistance to *H. lataniae* with tolerance to Psa, now require sufficient exposure to Psa in the field to identify individuals tolerant to Psa.

In this experiment it is interesting that there was no resistance to *H. lataniae* in three of the four elite males used in the cross to Female 6, despite having Female 6 in their pedigree (Table 3.1). The lack of resistance found in elite populations tolerant to Psa may signify that the alleles for resistance to *H. lataniae* are linked to deleterious traits. It is hoped that the cell size increases observed in tetraploid *A. chinensis* var. *chinensis* (Wu et al., 2012) do not interfere with the resistance mechanism of the wound periderm observed in 'Hort16A' (Hill et al., 2011). Thus, future studies will need to confirm the resistance to *H. lataniae* indicated by markers by

phenotyping the five families generated in this experiment to confirm the association of *H. lataniae* resistance with molecular marker results. Close attention will need to be paid when phenotyping these families since the variance in gene dosage between individuals with two copies of the resistance allele may alter the phenotype compared to a single dose of the resistance gene (Birchler & Veitia, 2014; Chen, 2007). However, the likelihood of linkage between Psa tolerance and *H. lataniae* resistance is low given the polygenic nature of Psa tolerance reported in Chapter 5. It is more likely that a genetic bottleneck in families using Female 6 as a parent lost the alleles for tolerance to *H. lataniae* (personal communication Paul Datson). Therefore, subsequent generations did not contain tolerance to *H. lataniae*.

Having larger populations will increase the diversity of potentially valuable alleles from these populations, decrease the effects of genetic drift and reduce the risk of inbreeding (Conolly et al., 2008). The diploid family with 217 individuals and the tetraploid family with 82 individuals developed here should maximise the capture of parental alleles for further breeding. However, the families with 12, 14 and 22 resistant individuals would benefit from re-genotyping the individuals in which genotyping failed to find more resistant individuals to add to the populations.

The experiments done in this chapter will enhance the breeding of resistance to *H. lataniae* into diploid and tetraploid populations alike. Breeding was assisted by providing methods to detect resistance faster and cheaper than phenotyping with Capture-Seq data. The incorporation of resistance to *H. lataniae* from diploid to elite tetraploid *A. chinensis* var. *chinensis* families with Psa tolerance will allow the integration of resistance to *H. lataniae* into new kiwifruit cultivars. This will minimise the costs associated with *H. lataniae*, creating a more efficient and environmentally friendly form of fruit production benefiting organic growers and conventional producers of kiwifruit alike.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the student and the student's main supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the student's contribution as indicated below in the Statement of Originality.

Student name:	Casey Flay
Name and title of main supervisor:	Vaughan Symonds
In which chapter is the manuscript/published work?	Chapter 5
What percentage of the manuscript/published work was contributed by the student?	100% unless otherwise stated in text

Describe the contribution that the student has made to the manuscript/published work:

Some resources were existing at Plant and Food Research which the student used to start their work.

Please select one of the following three options:

- ○ The manuscript/published work is published or in press
  Please provide the full reference of the research output:
- ○ The manuscript is currently under review for publication
  Please provide the name of the journal:
- ☑ It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Student's signature:	Flay	Main supervisor's signature:	Vaughan Symonds

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

Doctoral Research Committee
July 2022

4 QTL mapping kiwifruit (*Actinidia chinensis* var. *chinensis*) for resistance to the brown-headed leafroller (*Ctenopseustis obliquana*)

4.1 Abstract

The brown-headed leafroller moth (*Ctenopseustis obliquana*) causes significant problems to the NZD $2.7 billion per annum New Zealand kiwifruit (*Actinidia chinensis*) industry by directly damaging fruit and causing export quarantine issues. Kiwifruit cultivars resistant to *C. obliquana* could overcome problems of both the damage caused by *C. obliquana* and the presence of the insects on fruit. However, the genetic architecture for resistance to *C. obliquana* has not been assessed in *A. chinensis*, making its inclusion in breeding programmes difficult. In this chapter, phenotypic data were collected from a family of *A. chinensis* var. *chinensis* individuals that had one parent resistant and the other susceptible to *C. obliquana.* Linkage maps were assembled using high-density genotyping by sequencing (GBS) data and a quantitative trait loci (QTL) mapping approach was taken to identify regions of the genome associated with resistance. However, no significant QTL were detected, likely due to issues with the bioassay.

4.2 Introduction

The brown-headed leafroller moth (*Ctenopseustis obliquana*) is a highly polyphagous pest of kiwifruit (*A. chinensis*) in New Zealand. The larvae of *C. obliquana* feed on the fruit and leaves of kiwifruit, causing up to 33 % of fruit to be damaged if no control measures are applied (McKenna & Stevens, 2007). The damage caused to fruit is an issue for growers, but as *C. obliquana* is endemic to New Zealand, its presence on exported kiwifruit is a significant market access issue. The Zespri KiwiGreen® spray programme prevents numbers of *C. obliquana* from reaching damaging levels and reduces the risk of finding infested fruit post-harvest. Despite the control measures, each fruit still needs to be inspected to determine if it is clear of *C. obliquana* before being exported.

Breeding plants resistant to *C. obliquana* will be the most cost-effective and environmentally friendly route to control *C. obliquana* on kiwifruit (Brown & Rant, 2013). However, there is no resistance to *C. obliquana* in the leading gold-fleshed cultivar 'Zesy002' (personal communication; Cathy McKenna). The lack of *C. obliquana* resistance in 'Zesy002' may be attributed to commercial breeding operations being unable to justify the high cost of phenotyping for resistance to *C. obliquana*. The high cost of phenotyping for resistance to *C. obliquana* results from the time-consuming and labour-intensive leaf disk bioassay performed to assess larvae fitness on kiwifruit leaves (Wearing et al., 2003). The leaf disk bioassay is used because adult female *C. obliquana* do not exhibit egg-laying preference between resistant and susceptible individuals that would allow a cheaper host preference test to be completed (Wearing et al., 2003). Instead, the resistance of kiwifruit to *C. obliquana* takes the form of reducing larval feeding or growth on leaves.

The distribution of host plant resistance within a family segregating for resistance often depends on whether the resistance is qualitative or quantitative (Willocquet et al., 2017). Qualitative resistance is usually conferred by the products of a single gene directly or indirectly by recognising the pests avirulence gene products (Keller et al., 2007; Marone et al., 2013). Breeding for these single genes allows resistance to be bred into a population quickly and easily, enabling its widespread utilisation. Breeding for single genes also has the advantage of conferring strong resistance to a genotype (Willocquet et al., 2017). However, single-gene traits for strong resistance exert intense selection pressure on the pest or pathogen they

target, often leading to the development of resistance to the resistance gene (McDonald & Linde, 2002; Zhan et al., 2014).

On the other hand, quantitative resistance is often driven by several minor-effect resistance genes. The resistance expressed by the minor genes is not usually as strong as monogenic resistance. However, because more genes influence different pathways to confer resistance, they have greater durability against the evolution of insect resistance (Willocquet et al., 2017). While quantitative resistance can convey a durability advantage, conventional breeding for quantitative resistance is more costly and time-consuming than breeding single gene traits. The increased cost of quantitative resistance breeding often results from the increased phenotyping difficulty and because the genes responsible for resistance are often distributed throughout the genome (Poland et al., 2009). Molecular techniques show the most promise for breeding quantitative resistance into cultivars due to the ability of markers to identify plants with multiple resistance genes, which are challenging to separate phenotypically. These markers can then be utilised in conventional breeding programmes as selection criteria to incorporate resistance markers into a single cultivar (Keller et al., 2007).

The work presented here analyses an *A. chinensis* var. *chinensis* family known as B1, which was also utilised in Chapter 2 to study resistance to the armoured scale insect (*H. lataniae*). The B1 family used for this experiment was large, with three clonal replicates, each comprising of 190 individuals for 570 plants in total. The female parent of the B1 family is resistant to *C. obliquana* (McKenna & Stevens, 2007), and the male parent is susceptible (unpublished data C. McKenna). Loci conferring resistance were investigated using a QTL mapping approach.

4.3 Methods

4.3.1 B1 family of *A. chinensis* var. *chinensis* plants

For the *C. obliquana* resistance bioassay, plants were sourced from the *A. chinensis* var. *chinensis* B1 family used in Chapter 2. The B1 family had 190 plants of sufficient size to hold enough leaves for the *C. obliquana* bioassay. The B1 family was developed from the *H. lataniae* resistant Female 6 as mother and the susceptible Male 7 as father (Figure 3.2). The plants making up the population were over 1.0 m high and grown in 4.5-L planter bags. Three clonal plants were derived from each original B1 plant. All plant replicates were fertilised with slow-release Osmocote® Exact Standard fertiliser tablets, and 'Yates® Thrive all-purpose liquid plant food' was applied at three-month intervals at the recommended dose. Each family replicate was kept in different conditions to identify whether these conditions affected the expression of resistance to *C. obliquana*. Two replicates were maintained at the Plant and Food Research, Te Puke site, with one replicate in an actively vented glasshouse and the other in a passively vented plastic house. The third replicate was maintained in a passively vented plastic house at Plant and Food Research, Palmerston North, New Zealand.

4.3.2 Phenotyping for resistance to larval *C. obliquana* in an F_1 population of *A. chinensis* var. *chinensis*

The life stage of *C. obliquana* that causes damage to kiwifruit plants is its caterpillar-like larvae. These 1-mm long larvae hatch from an egg and find a place to feed. Since the feeding and development of larvae are closely related to the characteristics of the leaves they are feeding on (Wearing et al., 2003), the variability in leaf age was kept to a minimum by defoliating each plant 6-16 weeks before the commencement of phenotyping. When the leaves were growing back after defoliation, an infestation of the two-spotted mite (*Tetranychus urticae*) was detected on each replicate. Due to the residual nature of pesticides and the susceptibility of *C. obliquana* larvae to pesticides, the two-spotted mites were controlled with 4000 predatory mites (*Phytoseiulus persimilis*), supplied by Bioforce Limited. The predatory mites suppressed the two-spotted mite populations on the new leaves, but some mites remained.

The effects of different environments on the expression of resistance in plants from the B1 family were considered by housing each of the three clonal replicates of 190 plants in a

different environment for at least eight weeks before harvesting leaves for the experiment (described in Chapter 2). To estimate the environmental variance on replicates, two clonal individuals of the resistant parent, Female 6, were used as controls within each replicate. If more plants of the parent Female 6 or other resistant individuals had been available, they would have been used within each row of plants in the glasshouse to account for within-replicate environmental variation. However, more resistant controls would need to have been propagated by tissue culture if they were to have been used, because the resistant parent was not available in the field. The propagation of plants from tissue culture to a similar size to the trial plants would have taken two to three years, which was too long to wait before commencing this study. This was also the case for the susceptible parent Male 7, so leaves from the cultivar 'Hayward' were sourced from the field as a replacement susceptible control.

Because the effect of plant resistance on insect health or survival from Female 6 was unknown, several parameters of insect health were measured during the three-week leaf disk bioassay. On the day of the experiment, ten leaf disks of 20 and 35 mm diameter were cut from two leaves from each individual to be phenotyped. The smaller leaf disk was placed on top of the larger disk, and both leaf disks were placed on top of a 60-mm diameter filter paper. The filter paper contained 150 µl of water in a 100-mm Petri dish without vents. Two first instar *C. obliquana* larvae were gathered from a container for placement on leaf disks. The delicate larvae were transferred without contact using a fine paintbrush to lift the silk thread that larvae use to suspend themselves when hanging from the hatching container's lip. The Petri dishes were sealed with parafilm and transferred to a 22-°C growth chamber for seven days. The growth chamber was lit with mild intensity fluorescent bulbs that remained on for the duration of the bioassay. Petri dishes were not randomised, to minimise the disturbance of the larvae within each dish. If two larvae remained alive from the first week, one larva was retained while the other was discarded. After each seven-day period, for the three-week duration of the bioassay, larval survival was assessed by visually inspecting each of the ten replicate Petri dishes. A score out of ten was given based on the number of Petri dishes with surviving individuals, e.g., if four Petri dishes had surviving individuals, a score of four would be given for that plant genotype at that week of recording. Because leaf disks do not remain fresh for more than two weeks in the sealed Petri dishes, swapping larvae to fresh leaves and moistened filter paper was also completed every week when larval survival was assessed. All the larvae

remaining on leaves after 21 days were placed into Eppendorf tubes with 70 % ethanol for storage. Head width was measured, and the total weight after 4 hours of oven-drying at 60 °C was recorded for all surviving individuals from each plant.

The leaf disk bioassay produced data on six attributes measuring insect health and survival in plants from the B1 family. These included insect survival at 7, 14, and 21 days, average weight, average larval head width, and average larval instar. The average weight, head width and instar were calculated by averaging data from larvae surviving within the ten replicates from each plant that survived until day 21. Because the effect of plant resistance on these traits is unknown, each trait was analysed as a separate phenotype.

Due to the supply of *C. obliquana* 1st instar larvae and the workload involved in setting up the bioassay, only ten replicate leaf disks from 32 plants could be set up per day for three days per week with a four-person team. The 96 individuals bioassayed each week were maintained for the assay duration, and then another set of 94 individuals started after the completion of the first set. The first replicate was phenotyped from March to April 2019, the second from September to October 2019, and the third from October to November 2019. The three replicates were analysed together, averaged, and analysed as individual replicates for QTL mapping.

The resistant parent (Female 6) was included as a resistant control. Four resistant controls were sourced for each replicate, two of which were bioassayed, along with the trial plants each week the bioassays were initiated. In the absence of the resistant father of the B1 population, the cultivar 'Hayward' was used as a susceptible control in each of the three replicates. Also, as control plants were not available in the glasshouse, leaf material from the susceptible control was obtained from the field to determine whether *C. obliquana* larvae behaved as expected between controls for each replicate. Therefore, the susceptible controls could not be used to measure the environmental impact on the plant resistance phenotype. Instead, this control was intended to assess the fitness of *C. obliquana* larvae.

4.3.3 Genotyping

The genotype data used in this study was produced using a GBS protocol modified by Tahir et al. (2019). However, their study focused on resistance to *Pseudomonas syringae* pv. *actinidiae* (Psa) in an expanded set of the B1 population. The experiment described here re-analysed the

raw GBS genotype data using a bioinformatics pipeline widely applied to the analysis of next-generation sequencing data. Genotypes were aligned to a reference genome, based on the susceptible male parent of the B1 population. This reference was also chosen because it was compiled with long-read (Hi-C) and short-read data from Illumina sequencing to enhance alignment accuracy.

GBS data analysis started by removing the GBS restriction enzyme residues from (BamHI) at the cut site. The data from each plate were then de-multiplexed using the "ea-utils" toolkit fastq-multix (Aronesty, 2013). Filtering of SNPs and indels was completed with "BCFtools view" (Li, 2011) using the "include" expression to include alleles with a minor allele frequency of >0.05, an average depth >16 and a quality score >20. Data were piped to "BCFtools filter" with the exclude expression excluding all genotypes with samples missing. Included in the "BCFtools filter" was an option to exclude SNP calls from 9 bp around indels. "BCFtools stats" showed that filtering with these options yielded 2298 SNPs and 146 indels. Two thousand and sixty-three of those were multi-allelic sites, with 1994 multi-allelic SNP sites. The unfiltered variant call file (VCF) had over one million SNPs and 28366 indels.

The number of incorrect variant calls was reduced by converting the filtered VCF file to a HapMap format text file using Tassel 5 (Gibbs et al., 2003). An example of data transformation from the HapMap format to the JoinMap format is provided in Table 4.1. The HapMap format text file was opened in Microsoft Excel, and columns were made for the frequency of each allele using a "countif" function, including each sample in the population. Additional columns were made with nested "if" functions to group polymorphic markers in the parents. Loci with non-polymorphic sites in both parents were removed. Alleles with a frequency below 10 % were assumed to be caused by sequencing error and were re-coded as missing data. Loci heterozygous in the mother Female 6 and homozygous in the father Male 7 were female informative. Female informative heterozygous markers were assigned as <lm>, and homozygous female markers were assigned as <ll>. Similarly, heterozygous loci in the male parent and homozygous in the female parent were male informative. Male informative heterozygous alleles were assigned the code <np> while homozygous alleles were assigned <nn>. Loci with a population distribution consistent with a null allele in one of the parents were called heterozygous for the parent carrying the null allele and coded accordingly. If a locus was heterozygous in both parents with an allele ratio of 1:2:1, it was called partially informative.

The partially informative loci with 50 % heterozygous genotypes were coded as <hk>, and the two homozygous genotypes with a 25 % frequency in the population were coded as <hh> and <kk>. Another set of partially informative loci presented a 1:2:1 pattern of segregation. If either of the parents had a homozygous genotype, it was concluded that the homozygous parent was heterozygous with a null allele. This miscoding can happen because null alleles do not appear in the HapMap file format and are instead presented as homozygous alleles. In these partially informative loci, the heterozygous allele was coded as <hk>, the informative allele with a null was coded as <h->, and the remaining allele was coded with a <k-> (Table 4.1).

Table 4.1. An example of *A. chinensis* var. *chinensis* allele frequencies used to transform genotype data from a HapMap format file to JoinMap format. Parent and population genotypes were retrieved from the HapMap format output. HapMap coding for alleles is shown in square braces [] and JoinMap coding for the CP population type in less than and greater than symbols <>.

Parent alleles	Parent genotype	Expected alleles in B1 population	Expected allele frequency	Allele coding, [Hapmap], <JoinMap>		
a- x tt	[A] x [T]	at + t-	at = 50%, t- = 50%	at = [W] = <nn>	t- = [T] = <np>	
at x tt	[W] x [T]	at + tt	at = 50%, tt = 50%	at= [W] = <lm>	tt = [T] = <ll>	
aa x at	[A] x [W]	aa + at	aa = 50%, at = 50%	aa = [A] = <nn>	at = [W] = <np>	
at x at	[W] x [W]	at + tt + aa	at = 50%, tt = 25%, aa = 25%	at = [W] = <hk>	aa = [A] = <hh>	tt = [T] = <kk>
a- x at	[A] x [W]	aa+a- + at + t-	aa+a- = 50%, at=25%, t- = 25%	aa+a- = [A] = <h->	at = [W] = <hk>	t- = [T] = <k->
a- x t-	[A] x [T]	at+a- + t- + --	at = 25%, a- = 25%, t- = 25%, -- = 25%	at = [W]	a- = [A]	t- = [T] = <k->

Linkage maps from the cleaned GBS genotypes were made in JoinMap 5. Markers that were >95 % identical to others were excluded to process only unique loci. Linkage groups were assembled using regression mapping, and markers with a LOD of under one were removed. Markers within the suspect linkages tab were removed with a chi-square test result of over six. Data were exported as separate LOC files for the parents and the consensus map data. Data from JoinMap 5 were loaded into R for QTL mapping. Before mapping was started, the genetic map was checked and showed some regions with a higher marker density than others, but with good coverage of most of the chromosomes when measured in cM (Figure 4.1, Table 4.2). The pairwise recombination fraction and LOD scores were checked but did not show significant misalignment of markers between chromosomes (Figure 4.2).

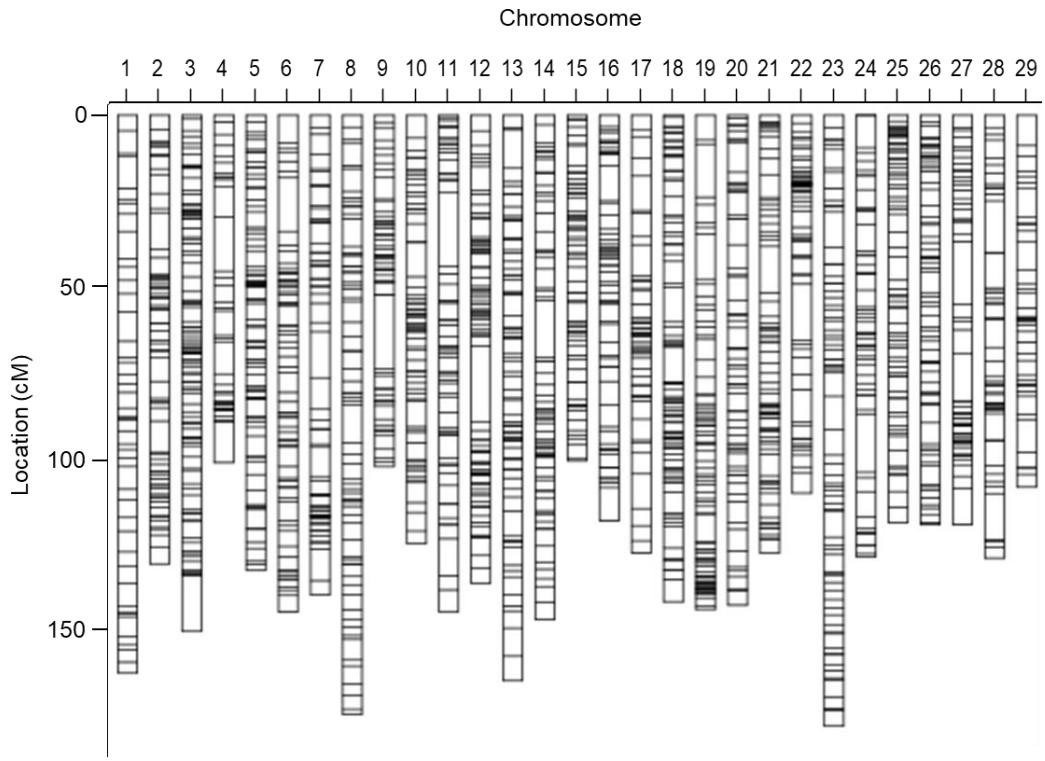


Figure 4.1. Genetic map of GBS markers from the B1 population. The marker positions (cM) are shown by the black bands within each chromosome.

Table 4.2. Distribution of the 2773 genotyping by sequencing markers across chromosomes from the B1 population of *A. chinensis* var. *chinensis* aligned against the Male 5 *A. chinensis* var. *chinensis* reference genome. The number of markers and average distance between markers (cM) are displayed.

Chromosome	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Number of markers	111	91	106	75	98	108	75	131	93	103	86	128	109	99	105
Average distance between markers (cM)	2.1	2.1	2	2.1	2	2.3	2.8	2.5	1.6	2	2.3	1.8	2.3	1.9	2

Chromosome	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**
Number of markers	70	89	129	110	87	120	106	97	77	68	112	67	57	66
Average distance between markers (cM)	2.4	2.8	1.9	2.2	2.3	1.7	1.9	2.6	2.3	2.3	1.8	2.5	3	2.3

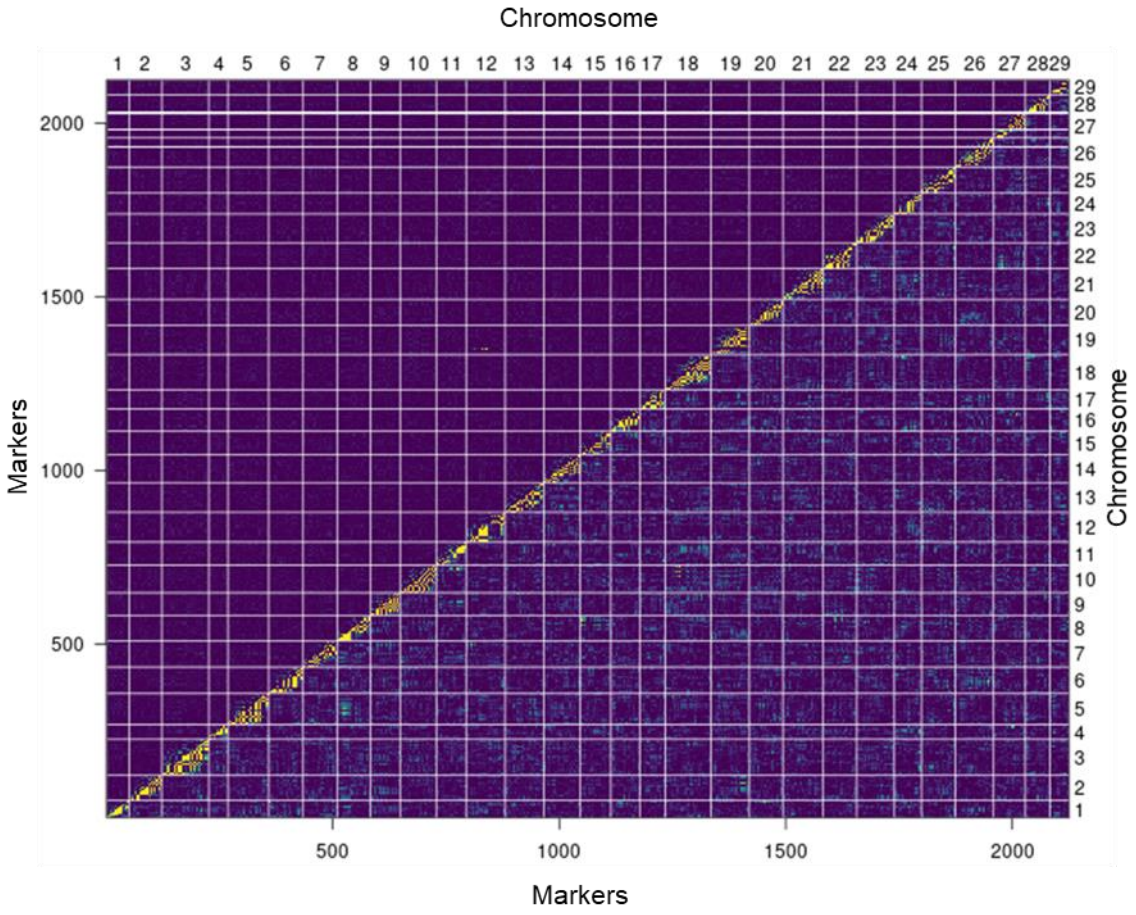


Figure 4.2. Pairwise recombination fraction and LOD scores from the B1 population of *A. chinensis* var. *chinensis*. The estimated recombination fractions (upper-left triangle) of each square and LOD scores (lower right triangle) for all pairs of markers. Yellow indicates linked (large LOD score or small recombination fraction) and blue indicates not linked (small LOD score or large recombination fraction).

4.3.4 QTL mapping

Before being analysed, data from each of the six *C. obliquana* larval performance parameters measured were box-cox transformed to fit a normal model. After data transformation, QTL mapping was completed in an R package called R/qtl (Broman et al., 2003). Data were imported as a back-cross cross-type to reflect the heterozygous nature of the F_1 cross. The function Scanone was run using the interval mapping QTL model (Lander and Botstein 1989). Where a Shapiro-Wilk test for normality was greater than $p = 0.05$, the normal model was used, and

where it was less than $p = 0.05$, a non-parametric model was used. The EM algorithm was applied to both model types. A marker regression model was performed simultaneously with the QTL map to compare the results of the two models. R/qtl uses a hidden Markov model to calculate QTL genotype probabilities. The R/qtl program also simulates the genotype distribution and calculates the most likely sequence of underlying genotypes to generate a threshold for significance. Data quality checking identified that many of the markers were not ordered correctly, leading to many markers having a higher recombination frequency than expected. Re-ordering of markers was done with ripple and switch.order for each chromosome. However, re-ordering with ripple and switch.order resulted in the genetic map expanding to greater than 20000 cM from the original 160 cM on chromosome one. Because of this, the original ordering from JoinMap was used.

Because of the weak effect of the resistance bioassay, each of the six different phenotypes was analysed in three different ways. Firstly, the three replicates were analysed separately, then the three replicates were analysed together, and the final analysis averaged the phenotype data for the QTL analysis. Because each replicate was a clonal copy, the genotype data remained the same for each replicate. An empirical significance threshold was established for each analysis at $p = 0.05$ using a simulation with 1000 permutations of genotypes. The data alignment and R/qtl methodology were tested using the *H. lataniae* phenotype from Chapter 2. The *H. lataniae* phenotype was analysed alongside the *C. obliquana* larval phenotypes in R/qtl.

4.4 Results

Resistance of *A. chinensis* var. *chinensis* to larval *C. obliquana* was measured using six parameters to assess larval fitness. The six phenotypes were associated with GBS markers from the B1 population of *A. chinensis* var. *chinensis* to identify QTL at sites with a significant association. However, no reliable QTL were associated with any of the phenotypic attributes recorded (Table 4.3, Appendix, Figures 6.1 to 6.6). A combination of an imprecise bioassay, a reduction in the number of individuals phenotyped for each replicate, and genotyping error are thought to have been the main contributors to the lack of association of genomic markers with phenotype.

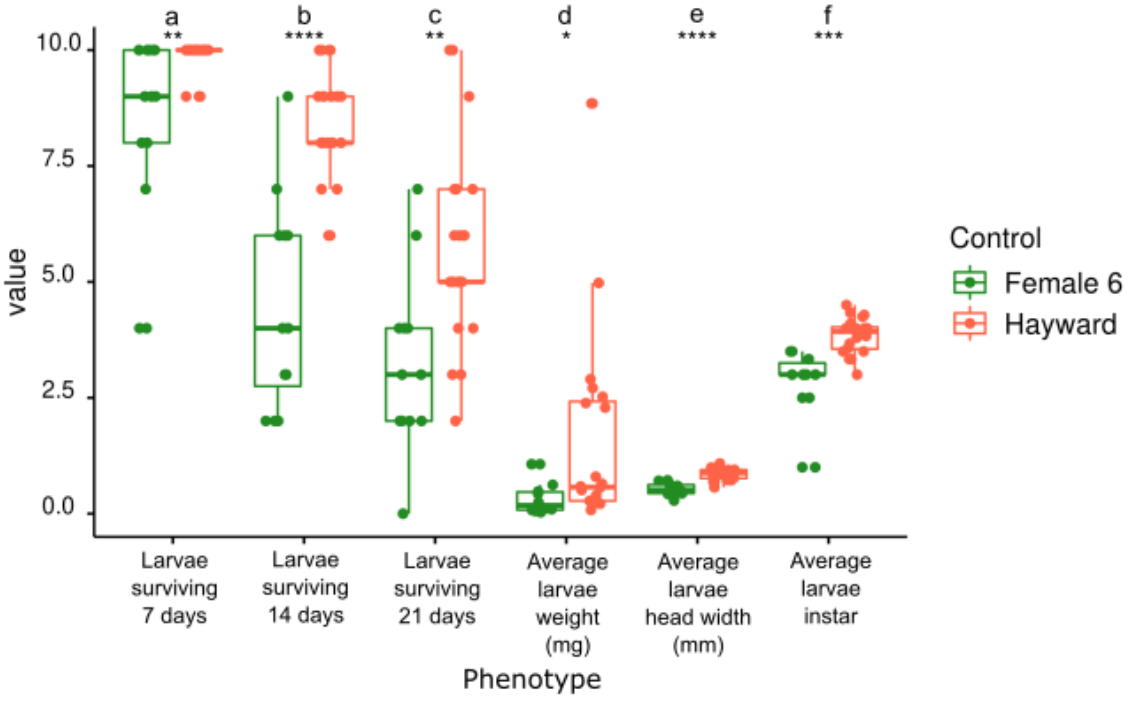


Figure 4.3. Health attributes of larval *C. obliquana* on the resistant *A. chinensis* var. *chinensis* parent, Female 6, and a susceptible *A. chinensis* var. *deliciosa* cultivar, 'Hayward'. The number of larvae surviving at 7, 14, and 21 days was significantly higher for the susceptible 'Hayward' cultivar than for the resistant parent; Female 6; with Kurskal-Wallis test values of $p = 0.01$, $p = 0.0001$, and $p = 0.01$, respectively. The average larval weight, average head width and average larval instar were significantly higher in the susceptible 'Hayward' than the resistant Female 6 control with KWT values of $p = 0.05$, $p = 0.0001$ and $p = 0.001$, respectively. Phenotypes under letters a, b, and c use y-axis values that indicate the number of surviving larvae out of the ten

individuals applied to leaves of control plants at 7, 14 and 21 days. The phenotypes of average larval weight (d), average larval head width (e), and average larval instar (f) have y-axis values that indicate the average among all individuals surviving 21 days on each control parent. One, two, three, and four asterisks indicate $p = 0.05$, $p = 0.01$, $p = 0.001$ and $p = 0.0001$ respectively.

4.4.1 Phenotyping

Because no QTL were associated with any of the *C. obliquana* phenotypes, the bioassay accuracy was suspected as the main source of error. Confidence that variance from the bioassay was affecting QTL detection was increased when no QTL were associated with any *C. obliquana* larval fitness attributes when each replicate was analysed individually. The genotypic data were unlikely to have affected this as the phenotypic data for resistance to *H. lataniae*, run with the *C. obliquana* phenotypes in R/qtl, identified QTL over the significance threshold. Also, the resistant control plant Female 6 showed substantial phenotypic variance for resistance, which was not consistent with previous work. Physiological differences between leaves within the control may have contributed to this variance.

Unfortunately, the resistance phenotype also was confounded with substantial variance in the measured *C. obliquana* larval fitness parameters on the susceptible 'Hayward' control (Figure 4.3). The phenotypic variance on the susceptible control strongly indicated that the *C. obliquana* larvae used in this study had significant variance between individuals. There was also a substantial overlap between the susceptible and resistant phenotypes for control plants (Figure 4.3).

The confounded variance in plant resistance and *C. obliquana* larvae likely prevented QTL detection (Table 4.3). If each phenotype was consistent, e.g., if each replicate showed the same ranking of plants, there should have been an increase in LOD scores when replicates were run together in the analysis. However, due to the weak effect of resistance observed for this bioassay, the environmental effects between each replicate may have caused a significant re-ranking of phenotypes. The variance between replicates, shown in Figure 4.4, highlights the high variance in the average phenotypic value of individual replicates. The low power to detect QTL would have been exacerbated due to incomplete phenotype data for each replicate (Figure 4.4), the variability in plant survival and the supply of *C. obliquana* larvae.

Unfortunately, because leaves for the susceptible control were obtained from the field instead of from control plants within the glasshouse with the trial plants, the insect performance on control plants does not indicate whether the environment influenced plant susceptibility.

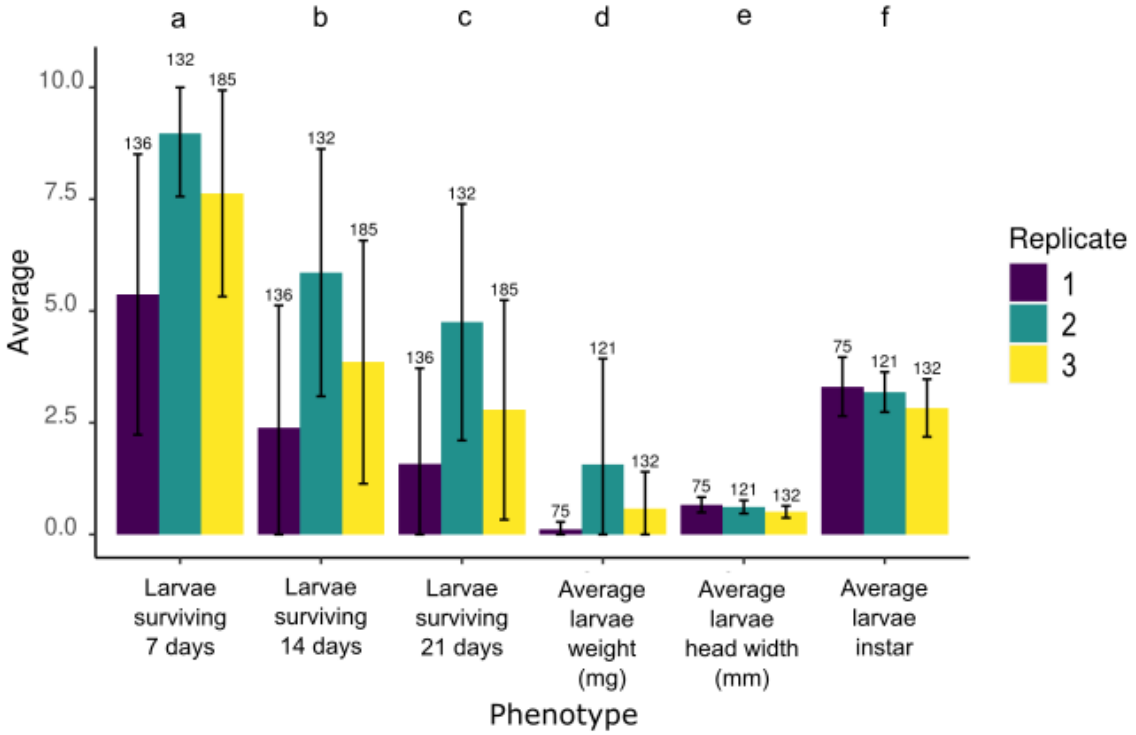


Figure 4.4. Health attributes of *C. obliquana* larvae on *A. chinensis* var. *chinensis* between replicates of the B1 population. Phenotypes were collected from the number of individuals presented above the standard deviation bars. Phenotypes under letters a, b, and c use y axis values that indicates the number of surviving larvae out of ten between replicates of the B1 population at 7, 14 and 21 days. The average number of individuals surviving on replicate two was higher than that of replicates one and three on all measurement days. However, there were no significant differences between replicate means of larvae surviving 7, 14 or 21 days KWT $p = 0.05$. The phenotypes of average larval weight (mg) (d), average larval head width (mm) (e), and average larval instar (f) have y axis values that indicate the average among all surviving individuals from each replicate after 21 days of that phenotype. There were no significant differences between replicate means of average weight, average head width or average instar KWT $p = 0.05$.

4.4.2 Mapping QTL for resistance to larval *C. obliquana* using GBS genotypes from the B1 family of *A. chinensis* var. *chinensis*

To search for QTL in *A. chinensis* var. *chinensis* associated with resistance to *C. obliquana*, markers from GBS genotype data were sought from the B1 family. The analysis found 2773 markers which were used to map QTL for the six *C. obliquana* larval fitness parameters as separate phenotypes. However, only two QTL were detected above the threshold of significance. The QTL above the threshold of significance detected in replicate three (Table 4.3), for average head width, are likely to have been created by noise in the analysis, since peaks at this position were not found in any of the other replicates, in the average of replicates, or when replicates were analysed together. Moreover, there were no chromosomes on which QTL peaks occurred more frequently (Table 4.3). This inconsistency and the lack of QTL detected in the averaged replicates, or all replicates, cast doubt on the reliability of this QTL. The genomic data were checked by mapping the *H. lataniae* phenotype alongside the *C. obliquana* larval fitness phenotypes in R/qtl. The QTL peak on chromosome 10 with a LOD of 24 was well above the threshold LOD of 3.7 (Appendix, Figure 7.7). The significant association indicates that the genomic data were suitable to identify QTL of strong effect. The lack of consistent QTL peaks between replicates, the lack of increased power with the average of replicates, and the acceptable genotype data all indicate a problem with the *C. obliquana* bioassay.

Table 4.3. QTL analysis of resistance to *C. obliquana* in *A. chinensis* var. *chinensis* results from R/qtl. No QTL were detected above the threshold of significance, and the highest minor peaks shared no chromosomes in common between replicates. LOD highlighted in salmon colour are above the significance threshold.

Phenotype	Replicate	LOD	Threshold	Chromosome	Position (cM)
Larvae alive at day 7	Rep. 1	2.36	3.66	12	97.35
Larvae alive at day 14	Rep. 1	2.05	3.53	25	118.45
Larvae alive at day 21	Rep. 1	2.68	3.52	5	130.23
Average weight (mg)	Rep. 1	2.17	3.42	4	0
Average head width (mm)	Rep. 1	1.99	3.46	25	21.27
Average Instar	Rep. 1	2.23	3.53	28	55.08
Larvae alive at day 7	Rep. 2	3.15	3.63	9	28
Larvae alive at day 14	Rep. 2	3.22	3.62	22	23.39
Larvae alive at day 21	Rep. 2	2.91	3.65	19	76.12
Average weight (mg)	Rep. 2	2.45	3.59	23	138.17
Average head width (mm)	Rep. 2	2.67	3.55	8	166
Average Instar	Rep. 2	3.20	3.70	2	117.78
Larvae alive at day 7	Rep. 3	2.22	3.72	8	52
Larvae alive at day 14	Rep. 3	2.85	3.76	27	88
Larvae alive at day 21	Rep. 3	2.85	3.66	9	32.85
Average weight (mg)	Rep. 3	2.73	3.62	3	100.31
Average head width (mm)	Rep. 3	3.70	3.58	3	104.93
Average Instar	Rep. 3	3.88	3.63	10	80.92
Larvae alive at day 7	Avg. of reps	2.25	3.76	3	70.31
Larvae alive at day 14	Avg. of reps	3.11	3.74	12	0
Larvae alive at day 21	Avg. of reps	3.00	3.71	12	0
Average weight (mg)	Avg. of reps	1.97	3.74	19	61.53
Average head width (mm)	Avg. of reps	3.47	3.78	3	104.93
Average Instar	Avg. of reps	2.01	3.71	26	8.99
Larvae alive at day 7	All reps	2.69	3.73	3	70.31
Larvae alive at day 14	All reps	3.14	3.70	12	0
Larvae alive at day 21	All reps	2.71	3.69	12	0
Average weight (mg)	All reps	2.85	3.68	9	100
Average head width (mm)	All reps	2.83	3.64	3	104.93
Average Instar	All reps	2.64	3.60	23	70.90

4.4.3 Discussion

Production of kiwifruit (*A. chinensis*) is a multibillion-dollar per annum industry in New Zealand (MPI, 2022) that is impacted by the brown-headed leafroller moth (*Ctenopseustis obliquana*). Nevertheless, the molecular genetics of kiwifruit resistance to *C. obliquana* has not been explored. Identifying loci for resistance to *C. obliquana* may enable breeders to utilise markers in the early screening of susceptible individuals, thereby enhancing the probability of new cultivars carrying loci for resistance to *C. obliquana*. In this experiment, loci responsible for resistance to *C. obliquana* in *A. chinensis* var. *chinensis* were sought through an association mapping approach to produce QTL. However, no reliable QTLs were identified, despite

collecting detailed phenotypes on the growth and survival of *C. obliquana* larvae and analysing high-density genotype data generated by GBS. The lack of QTL associated with resistance to *C. obliquana* indicates that the bioassay was unsuitable for collecting the accurate phenotypic data required for QTL analysis.

Identifying QTL requires a large dataset with low error and accurate phenotypic and genotypic data. This experiment analysed three clonal replicates of 190 individuals from the B1 population. The same B1 population and GBS genotyping data were used in the work described in Chapter 2 to detect QTL for resistance to *H. lataniae*. However, this experiment detected no QTLs for resistance to *C. obliquana*. The lack of QTL detection warranted separating each replicate for individual analysis. Separating each replicate in this way lowered the power of the analysis, but also eliminated any variance between replicates. Still, no significant QTL were detected within any of the replicates. This result was consistent with the complete analyses that included all replicates with repeated markers and the analyses in which phenotypes were averaged for each replicate. The lack of QTL detection in any of the five analyses indicated that error existed in the data.

The lack of QTLs with a significant effect and the quantitative nature of *C. obliquana* larval growth suggests that a single major gene is unlikely to be responsible for the resistance to *C. obliquana*. Even with the significant variance, we should expect some signal if there was a large effect gene. It is more likely that many minor genes contribute to the resistance of Female 6 to *C. obliquana*. It is common for multiple genes to confer resistance to chewing insects because, unlike sucking insects, chewing insects consume the site they feed on, leaving less time for induced resistance mechanisms to affect the feeding site (Dixit et al., 2020).

The quantitative nature of resistance cannot explain the lack of QTL detected. Instead, the variance in the measured phenotypes may have contributed significantly to error in the analysis. The susceptible cultivar 'Hayward' had enough variance to overlap with the measured phenotypes on the resistant cultivar. The variance in measured phenotypes between resistant and susceptible individuals is likely to be a causal factor in the lack of QTL detection. However, because the variance sources are confounded in the measured phenotypes, the variance could not be modelled to remove it from the measured phenotype data.

A portion of the phenotypic variance may have come from the variable quality of *C. obliquana* larvae. The variance in *C. obliquana* larval quality may have been caused by a diet component missing from the *C. obliquana* colony. The effect of the missing diet components was only noticed toward the end of phenotyping the third replicate. The lack of nutrients in the *C. obliquana* colony's diet reduced the fertility of the individuals produced simultaneously with larvae used in this experiment. *C. obliquana* larvae may also have natural genetic variation within their population to overcome plant resistance. Heterogeneity for traits that allow them to overcome the resistance observed in Female 6 may explain the variance observed in the resistant controls. However, it would not explain the variance within susceptible controls.

The variance observed within control and trial plants may have been introduced to the experiment from differences between leaves on each plant. The variance within the resistant control trial plants could have been from variation in the properties of harvested leaves due to damaged caused by the two-spotted mite. While this variable was controlled by collecting leaves free from damage, some two-spotted mite damage can go unnoticed. The damage caused by the piercing/sucking insect could have primed the plant's defence system for this type of attack. The priming of plant defence is known to occur in response to piercing/sucking insects by inducing the salicylic acid defence pathway (Costarelli et al., 2020; Hill et al., 2016; Pieterse et al., 2012). The induction of the salicylic acid resistance pathway can also affect resistance to chewing insects such as *C. obliquana* (McConn et al., 1997). Salicylic acid can reduce the resistance of plants to chewing insects because it can antagonise the jasmonic acid resistance pathway used in plant defence against chewing insects (Costarelli et al., 2020; McConn et al., 1997).

Further, the variable quality of leaves available for sampling within an individual also may have contributed to the bioassay variance. The variance between leaves may have been created because the loss of 10 plant leaves over the sampling period required for the bioassay affected the plants' nutrient status. The loss of ten leaves to a larger plant is not an issue, but few plants had more than 12 large leaves. The loss of these 13-24-cm-wide leaves affects nutrient availability within plants. Plants were re-potted and fertilised before the experiment, but a lack of nitrogen was observed during the trial. The dry weight of each leaf disk was recorded to measure this variable upon completion of each replicate. However, no correlation was found between the dry weight of leaves and the average weight of insects (R^2 = 0.0017), or the dry

weight of leaves and insect survival (R^2 = 0.0029). The dry weight of leaves indicates nutrient and carbon availability when the leaves had grown, but it does not indicate the nitrogen content of leaves when they were sampled. Since a lack of available nitrogen can lead to a lower concentration of jasmonic acid, the main trigger of resistance to chewing insects, this may have influenced the resistance response within plants (Martinez Henao et al., 2020).

The introduction of errors affecting QTL detection in this experiment also may have come from the loss of information from missing data. The loss of information would have reduced the statistical power, decreased the chances of identifying real resistance loci and increased the chances of presenting QTL at loci that are not associated with resistance (Kang, 2013). GBS data could also introduce errors to the QTL mapping results. Sites with insufficient sampling depth can cause heterozygous SNPs to be falsely called homozygous, because the alternate allele does not appear in the dataset, or its frequency is too low to appear in the dataset. The noise introduced by the genotyping error is likely to have contributed to the non-significant QTL. However, for QTL to be hidden by this noise, they would have to be of weak effect. Because other research using the same population found a QTL for the pathogen Psa, which is caused by many low-effect QTL, it is likely that the bioassay was at fault in this research.

Two maternal controls for *C. obliquana* resistance were included in each trial, and ten additional maternal controls were included in the same environment as the trials. The maternal controls indicated whether the fitness of *C. obliquana* larvae declined over time relative to the susceptible control. However, without the ability to quantify the effect of the environment on each individual within and between replicates, there was no way to identify re-ranking among individuals caused by environmental effects. The inability to identify re-ranked individuals could have been corrected if sufficient maternal and paternal controls had been included in each replicate row of plants in the glasshouse. If sufficient parent controls were included, the spatial effects of environmental variance within and between replicates could have been considered when analysing phenotypes using Bayesian analysis (Denwood et al., 2011; Filipe & Kyriazakis, 2019). Because leaves from susceptible controls were sourced from the field, the control helped determine whether insects were growing as expected within and between replicates. However, it did not allow the effect of the environment within or between replicates to be estimated.

A combination of variance from the bioassay, variable *C. obliquana* larval quality, variance from genotyping, and the polygenic nature of this resistance likely led to insignificant QTL results in this experiment. The variable contributing the most error to this experiment was likely the *C. obliquana* larval quality. However, this could not be confirmed because of other factors affecting the bioassay. The bioassay for resistance to *C. obliquana* could be improved in the future by applying larvae directly to leaves while attached to plants in the glasshouse. Leaves could be folded in half and surrounded in mesh bags that are closed at the leaf petiole to protect larvae and catch emerging adults. Adult moth emergence could then be measured on multiple leaves from each plant over time. This bioassay method would take longer but would give resistant plants the entire duration of larval development to display resistance attributes. Using this method would significantly reduce the lab-based workload and enable the measurement of variance between leaves on each plant. This method would expose *C. obliquana* larvae to a variable temperature and humidity glasshouse environment. But, with clones of the resistant and susceptible parents used as controls in each row of plants in the glasshouse, the variance coming from the glasshouse temperature could be accounted for by using the susceptible parent as the control for insect variance and the resistant parent as the control for plant variance. An adapted bioassay should allow identification of the QTL for resistance to *C. obliquana*. Markers could then be developed for the loci affecting QTL to assist the breeding of resistance to *C. obliquana* into new kiwifruit cultivars.

5 Bulked segregant analysis identifies QTL associated with resistance to Psa (*Pseudomonas syringae* pv. *actinidiae*) in kiwifruit (*Actinidia chinensis*)

5.1 Abstract

Pseudomonas syringae pv. *actinidiae* (Psa) is a bacterial pathogen of *A. chinensis*. It causes symptoms including leaf-spotting, cane dieback, wilting, and cankers (lesions) and can lead to plant death. Families of diploid *A. chinensis* seedlings grown in the field show a range of susceptibility to the pathogen with up to 100 % of seedlings in some families succumbing to Psa. But the effect of selection for field resistance to Psa on the alleles that remain in surviving seedlings has not been assessed. In this project, the effect of Psa on the allele frequency of an incomplete-factorial-cross population was analysed. This population was founded using a range of genotypically distinct diploid *A. chinensis* var. *chinensis* parents to make 28 F_1 families. However, because of the diversity of these families, low numbers of surviving individuals, and a lack of samples from dead individuals, standard quantitative trait loci (QTL) mapping approaches were unlikely to yield good results. Instead, a modified bulked segregant analysis (BSA) overcame these drawbacks while reducing the costs of sampling and sample processing, and the complexity of data analysis. Because the method was modified, the first part of this project was used to determine the signal strength required for a QTL to be detected with BSA. Once QTL detection accuracy was known, the second part of this project analysed the 28 families from the incomplete-factorial-cross population that had multiple individuals removed due to Psa infection. Each family was assigned to one of eight bulks based on a single parent that contributed to the families. DNA was extracted in bulk by grinding sampled leaf disks together before DNA extraction. Each sample bulk was compared against a bulk made up of whole genome sequence (WGS) data from the parents contributing to the sample bulk. The deviation in allele frequency from the expected allele frequency within surviving populations using the modified BSA method was able to highlight 11 QTL for Psa that were present in at least two analyses. The identification of these Psa resistance loci using the modified BSA method will be vital for future kiwifruit breeding.

5.2 Introduction

Psa (*Pseudomonas syringae* pv. *actinidiae* biovar 3), also known as the virulent form of Psa (Psa-V), is a bacterial pathogen that is devastating to many cultivars of kiwifruit (McCann et al., 2013). This pathogen is particularly destructive to *A. chinensis* var. *chinensis* (Datson et al., 2013) genotypes, but also affects *A. chinensis* var. *deliciosa* (Takikawa et al., 1989). It has been reported that Psa spread from Asia, where up to four pathovars were present (Koh et al., 1994). Each of the non-virulent biovars had different pathogenesis and molecular characteristics on different kiwifruit genotypes, but they did not cause the pathogenesis observed in the virulent Psa-biovar-3. Psa-biovar-3 was first reported in New Zealand in 2010 (Everett et al., 2011) and is now widespread in the North Island of New Zealand where most kiwifruit are produced. Plants infected with this bacterium often start showing symptoms, including leaf-spotting, cane dieback, wilting, or oozing a clear, brown or white liquid in spring or autumn from cankers (lesions). Plants with severe symptoms such as cankers, or multiple dead canes, were removed in line with regulations established by the national agency, Kiwifruit Vine Health (KVH). In highly susceptible genotypes, these symptoms occur on multiple canes leading to whole vine death. On more resistant genotypes, symptoms can involve flower bud browning, bud drop, flower wilting, and leaf spotting (Everett et al., 2011). During the initial outbreak in 2011, the leading yellow-fleshed *A. chinensis* var. *chinensis* cultivar in New Zealand, named 'Hort16A', and its pollenisers, were particularly affected by Psa, which led to a major drop in gold kiwifruit production (Dwiartama, 2017). Fortunately, a gold fleshed *A. chinensis* var. *chinensis* cultivar with the PVR name 'Zesy002' (fruit marketed as Zespri™ SunGold) had been in pre-commercial trials before the Psa outbreak and this was used to replace 'Hort16A' (Dwiartama, 2017; Everett et al., 2011). The 'Zesy002' cultivar proved to be more resistant to Psa but, Psa is still a major pathogen affecting 'Zesy002' production.

As Psa is a ubiquitous pathogen in most kiwifruit-producing areas of the North Island of New Zealand and many other kiwifruit-producing countries (Ferrante & Scortichini, 2010), incorporation of resistance to Psa is required for any kiwifruit exposed to field conditions. Breeding programmes currently retain moderate resistance to Psa by crossing individuals from Psa resistant families with other Psa resistant individuals. However, due to the highly polygenic nature of resistance to Psa, strong resistance has not yet been achieved in the gold-fleshed *A. chinensis* var. *chinensis* (Tahir et al., 2017; Tahir et al., 2019). QTL for resistance to Psa have

been identified in two families of *A. chinensis* var. *chinensis* resulting from crossing a resistant individual with a susceptible individual, but identifying these QTL required large replicated trials from a single family and detailed phenotyping (Tahir et al., 2019). The phenotyping requirement, and the requirement of large, replicated families to generate QTL for resistance to Psa, could be overcome by identifying alleles remaining in breeding populations after the selective sweep caused by severe Psa infection.

When Psa spread through kiwifruit families established at Plant and Food Research, Kerikeri, New Zealand, there was significant variation in seedling removal among these families due to Psa (personal communication Paul Datson). The selective sweep caused by Psa created an opportunity to analyse the genomic regions that remained in families using bulked segregant analysis (BSA) to understand the genetic architecture of resistance and susceptibility to Psa. This is because BSA examines the change in allele frequencies between populations that have segregated due to selection pressure, like a selection map (Li & Xu, 2022; Magwene et al., 2011; Michelmore et al., 1991; Shen & Messer, 2022; Wisser et al., 2008). The BSA technique has been used for the detection of QTL for target traits in various species, including dwarfing in watermelon (Dong et al., 2018), cotyledon colour in soybean (Song et al., 2017), cold resistance in rice (Sun et al., 2018), resistance to ascochyta blight in chickpea (Deokar et al., 2019), and kernel length-width ratio in wheat (Xin et al., 2020). A typical BSA investigates loci that differ between sample bulks segregating for a trait of interest, combining ideas from linkage mapping and GWAS (Li & Xu, 2022; Michelmore et al., 1991; Shen & Messer, 2022). Like classical linkage mapping, most BSA trials are designed using two parental strains with different phenotypes. The two strains are crossed to generate an F_1 population which is back-crossed or interbred for several generations to generate sufficient recombination to break up linkage from parents (Michelmore et al., 1991). Individuals from the last generation are selected to form two bulks that segregate for the phenotype of interest. Thus, alleles affecting the target phenotype should show a significant difference in frequency between the two bulks, while unselected alleles should remain in both bulks at similar frequencies (Michelmore et al., 1991; Shen & Messer, 2022). Diverging from the typical BSA, bulks can be analysed with BSA directly from F_1 populations (Dai et al., 2018; Guan et al., 2019). Regardless of the method, allele shifts not caused by the target selection pressure can be caused by the genetics of the founding parents (Chen et al., 2019; Conolly et al., 2008; James, 1970). Similarly, selection

mapping approaches compare a shift in allele frequency between two bulks created from samples of the population before and after a selection event altered the population's allele frequency (Johnsson, 2018; Wisser et al., 2008). The DNA that contributes to each of the bulks in BSA and selection mapping approaches are typically quantified for each individual, and an equivalent amount of DNA added to the bulk from each individual (Dong et al., 2018; Munjal et al., 2018; Song et al., 2017; Wang et al., 2021). While this approach ensures that a precise quantity of DNA is added from each individual, tracking samples and extracting DNA from individuals is costly and time-consuming.

An alternative approach to bulking DNA samples would be to bulk leaves of different individuals prior to DNA extraction. This approach would simplify sampling and reduce the cost and workload involved with DNA extraction by extracting DNA directly from a bulk of leaf samples. To help standardise the DNA contribution from each sample, the leaf sample growth stage and the amount of leaf material would need to be kept consistent. Samples from each individual then could be ground together for DNA extraction as a bulk. However, this approach precludes a precise balance of each individual's DNA contribution to the pool and may introduce greater variance into the bulks; individuals that potentially contribute a greater amount of DNA would make a greater contribution to the allele frequencies than others with less DNA extracted. This may decrease the power to identify allelic differences between bulks and thus QTL. Prior to applying such a modified method to an experimental population, a test of the approach to detect selection at a known site would need to be performed to determine its accuracy.

Testing the level of precision of the modified method of bulk sampling would require a population segregating for a simple control trait that is determined by a single well-characterised locus, and ideally with low interaction between the gene and the environment. To this end, within the dioecious *A. chinensis*, plant sex is a suitable trait as it is easy to phenotype, controlled by a single well-characterised dominant gene that is not affected by the environment. This kiwifruit sex gene, named *Shy Girl* (*SyGI*) exists on the male y chromosome and suppresses flower feminisation, producing males in plants possessing it (Akagi et al., 2018). It was assumed that if a shift in allele frequency could be detected using these methods in the monogenic *Shy Girl* gene, polygenic loci of strong influence on the population would also be able to be detected.

The work described here aimed first to test whether pooling leaves from multiple individuals prior to DNA extraction enables a BSA to be effectively carried out on the resulting DNA pool. This was done using a series of bulked pools that varied in the ratio of male and female *A. chinensis* var. *chinensis* individuals that contributed to the pools and investigated whether the QTL for plant sex could be identified on chromosome 25. The second part of this work aimed to identify any changes in allele frequency between bulks of sample pools of seedlings that had survived in the field following a Psa selective sweep and a bioinformatically generated bulk of data from parents contributing to each sample bulk. Regions of the genomes where alleles have a greater sample depth in the WGS of sample bulk data than expected from their parental bulk of data should highlight the regions of the genome under strong selection from Psa.

5.3 Methods

5.3.1 Population

For this work, a diverse population of diploid *A. chinensis* var. *chinensis*, named "18x12", was identified at Plant and Food Research, Kerikeri, New Zealand as suitable for both aims of this project. The seedling vines for this population were planted in 2015 and grown on a T-bar system, with 0.75 m between each plant and 3 m between rows. The population was spread over three blocks of 4000 m^2 with 6-m high hedging shelter belts dividing and surrounding the blocks. This population was exposed to Psa naturally present in the Kerikeri orchard when seedlings were planted. The exposure to Psa resulted in symptoms including tip dieback, cane death, oozing from infected cankers, and whole plant death in highly susceptible individuals. The population was managed for Psa symptoms by removing canes if tip death or cane death was present on a single cane. Where more than three canes were infected with cankers or cane dieback, the whole vine was removed from the orchard. The structure of the 18x12 population was made by crossing a diverse set of 18 female and 12 male parents from *Actinidia* germplasm in an incomplete factorial design, which resulted in 63 families. A variable number of 33, 48, or 56 seedling plants from each family were planted in the field after being established in pots. Fifty-nine families had individuals remaining after 4 years (Figure 5.1, Figure 5.2). Between 2015 and 2019, from 63 % to 100 % of individuals were removed from different families due to severe Psa infection. Of the families with surviving individuals, 25 had sufficient individuals to include in the current study. However, few of these had sufficient numbers on their own, prompting the bulking of families based on parents.

		Male							
		P1	P2	P3	P4	P5	P6	P7	P8
Female	P9		6% (3/48)		15% (7/48)		54% (26/48)	17% (8/48)	
	P10	4% (2/56)				13% (7/56)			27% (9/33)
	P11			27% (13/48)	44% (21/48)				
	P12	63% (30/48)	19% (9/48)				50% (24/48)		
	P13			8% (4/48)					
	P14				4% (2/56)				
	P15				2% (1/48)				
	P16		11% (6/56)						
	P17		7% (4/56)						
	P18	2% (1/56)		38% (18/48)					
	P19		6% (3/48)		13% (6/48)				
	P20			4% (2/48)					
	P21	6% (3/48)							
	P22	4% (2/48)	23% (11/48)						

Figure 5.1 The crossing structure of 25 families of *A. chinensis* individuals that survived exposure to *Pseudomonas syringae* pv. *actinidiae* (Psa) from the 18x12 population. For the second part of the project, eight bulks were made by sampling leaves from all surviving plants within families which shared the parent indicated in the blue columns and salmon rows. Female parents are shown on the left-hand side with male parents at the top. Percentages in cells indicate the number of individual F_1 seedlings remaining after being exposed to Psa for four years in the field. The numbers in brackets are the number of individuals remaining from the total number of individuals that were planted from the family. For example, 4 %, or two of 56 plants survived after four years in the field from the cross between P1 and P10.

5.3.2 Comparing the genomic difference between the parents contributing to bulks

Because the methods used in the second part of this project could be influenced by the similarity of parents, the genomic distance between parents needed to be tested. The genomic distance between individuals can be analysed with PCA by transforming genomic data into a Boolean vector, as described in Konishi et al. (2019). The variants from parents were used to identify the genomic distance between each parent with WGS data.

5.3.3 Sample collection, DNA extraction and sequencing

The field sampling, DNA extraction processes, and sequencing methods were the same for both parts of the project. Sampling was kept simple by placing a single leaf from each plant destined

for a bulk into a plastic bag labelled with the bulk's name. Leaf samples were taken from the third leaf from the growing cane tip and kept cold in a chilly box with ice while sampling. After field collection, the bulks of leaves were stored in a -80 °C freezer before processing. Each leaf had a 10-mm diameter leaf disk removed while frozen, with care taken to avoid the leaf midrib and large veins. All leaf disks from a bulk were finely ground together in liquid nitrogen with a pestle and mortar. DNA was extracted from the ground material with a Qiagen DNeasy® Plant Maxi kit. To remove pectin from samples, DNA was precipitated by adding 1/10 volume sodium acetate (3 M, pH 5.2) to two times the volume (calculated after addition of sodium acetate) of at least 95 % ethanol. Samples were incubated on ice overnight, then centrifuged at 14000 g for 30 min at 4° C. Supernatant was removed and rinsed with 70 % ethanol, then centrifuged at 14000 g for 15 min. The supernatant was discarded, and the pellet dissolved in TE buffer (pH 8.0). TE buffer was made by adding 100 mL of 1M Tris-Cl (pH 8.0) to 20 uL of 0.5 M EDTA (pH 8.0) to 9.880 mL of reverse osmosis water. DNA quality and quantity were checked using a Qubit® 2.0. In samples with a low DNA quantity, extraction was repeated. In samples with low quality DNA, identified by a 260/280 value of under 1.6, DNA was cleaned of pectin using a second ethanol precipitation step. In this step, DNA was precipitated in 98 % ethanol and the DNA pellet was lightly massaged with a spatula against the Eppendorf tube wall to remove pectin within the DNA precipitate. A minimum of 1400 ng of DNA from each bulk was sent to the Australian Genome Research Facility for library preparation and sequencing. The PCR free library preparation technique TruSeq was chosen to reduce the effects of PCR on allele depth with the number of samples included per bulk described in the next section. Whole-genome sequence (WGS) data were captured from each bulk at 30x coverage with 150 bp paired-end reads on an Illumina NovaSeq 6000.

Using this method of bulking leaf samples forgoes the step of extracting DNA from individual plants, quantifying the DNA from each extraction and adjusting the amounts so that each bulk contains an even amount from each contributing individual. However, it also introduces the risk of having a variable quantity of DNA added from each individual to a bulk and may therefore increase the error associated with analysing allele depth.

5.3.4 Bulking samples for part one of the project

Since the sensitivity of these methods to detect a shift in allele frequency was unknown, the sensitivity of the methods used to detect various magnitudes of allele frequency shift was measured in the first part of the project. This project used nine bulks of DNA from 19-20 individuals containing about 10 %, 20 %, 25 %, 40 %, 50 %, 55 %, 68.4 %, 78.9 % or 89.5 % male contribution to the bulk (Table 5.1). Male and female F_1 individuals were selected from a mix of families from the 18x12 population that were derived from parent P8 as the father (Figure 5.1, Figure 5.2). Female parents of families used in each bulk included P10, P14, P15, and P17.

Table 5.1. The number of male and female *A. chinensis* contributing to bulks in part one of the project. The percentage of males within each of the nine bulks ranged from 10.5 % in bulk one to 90 % in bulk nine.

Bulk	Number of males in bulk	Number of females in bulk	Total individuals in bulk	Percentage of males in bulk	Percentage of females in bulk
1	2	17	19	10.5	89.5
2	4	15	19	21.1	78.9
3	6	13	19	31.6	68.4
4	9	11	20	45	55
5	10	10	20	50	50
6	12	8	20	60	40
7	15	5	20	75	25
8	16	4	20	80	20
9	18	2	20	90	10

5.3.5 Bulking samples for part two of the project

The second part of this project investigated WGS data from bulks to detect whether there was a shift in allele frequencies within bulks of individuals that remained in families after exposure to Psa. This project was complicated because severe Psa infection had led to the removal of many individuals from all the families in the population. Because some families had very few individuals remaining, each of the eight sample bulks included resistant individuals from up to six families. These were bulked based on a single parent that contributed to all the families in

the bulk (Figure 5.1, Figure 5.2). For example, the bulk P1 contained F_1 families from crosses P1 x P10, P1 x P12, P1 x P18, P1 x P21, and P1 x P22.

This part of the project differed from a typical BSA because there was no DNA from individuals that were removed because of Psa. Instead, the frequency of alleles in surviving individuals was compared against a bioinformatically generated bulk of data from parents contributing to the sample bulk. The bioinformatically generated parent bulks were used in place of bulks of individuals susceptible to Psa. Bulks like this can be used because the alleles in the parental bulks were representative of the families included in the bulks without selection. This methodology is similar to that done for selection mapping (Matsumoto et al., 2017; Wisser et al., 2008), but it has the drawback of assuming no other influences on allele transmission. The bioinformatically generated parent bulks were made by merging parental BAM files before variant call files were made. However, not all parents that contributed to the families used in parental bulks had WGS data available (Figure 5.2). As a result, the bulks of parents that contained some parents without WGS data would give a less accurate representation of the population before selection. The loss of information was particularly apparent in the parental bulk of P9, which had 29.5 % of its theoretical DNA contribution missing from its P6 parent and 9.1 % missing from the P7 parent (Figure 5.2). The missing data from parents would have resulted in some alternate alleles present in these parents not being included in the analysis. Unfortunately, once the pools were established during field sampling, the families with missing WGS data could not be removed from pools.

			Sample bulk parent (male)				Sample bulk parent (female)			
	Parents with WGS	bulk parent crossed with	P1	P2	P3	P4	P9	P10	P11	P12
Male polleniser of female sample bulk parent	WGS	P1						5.6%		23.8%
	WGS	P2					3.4%			7.1%
	WGS	P3							19.1%	
	WGS	P4					8.0%		30.9%	
	noWGS	P5						19.4%		
	noWGS	P6					29.5%			19.0%
	noWGS	P7					9.1%			
	WGS	P8						25.0%		
Female parent used in cross with male sample bulk parent	WGS	P9		4.2%		9.5%				
	WGS	P10	2.6%							
	WGS	P11			18.1%	28.4%				
	WGS	P12	39.5%	12.5%						
	WGS	P13			5.4%					
	noWGS	P14				2.7%				
	WGS	P15				1.4%				
	noWGS	P16		8.3%						
	noWGS	P17		5.6%						
	noWGS	P18	1.3%		24.3%					
	noWGS	P19		4.2%		8.1%				
	noWGS	P20			2.7%					
	WGS	P21	3.9%							
	noWGS	P22	2.6%	15.3%						
Expected percentage of sample bulk DNA contributed by parents samples were bulked on			50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%
Expected percentage of sample bulk DNA contributed by parents with WGS data			96%	67%	73%	89%	61%	81%	100%	81%
Expected percentage of sample bulk DNA missing from parent bulk			4%	33%	27%	11%	39%	19%	0%	19%

Figure 5.2 Percentage of each *A. chinensis* parent's theoretical contribution to sample bulks. Bulks were based on the parents in columns, with parents contributing to the sample bulk in rows. Families from parents with grey-filled parent names were represented twice where the family was used in bulks based on male and female sample bulk parents. Parents with whole-genome sequence (WGS) data have cells filled in green, and those without WGS are filled in salmon. Parent bulks contained DNA only from the parents with green shading. The total theoretical DNA contribution missing from parents without WGS data in parent bulks is shown in the bottom row.

5.3.6 Bulked segregant analysis part one

To test the limits of the methods used to bulk samples and extract DNA to determine the architecture for Psa tolerance, WGS data for part one of this project were analysed with the QTLseqR package v0.7.5.2 (Mansfeld & Grumet, 2018). The analysis included nine bulks, with a varying number of males added to each bulk, were each compared with each other for 36

separate analyses, described further below. These comparisons were expected to present a QTL peak in the bulked segregant analysis at 1.6 Mb on chromosome 25. QTL were expected on chromosome 25 because it contains the heterozygous dominant sex-determining *Shy Girl* gene that suppresses the feminisation of flower production to generate male flowers and thus a male plant (Akagi et al., 2018). But detection of QTL at the *Shy Girl* gene locus could only occur if the methods used were tolerant enough of the sampling and bulking methods, the effect of Psa on the families, and the relationship between the samples for the bulks, since these would have an influence on frequency of alleles between bulks. For example, if a bulk with a 5 % difference in male number were compared and QTL were consistently detected in comparisons with different backgrounds, and with a similar difference in male percentage between pools, it could be assumed that the methodology added 5 % of error to the analysis.

Bulks of males were compared with each other because WGS data were unavailable for two of the five male bulk parents, P14 and P17. Thus, a bulk of these parents would not accurately represent the bulks of parental data. Instead, data from each of the nine bulks with a known percentage of males were compared with each other, resulting in 36 separate analyses that compared pairs of bulks. The difference between the percentages of males between bulk pairs varied between 5 and 79.5 %.

WGS data for both parts of this project were supplied by AGRF as compressed FASTQ formatted sequence files containing single read sequence output with Illumina quality scores. Binary alignment files (BAM files) were generated by aligning reads of parent P8 to the "Russell" reference genome (Tahir et al., 2022) by Roy Storey using BWA-MEM (Yao et al., 2020). The author completed the subsequent bioinformatics work using the R coding language. BAM files from separate flow cell lanes were merged with Picard "MergeSamFiles". Samtools was used for sorting and indexing BAM files. Variant call files were generated using BCFtools mpileup with options including setting a minimum base quality of 20 and disabling probabilistic realignment to help reduce false SNPs caused by misalignments. Indel calls were excluded. Optional tags included the depth at each site, the depth of each allele, and the Phred-scaled strand bias P-value. Uncompressed output was piped to BCFtools call, which included the genomic quality and genotype posterior probability format fields and the multiallelic caller option. The resulting variant call files were indexed using BCFtools index. BCFtools query was used to split data into separate comma-separated value text files for each chromosome and

exclude sites with a depth of less than 20 or greater than 200, and data were read into R/datatable.

The SNP index for each bulk pair to be analysed was calculated by dividing the alternate allele depth by the total read depth. The reference allele frequency was calculated by summing the reference allele depth of bulks being compared and dividing the result by the sum of the total depth of the bulks being compared. The delta-SNP index was calculated by subtracting the SNP index of the sample bulk from the parental bulk. The modified G statistic was calculated for each SNP based on the observed and expected allele depths (Magwene et al., 2011) and smoothed using a tricube smoothing kernel (Watson, 1964) in QTLseqR (Mansfeld & Grumet, 2018). The Gprime value was calculated from the tricube smoothed G statistic by taking the average weight of the physical distance across the neighbouring SNPs within the 1-Mb window. This approach accounted for linkage disequilibrium and minimized the noise attributed to SNP calling errors (Magwene et al., 2011). SNPs were filtered using the QTLseqR package selecting a reference allele frequency of 0.05, a minimum total depth of 60, a maximum total depth of 160, an allele depth difference of less than 50 between bulks, a minimum sample depth of 10 and a minimum genomic quality of 100. The QTLseqR analysis package had the bulk size set to 20 individuals with the Gprime window size set at 1 Mb. Because the adjusted p statistic threshold failed to detect peaks with a low difference in allele frequency, the top 0.5 % of SNPs were highlighted in green.

5.3.7 Bulk segregant analysis part two

Part 2 of this project used the same bioinformatics pipeline as part one of the project to generate VCF files of sample bulks. However, part two of the project differed from part one because a modified BSA approach was used for bulk creation and sample bulks were compared against the bioinformatically generated bulk of WGS data from parents that contributed to the bulked families (Table 5.1). These parent bulks were created by merging BAM files from parents using samtools-merge. VCF files were created with BCFtools-mpileup using the same options as in part one of this project.

Before the sample bulks and parent bulks could be compared, calculations based on SNP data from VCF files were performed. VCF files were read into R/datatable, and a SNP index was calculated for each bulk by dividing the alternate allele depth by the total read depth. The

reference allele frequency was calculated by summing the reference allele depth of the sample bulk and the parent bulk, and dividing the result by the sum of the total depth of the bulks being compared (Mansfeld & Grumet, 2018). The delta-SNP index was calculated by subtracting the SNP index of the sample bulk from the parental bulk (Mansfeld & Grumet, 2018).

The data preparation for analysis in QTLseqR was done similarly to that done in part one. First, BCFtools-query split data into separate .csv files for each chromosome and excluded sites with a depth of less than 20 or greater than 200. Data were read into R/datatable, and the SNP index per bulk was calculated by dividing the alternate allele depth by the total read depth. Unlike in part one, in part two the reference allele frequency was calculated using the sum of reference allele depths of sample bulks and dividing the result by the sum of the total depth of the parental bulks. The delta-SNP index was calculated by subtracting the SNP index of the sample bulk from the parental bulk.

Data from each of the eight sample bulks were compared with their parent bulk using the Gprime analysis portion of the QTLseqR package (Mansfeld & Grumet, 2018). Gprime analysis was used because the average G values across SNPs in the 1-Mb sliding window reveal the signal of divergence in allele frequency between bulks that are conserved between closely linked sites (Magwene et al., 2011; Mansfeld & Grumet, 2018). Using the G value reduces the influence of random noise due to variable sequencing read coverage (Mansfeld & Grumet, 2018). Within the QTLseqR package, SNPs were filtered by depth for each comparison depending on the data distribution. Minimum and maximum total depth were set to remove SNPs of extremely low and extremely high frequency (Table 5.2) (Mansfeld & Grumet, 2018). Filtering SNPs by read depth helps remove SNPs with low confidence due to low coverage, or remove SNPs in repetitive regions that would have an artificially inflated read depth (Mansfeld & Grumet, 2018). Settings for the Gprime analysis method were as follows: the sliding window size was set at 1 Mb, the outlier filter was set as “deltaSNP”, and the filter threshold was set at 0.4. The resulting Gprime values for each SNP site were plotted with the top 0.5 % of Gprime values highlighted in green.

Table 5.2. Depth filter settings applied whole genome sequence data from *A. chinensis* before Gprime analysis. Sample bulks P1-P12 retained the same reference allele frequency, allele depth difference, minimum sample depth and minimum genomic quality, but varied in the minimum total allele depth and maximum total allele depth depending on the distribution of depth data in each sample bulk.

Sample bulk	Refence allele frequency	Minimum total depth	Maximum total depth	Depth difference	Minimum sample depth	Minimum genomic quality
P1	0.05	40	85	50	10	100
P2	0.05	35	85	50	10	100
P3	0.05	20	85	50	10	100
P4	0.05	40	150	50	10	100
P9	0.05	40	130	50	10	100
P10	0.05	40	90	50	10	100
P11	0.05	60	170	50	10	100
P12	0.05	30	100	50	10	100

5.4 Results

To determine the genomic relationship between the parents used for this project, parents that had WGS data were compared against each other using a principal component analysis (PCA) (Konishi et al., 2019). The result showed a close relationship among the half-sibling individuals P8 and P9, with a greater distance between P8 and P9 and the other individuals at PC1. PC2 showed an even distribution of genomic relationship between the remaining individuals with the exception of P1 and P2, which had minimal genomic distance on PC2 (Figure 5.3).

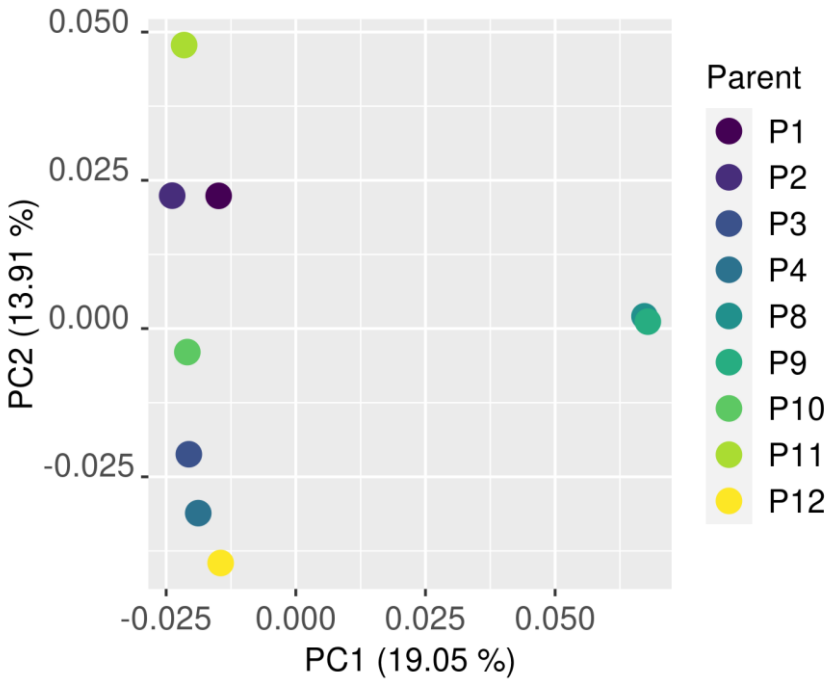


Figure 5.3. The genomic distance between *A. chinensis* parents that contributed to bulks analysed by principal component analysis. A close relationship among the half-siblings P8 and P9 was found with a greater distance between P8 and P9 and the other individuals.

The methods used for the BSA in this project differed from the standard methods used for BSA in QTLseqR (Mansfeld & Grumet, 2018). Therefore, the sensitivity of these methods to detect QTL in bulks of *A. chinensis* var. *chinensis* needed to be tested. The first part of this project tested the resolving power of the methods by making 36 pairwise comparisons at the sex loci on chromosome 25 among the nine bulks of individuals with a differing percentage of males.

However, QTLs were detected in only 12 of the 36 bulk comparisons when using the adjusted p = 0.05 threshold (Figure 5.4). Increasing the threshold to adjusted p = 0.1 included more QTL peaks, but also significantly increased the signal-to-noise ratio. Because the adjusted p-value based threshold could be caused by the alternative method of bulking multiple families or lack of inclusion of some parents in the bioinformatically generated bulk of parents, the significance threshold was changed to use the top 0.5 % of Gprime values. Using the top 0.5 % of Gprime values allowed QTL detection from BSA with greater accuracy in bulks, detecting QTL at the sex-linked gene locus in 19 of the 36 bulks analysed. However, using the top 0.5 % of Gprime values as a threshold of significance for QTL detection also has the disadvantage of missing smaller peaks for QTL in BSA plots where the signal for selection for some QTL was strong and covered a large range of loci. For example, the P3 bulk may have signal for selection on chromosomes 25 and 29, but the peaks on chromosomes 11 and 12 hold the top 0.5 % of Gprime values.

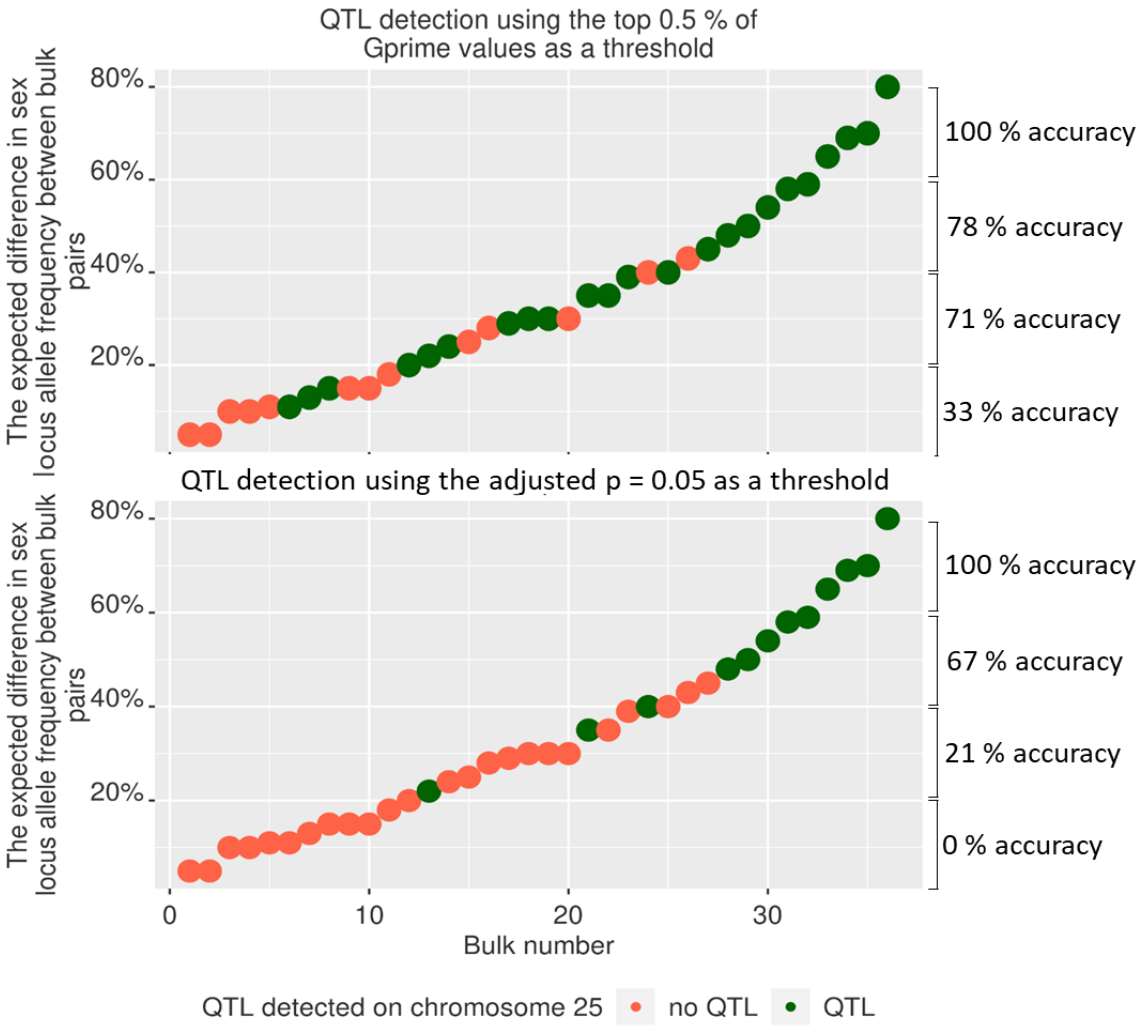


Figure 5.4. The detection of quantitative trait loci (QTL) on chromosome 25 of *A. chinensis* from pairwise comparisons between nine bulks with a different percentage of males added to each bulk. This figure presents only QTL from chromosome 25, but the analysis was completed over the whole genome. Using the top 0.5 % of SNPs over the whole genome (top) to detect QTL peaks gave fewer false-negative results than QTL peaks, which were deemed statistically significant using an adjusted p-value of greater than 0.05. When using the top 0.5 % threshold, the detection accuracy was estimated to be 33 %, 71 %, 78 %, and 100 % for an expected difference in allele frequency between bulks of 0-20 %, 20-40 %, 40-60 %, and 60-80 %, respectively. When using the adjusted p=0.05 threshold, the detection accuracy was estimated to be 0 %, 21 %, 67 %, and 100 % for an expected difference in allele frequency between bulks of 0-20 %, 20-40 %, 40-60 %, and 60-80 %, respectively.

To determine the effect of Psa on *A. chinensis* var. *chinensis* alleles in an incomplete factorial population, in part two of the project, samples that survived Psa were assigned to bulks based on families with a parent in common (Figure 5.1, Figure 5.2). Using BSA, the eight sample bulks were compared against bioinformatically created bulks of parental WGS data. The resulting BSA (Figures 5.5 - 5.12) identified sites of higher frequency in the sample bulk compared to the parent bulk, potentially caused by selection for resistance to Psa. The QTL presented as higher Gprime values in the resulting BSA plots, with the top 0.5 % of Gprime values considered significant QTL.

In theory, the variants for resistance to Psa had a maximum potential selection of 50 %. For example, in the case of a cross ab x cc with resistance associated with the 'a' variant, if there was strong selection for the 'a' variant in all seedlings, the resulting family containing ac variants would have a variant frequency 50 % higher than if there was no selection producing ac and bc variants. The b variant will also decrease in frequency by 50 %.

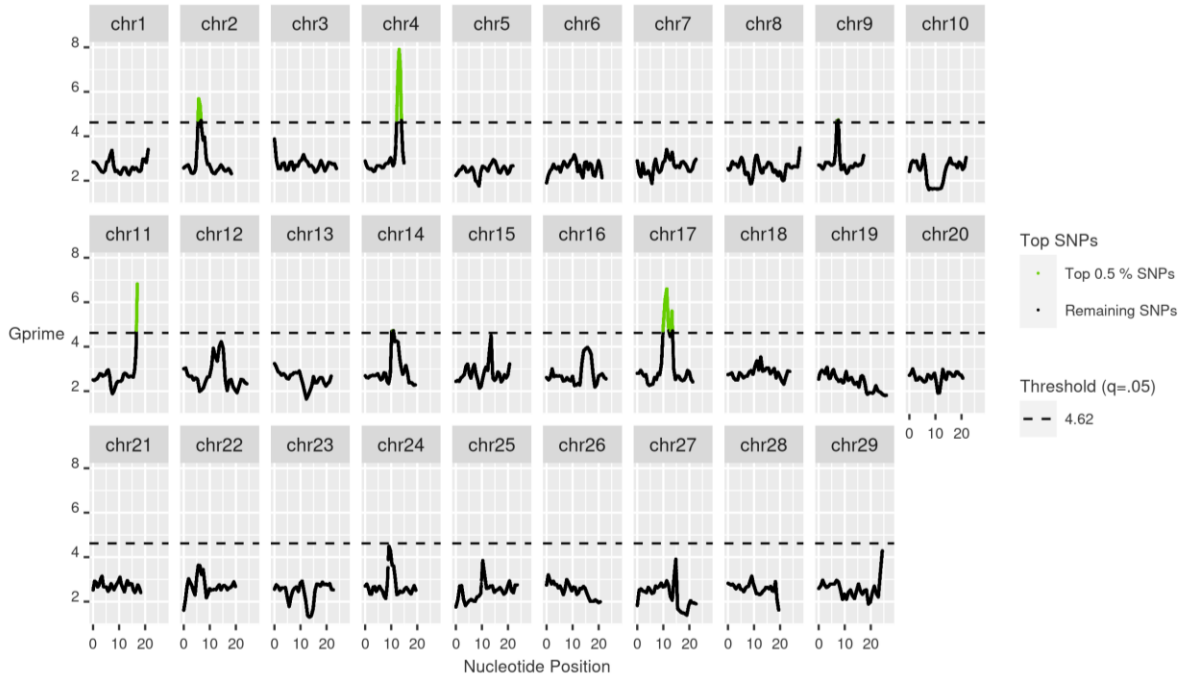


Figure 5.5. Bulked segregant analysis, B1, between sample bulk P1 and a bulk of parental whole genome sequence (WGS) data contributing to P1. Each analysed single nucleotide polymorphism site presents as a Gprime value analysed in a 1-Mb sliding window. Higher Gprime values reflect a higher frequency of P1 bulk alleles at certain genome positions than the bulk of parental WGS data. Green points represent the top 0.5 % of Gprime values, with

the remaining points coloured black. The threshold of the adjusted p = 0.05 is shown as a dashed line. Quantitative trait loci peaks for B1 were found on Chromosome 2 at 5.45 Mb, Chromosome 4 at 12.1 Mb, Chromosome 9 at 7.35 Mb, Chromosome 11 at 16.65 Mb, Chromosome 14 at 10.65, and Chromosome 17 at 10 Mb.

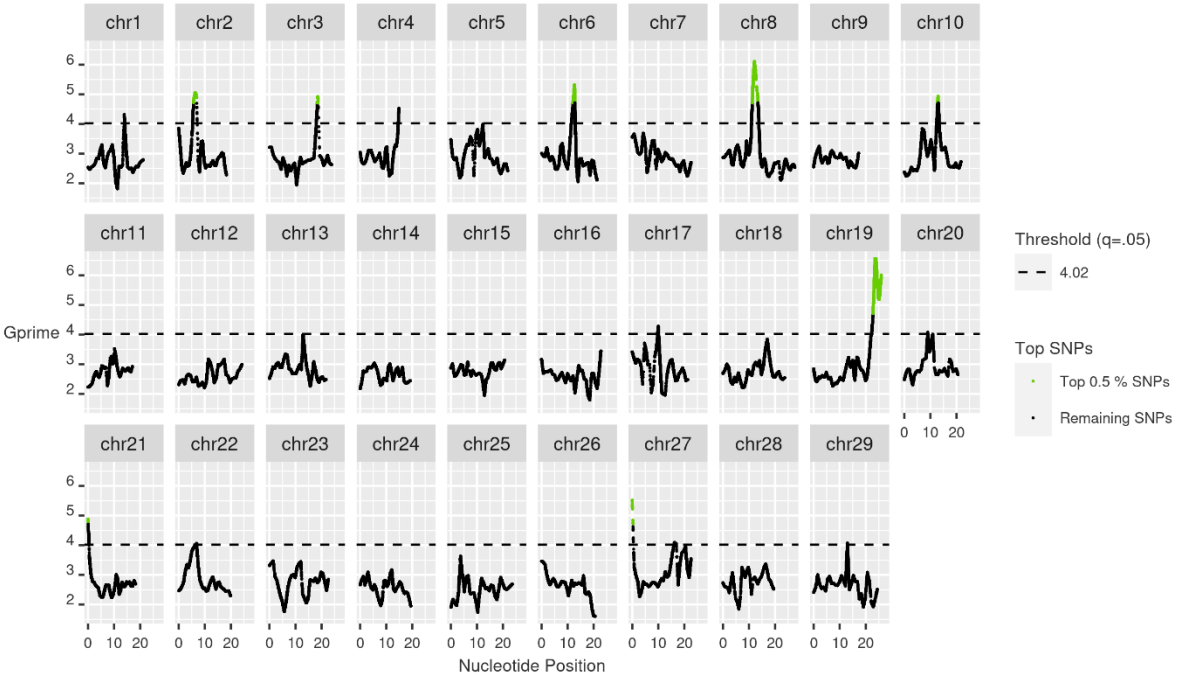


Figure 5.6. Bulked segregant analysis, B2, between sample bulk P2 and a bulk of parental genome sequence (WGS) data contributing to P2. Each analysed single nucleotide polymorphism site presents as a Gprime value analysed in a 1-Mb sliding window. Higher Gprime values reflect a higher frequency of P1 bulk alleles at certain genome positions than the bulk of parental WGS data. Green points represent the top 0.5 % of Gprime values, with the remaining points coloured black. The threshold of adjusted p = 0.05 is shown as a dashed line. Quantitative trait loci peaks for B2 were found on Chromosome 2 at 5.67 Mb, Chromosome 3 at 18.42 Mb, Chromosome 6 at 12.12 Mb, Chromosome 8 at 11.25 Mb, Chromosome 21 at 0.01 Mb, and Chromosome 27 at 0.01 Mb. The threshold of significance was lower than the top 0.5 % of SNPs.

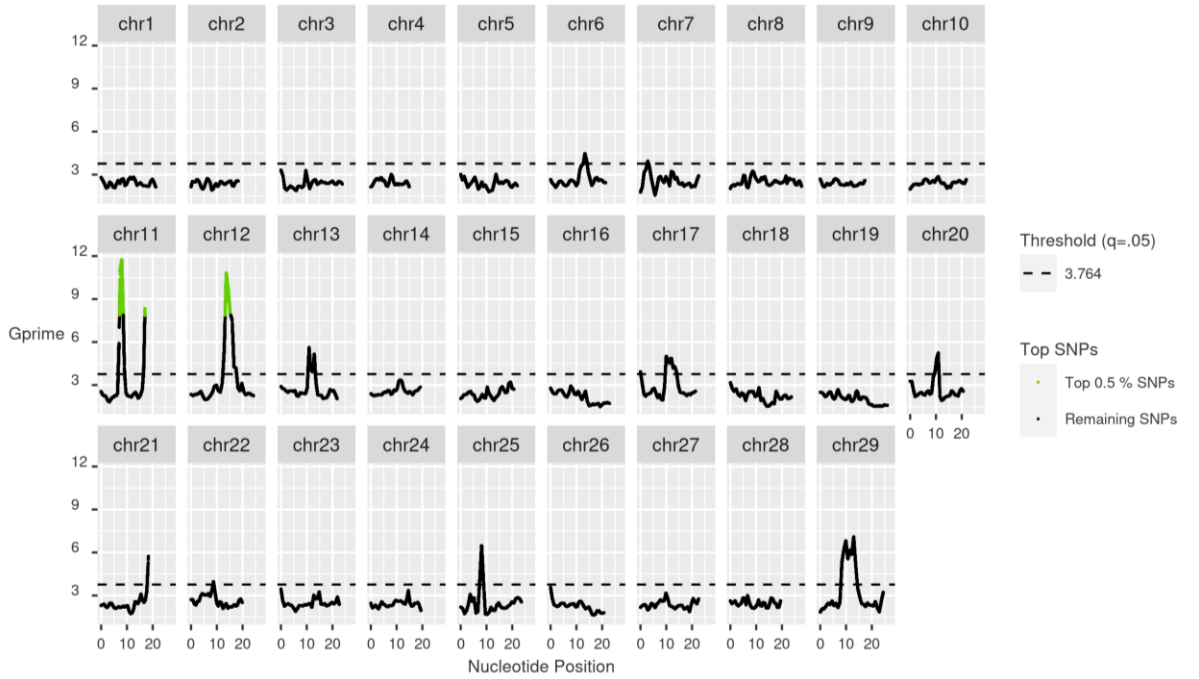


Figure 5.7. Bulked segregant analysis, B3, between sample bulk P3 and a bulk of parental genome sequence (WGS) data contributing to P3. Each analysed single nucleotide polymorphism site presents as a Gprime value analysed in a 1-Mb sliding window. Higher Gprime values reflect a higher frequency of P1 bulk alleles at certain genome positions than the bulk of parental WGS data. Green points represent the top 0.5 % of Gprime values, with the remaining points coloured black. The threshold of adjusted p = 0.05 is shown as a dashed line. Three quantitative trait loci (QTL) peaks for B3 were found: two on Chromosome 11 peaking at 7.00 Mb and 16.95 Mb, and one at 13.22 Mb on Chromosome 12. The threshold of significance was lower than the top 0.5 % of SNPs for this analysis, reflecting the strength of QTL on Chromosomes 11 and 12.

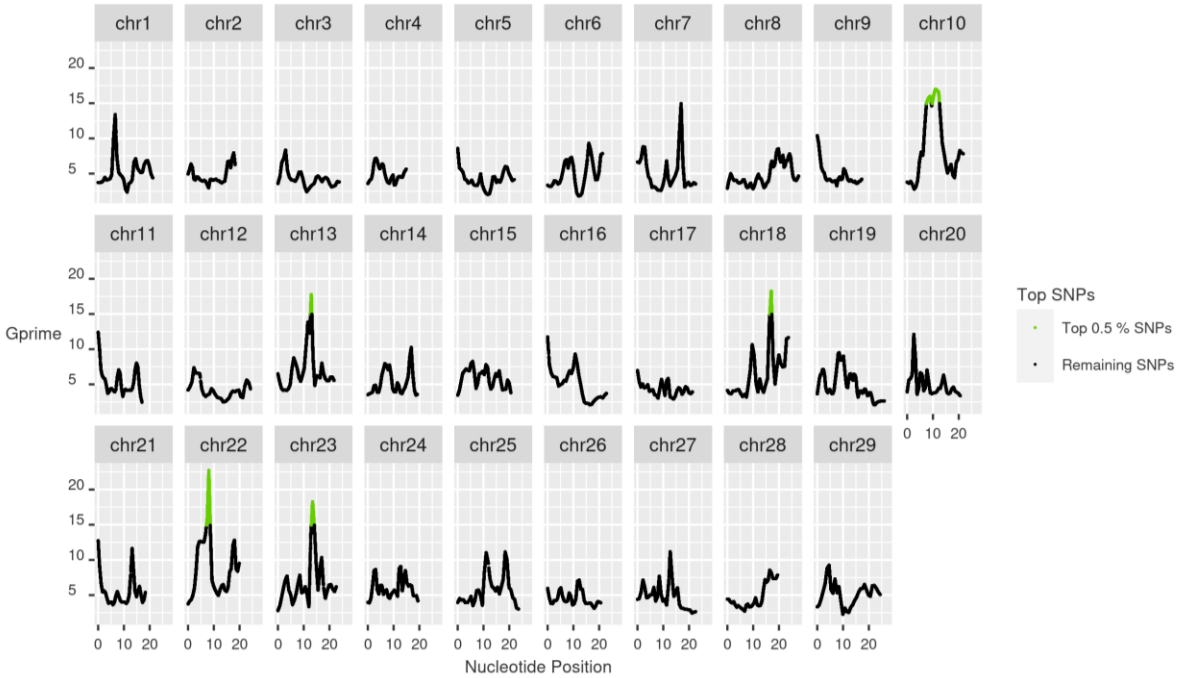


Figure 5.8. Bulked segregant analysis, B4, between sample bulk P4 and a bulk of parental genome sequence (WGS) data contributing to P4. Each analysed single nucleotide polymorphism site presents as a Gprime value analysed in a 1-Mb sliding window. Higher Gprime values reflect a higher frequency of P1 bulk alleles at certain genome positions than the bulk of parental WGS data. Green points represent the top 0.5 % of Gprime values, with the remaining points coloured black. Six Quantitative trait loci peaks were found in B4: on Chromosome 7 at 16.91 Mb, on Chromosome 10 at 10.0 Mb, on Chromosome 13 at 12.95 Mb, on Chromosome 18 at 16.45 Mb, on Chromosome 22 at 7.17 Mb, and on Chromosome 23 at 12.97 Mb. The significance threshold was not presented for this bulk comparison as it was much higher than the top 0.5 % of SNPs.

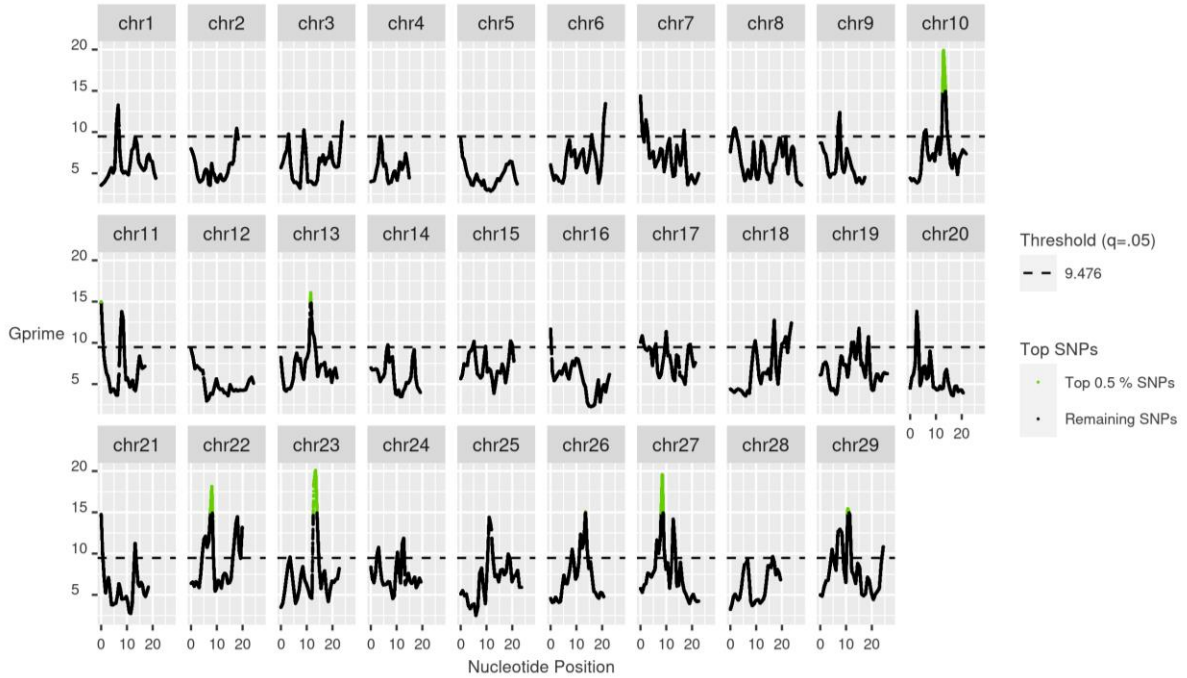


Figure 5.9. Bulked segregant analysis, B9, between sample bulk P9 and a bulk of parental genome sequence (WGS) data contributing to P9. Each analysed single nucleotide polymorphism site presents as a Gprime value analysed in a 1-Mb sliding window. Higher Gprime values reflect a higher frequency of P1 bulk alleles at certain genome positions than the bulk of parental WGS data. Green points represent the top 0.5 % of Gprime values with the remaining points coloured black. The threshold of adjusted p = 0.05 is shown as a dashed line. Eight Quantitative trait loci peaks were found in B9: on Chromosome 10 at 12.61 Mb, on Chromosome 11 at 0.01 Mb, on Chromosome 13 at 11.55 Mb, on Chromosome 22 at 7.53 Mb, on Chromosome 23 at 12.56 Mb, on Chromosome 26 at 13.52 Mb, on Chromosome 27 at 8.02 Mb, and on Chromosome 29 at 10.64 Mb. The significance threshold was much lower than the top 0.5 % of SNPs.

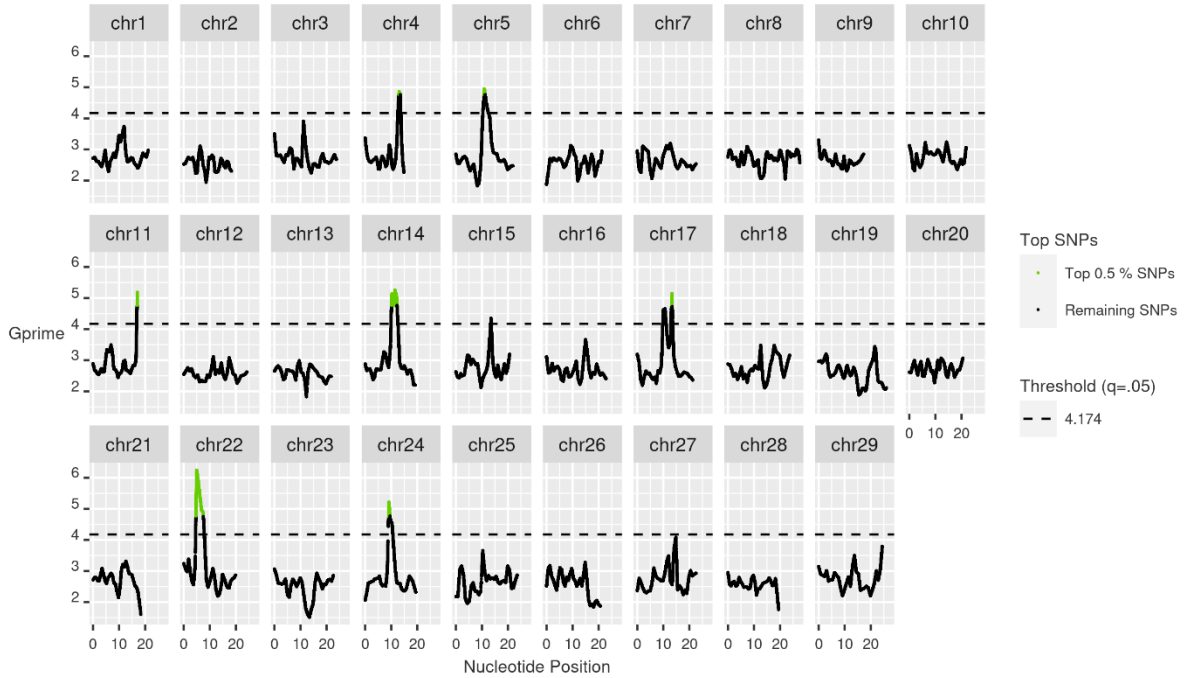


Figure 5.10. Bulked segregant analysis, B10, between sample bulk P10 and a bulk of parental genome sequence (WGS) data contributing to P10. Each analysed single nucleotide polymorphism site presents as a Gprime value analysed in a 1-Mb sliding window. Higher Gprime values reflect a higher frequency of P1 bulk alleles at certain genome positions than the bulk of parental WGS data. Green points represent the top 0.5 % of Gprime values, with the remaining points coloured black. The threshold of adjusted p = 0.05 is shown as a dashed line. Seven Quantitative trait loci peaks were found in B10: on Chromosome 4 at 12.96 Mb, on Chromosome 5 at 10.81 Mb, Chromosome 11 at 16.90 Mb, on Chromosome 14 at 10.11 Mb, on Chromosome 17 at 13.15 Mb, on Chromosome 22 at 4.67 Mb, and on Chromosome 24 at 8.99 Mb. The significance threshold was lower than the top 0.5 % of SNPs.

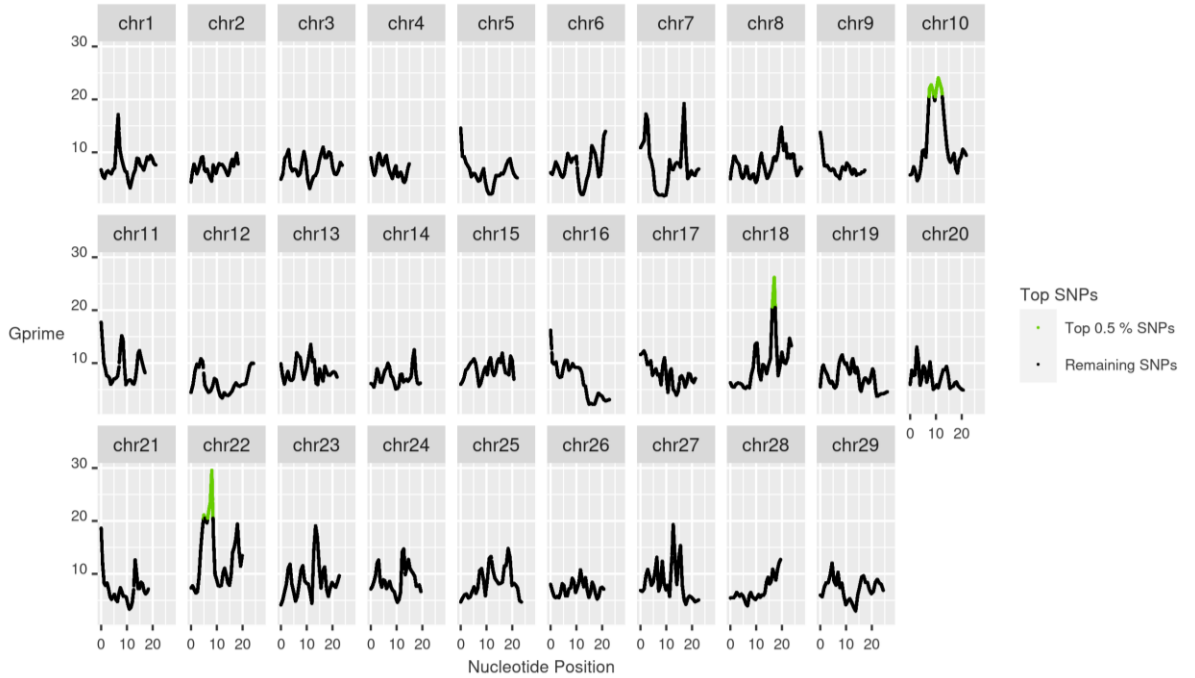


Figure 5.11. Bulked segregant analysis, B11, between sample bulk P11 and a bulk of parental genome sequence (WGS) data contributing to P11. Each analysed single nucleotide polymorphism site presents as a Gprime value analysed in a 1-Mb sliding window. Higher Gprime values reflect a higher frequency of P1 bulk alleles at certain genome positions than the bulk of parental WGS data. Green points represent the top 0.5 % of Gprime values, with the remaining points coloured black. Three Quantitative trait loci peaks were found in B11: on Chromosome 10 at 10.0 Mb, on Chromosome 18 at 16.27 Mb, and Chromosome 22 at 4.84 Mb. The threshold of significance was not presented in this bulk comparison as it was much higher than the top 0.5 % of SNPs.

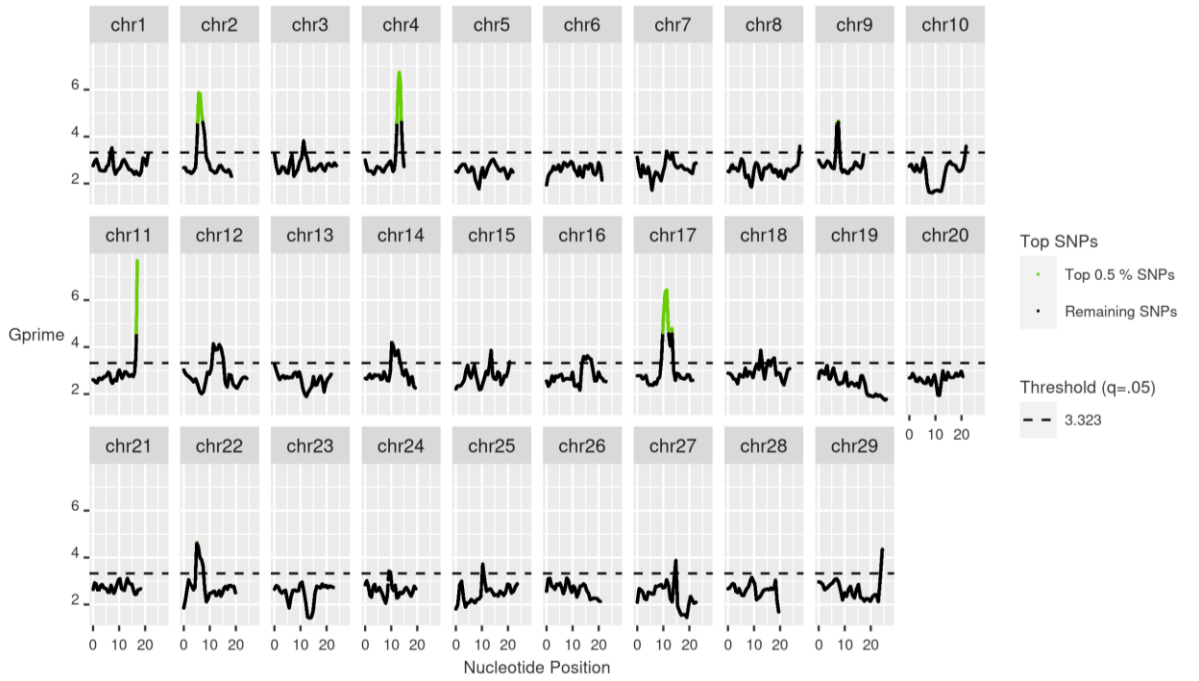


Figure 5.12. Bulked segregant analysis, B12, between sample bulk P12 and a bulk of parental genome sequence (WGS) data contributing to P12. Each single nucleotide polymorphism site presents as a Gprime value analysed in a 1-Mb sliding window. Higher Gprime values reflect a higher frequency of P1 bulk alleles at certain genome positions than the bulk of parental WGS data. Green points represent the top 0.5 % of Gprime values, with the remaining points coloured black. The threshold of adjusted p = 0.05 is shown as a dashed line. Six quantitative trait loci peaks were found in B12: on Chromosome 2 at 5.35 Mb, on Chromosome 4 at 12.22 Mb, on Chromosome 9 at 7.31 Mb, on Chromosome 11 at 16.60 Mb, on Chromosome 17 at 10.00 Mb, and on Chromosome 22 at 4.97 Mb. The significance threshold was much lower than the top 0.5 % of SNPs.

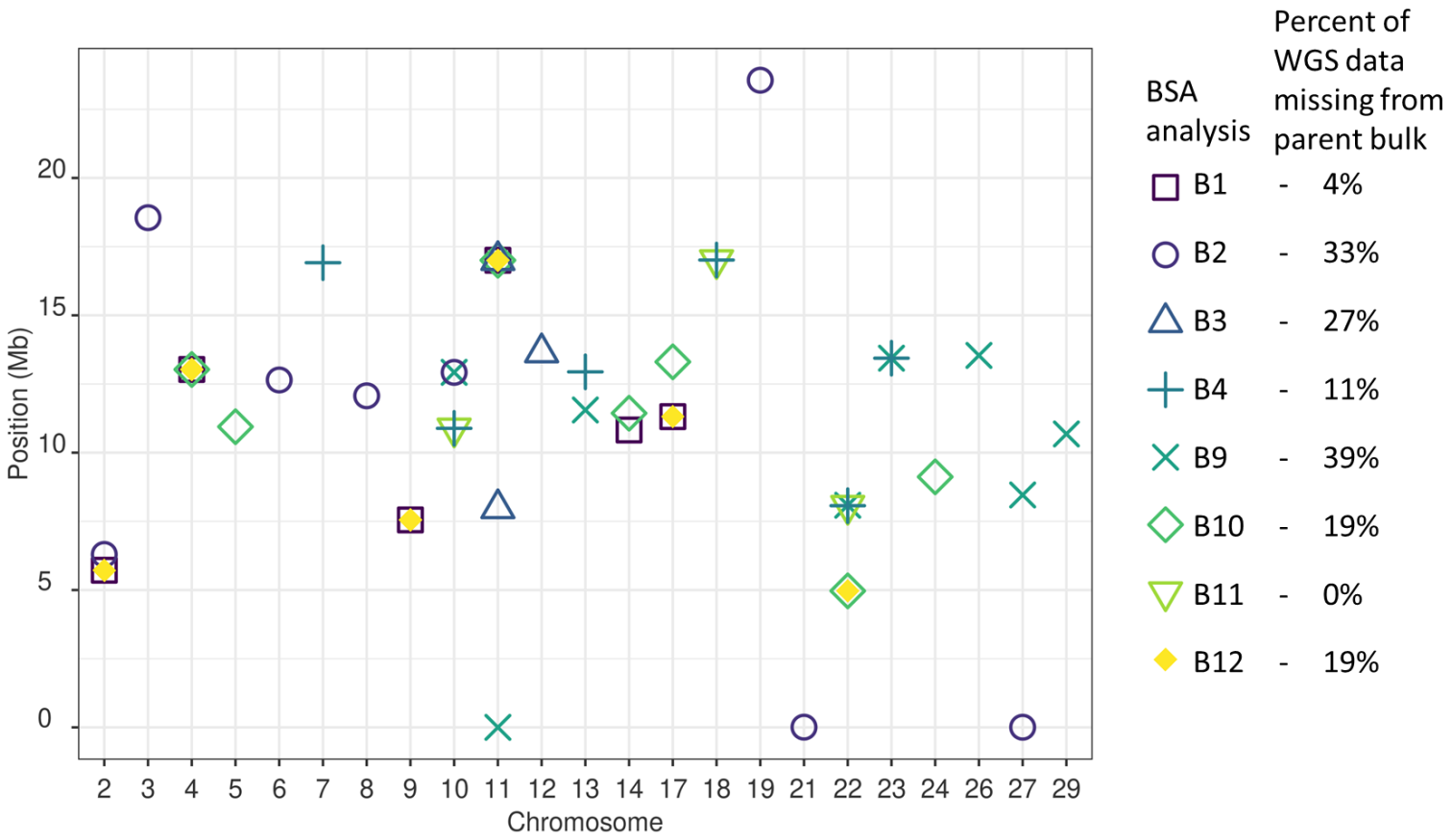


Figure 5.13. Bulk segregant analysis (BSA) of *Actinidia chinensis* quantitative trait loci (QTL) peak positions for *Pseudomonas syringae* pv. *actinidiae* (Psa) resistance between all analyses. Thirty QTLs were detected among the eight bulks analysed. The BSA presenting the most unique QTL were B2, B9, and B10. The BSA presenting no unique QTL were B11 and B12. A single QTL site was common between four BSA on Chromosome 11 at 16.95 Mb from B1, B3, B10 and B12. Four QTL sites were common between three BSA on chromosome 2 at 5.35 Mb, chromosome 4 at 13.02 Mb, Chromosome 17 at 11.31 Mb, and Chromosome 22 at 8.07 Mb. Six QTL sites were found in common between two BSA on Chromosome 9 at 7.55 Mb, Chromosome 10 at 10.89 and 12.92 Mb, Chromosome 14 at 11.13 Mb, Chromosome 17 at 11.31 Mb, Chromosome 18 at 17.01 Mb, Chromosome 22 at 4.97 Mb, and Chromosome 23 at 13.44 Mb. The 17 remaining QTL sites were found in a single BSA.

5.5 Discussion

Psa is one of the most destructive pathogens affecting kiwifruit, with a broad range of susceptibility and tolerance in *A. chinensis* var. *chinensis* genotypes. However, QTL for Psa resistance have been published within only two families to date (Tahir et al., 2020; Tahir et al., 2019), potentially leaving many loci for resistance to Psa undiscovered. Typically, QTL mapping methods would be used to investigate loci for a polygenic trait such as resistance to Psa (Jansen, 1996; Lefebvre & Palloix, 1996). However, accurately identifying traits influenced by more than one gene with QTL mapping is a costly and resource-intensive process requiring large replicated families specially developed for this purpose (Gupta et al., 2019; Soto-Cerda & Cloutier, 2012; Tahir et al., 2017; Tahir et al., 2019; Wisser et al., 2008). This project overcame the limitations of a typical QTL map by using bulks of diverse F_1 families in a modified BSA. This approach further increased the utility and cost-effectiveness of the typical BSA methods by analysing multiple small families in a single bulk, decreasing the sampling complexity, reducing the DNA extraction time and cost, reducing sequencing costs, and increasing the breadth of Psa resistance alleles that could be detected within a bulk.

Bulked DNA has been analysed using BSA methods in a wide range of species and traits (Dong et al., 2018; Li & Xu, 2022; Michelmore et al., 1991; Shen & Messer, 2022; Song et al., 2017; Sun et al., 2018; Xin et al., 2020). However, this project used BSA analysis methods to identify QTL for Psa resistance by measuring the shift in allele frequency of an *A. chinensis* var. *chinensis* population that had many individuals removed from established families due to Psa. The selective sweep of susceptible individuals provided a prime opportunity to measure the effect of this selection on the alleles that remained within those *A. chinensis* var. *chinensis* families. However, DNA was not captured from susceptible individuals for this study, which a typical BSA would use as the comparison bulk (Li & Xu, 2022; Michelmore et al., 1991). The lack of a susceptible bulk was compensated for in this modified BSA by using a bulk of parent WGS data as a population not exposed to selection pressure for Psa resistance. To my knowledge this is the first time a bulk of parents has been used as a substitute for a bulk in analyses done with BSA.

5.5.1 Testing the sensitivity of the modified methods

The first part of this project tested the ability of the modified BSA method to analyse the effects of Psa in kiwifruit by identifying a known locus through detecting a shift in allele frequency between bulks. This test was required because the families were sampled and extracted as bulks in a unique way (Figure 5.1). The modified method involved sampling leaves for a bulk from the field into a single bag and extracting DNA from all leaf samples within the bulk by first grinding them all together before DNA extraction. The sampling method involved taking a disc from each leaf of the same age for DNA extraction. This significantly increased the speed of sampling, sample tracking and DNA extraction, but variance among leaves could still occur due to differences in the number and size of cells, ratio of mitotic to interphase nuclei, or differing structures or biochemical composition of the plant cells (Kefelegn et al., 2021; Marsal et al., 2013). As a result, changes in allele frequency could have been created between samples due to a variable amount of DNA extracted from each leaf sample. The effect of the sampling methods and their integration with the analysis methods on QTL detection were tested by making bulks with a known percentage of males that differed between bulks. The shift in allele frequency, caused by the variable number of males between bulks at the region around the male-specific *Shy Girl* gene (Akagi et al., 2018) indicated that the minimum shift in allele frequency that could be detected between bulks was 10 % (Figure 5.4). The detection of an allele frequency shift of over 10 % was effective for detecting QTL under strong selection for Psa resistance for the second part of the project.

The modified technique tested in the first part of this project included the pooling of multiple families, but the threshold of detection for QTL also needed to be altered to fit the altered methods. To retrieve QTL peaks that indicated a significant shift in allele frequency between bulk pairs in the modified BSA, the peaks that contained the top 0.05 % of Gprime values were used instead of the adjusted p = 0.05 statistic adjusted for repeated measures. The use of the top 0.05 % of Gprime values increased the accuracy of detection of the sex locus between bulks with a variable number of males by 33.33 % compared to the adjusted p = 0.05 threshold value for the expected allelic variance of 20 - 50 %. This analysis allowed the detection of a 20 - 40 % shift in male allele frequency with an accuracy above 71 % (Figure 5.4), and a shift in allele frequency above 60 % could be detected with 100 % accuracy. This confirmed that the modified technique

could detect large changes in allele frequency. Because strongly selected alleles can cause a shift in allele frequency of up to 50 % within a family, this increased the confidence that sites of strong selection could be detected in the second part of the project.

The method of analysing the shift in frequency between bulks in the first part of the project may have restricted the power of the BSA to detect QTL caused by minor shifts in male allele frequency below 20 %. This is because the samples for the first part of the project were obtained from the diverse populations used in the second part of the project. The poor detection of minor shifts in allele frequency could have been caused by a larger shift in allele frequency between bulks at sites other than the sex-linked Chromosome 25. This may have happened because although the bulks with a differing percentage of males all had the same male parent, they included different female parents to make up the bulk. This was useful to get an idea about the allelic variance that would have to be managed in part two of the project, but it may have altered the allele frequency at certain sites between the bulks tested for a shift in male allele frequency. If the shift in allele frequency at these sites was greater than that between the number of males added to each pool, and the peak included over 0.05 % of the genome's alleles, the shift in male allele frequency would not be included in the top 0.05 % of Gprime values, and the peak would be missed using this method of QTL peak detection.

5.5.2 QTL for resistance to Psa

A typical BSA groups together loci that differ in allele frequency between sample bulks segregating for a trait of interest (Li & Xu, 2022; Michelmore et al., 1991). However, part two of this project modified the bulking strategy of a typical BSA. Each sample bulk was made up of multiple diverse families that had DNA extracted from all individuals in the bulk simultaneously as a single sample. The second part of this project also differed from a typical BSA because no individuals culled due to Psa infections were sampled to make a comparison bulk. Instead, a bulk of parental alleles was bioinformatically generated using WGS data from parents. The alleles with a higher frequency in the sample bulk compared to the parent bulk were assessed against the significance thresholds established in the preliminary trial to identify QTL for Psa resistance. Between the eight sample bulks that each contained 3-7 families, 30 QTLs for resistance to Psa were identified. Twelve of the

30 QTLs were detected in more than one bulk, with one locus on Chromosome 11 at 16.95 Mb detected among four bulks.

QTL for Psa resistance have previously been published in a large diploid *A. chinensis* var. *chinensis* family (Tahir et al., 2019). That study identified two QTL using field scores for Psa resistance. One of the loci identified was in an identical location to the locus found in this project on Chromosome 22 at 4.967790 Mb from two bulks, B10 and B12. Because the parent P8 was used in the bi-parental population by Tahir et al. (2019), and the bulk B10 from this project also contained the parent P8, it seems likely that this locus for resistance to Psa is coming from parent P8. The bulk B12 also had a locus for Psa resistance on Chromosome 22 at 4.97 Mb. The other locus for Psa resistance detected by Tahir et al. (2019) using field scores, found on Chromosome 27 at 4.305319 Mb, was not detected in this study. The lack of detection of this locus may have been because the cultivar 'Hort16A' that identified as the parent that contributed Psa resistance to the bi-parental family was not included in this study. The parent of 'Hort16A', included in this study, named parent P13, was included in bulk B3 but the contribution of P13 to this bulk was low at 5.4 %. It would be expected to have a peak if the QTL was strongly selected for in the B3 bulk, but it is more likely that the other parent of 'Hort16A', CK15_01, was the contributor of the resistance QTL found by Tahir et al. (2019) on Chromosome 27 at 4.305319 Mb.

The commonality of QTL sites among some of the different BSA bulks may reflect the inclusion of common parents that contributed to those bulks. This commonality of QTL sites can give insight into which of the parents were likely to be contributing the alleles under selection at some of the QTL. Looking at the peaks which are at the same site on the same chromosomes in different bulks (Figure 5.13) shows the most likely parents that were contributing the alleles to those QTL. For example, the QTL on Chromosome 2, at 5.35 Mb, is shared between bulks B1, B2, and B12, indicating that parent P12 is likely to be the source of alleles in higher frequency in those bulks (Figure 5.13). Similarly, the parent P4 is likely the source of the QTL on Chromosome 22 at 8.07 Mb and the parent P1 is likely the source of the QTL on Chromosome 4 at 13.02 Mb. The parent P1 also appears to be the contributing parent to the QTL on Chromosome 22 at 4.97 Mb. However, this prediction of the P1 parent contribution to QTL at 4.97 Mb on Chromosome 22 may be inaccurate since the P1 parent makes up only a small proportion of the bulk B10 (5.6%). Instead,

both the parents P6 and P8 may have contributed this QTL to these bulks. Parent P8 was used as one of the parents in a biparental mapping family for Psa resistance in a study by Tahir et al. (2019) and parents P6 and P8 are related. Tahir et al.'s study (2019) also identified the same QTL at 4.97 Mb on Chromosome 22 derived from parent P8. Eight other QTLs could have their parent contributors narrowed down to only two parents since both were shared between bulks and sites (Figure 5.14).

BSA QTL chromosome and position	Bulk name	Parents contributing to bulks																					
		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
Chr. 2, 5.35 Mb	B1	P1									P10		P12						P18			P21	P22
	B2		P2							P9			P12				P16	P17		P19			P22
	B12	P1	P2				P6						P12										
Chr. 4, 13.02 Mb	B1	P1									P10		P12						P18			P21	P22
	B10	P1				P5			P8		P10												
	B12	P1	P2				P6						P12										
Chr. 9, 7.55 Mb	B1	P1									P10		P12						P18			P21	P22
	B12	P1	P2				P6						P12										
Chr. 10, 10.89 Mb	B4				P4					P9		P11			P14	P15				P19			
	B11			P3	P4							P11											
Chr. 10, 12.92 Mb	B2		P2							P9			P12				P16	P17		P19			P22
	B9		P2		P4		P6	P7		P9													
Chr. 11, 17.00 Mb	B1	P1									P10		P12						P18			P21	P22
	B3			P3								P11		P13					P18		P20		
	B10	P1				P5			P8		P10												
	B12	P1	P2				P6						P12										
Chr. 14, 11.13 Mb	B1	P1									P10		P12						P18			P21	P22
	B10	P1				P5			P8		P10												
Chr. 17, 11.13 Mb	B1	P1									P10		P12						P18			P21	P22
	B12	P1	P2				P6						P12										
Chr. 18, 17.01 Mb	B4				P4					P9		P11			P14	P15				P19			
	B11			P3	P4							P11											
Chr. 22, 4.97 Mb	B10	P1				P5			P8		P10												
	B12	P1	P2				P6						P12										
Chr. 22, 8.07 Mb	B4				P4					P9		P11			P14	P15				P19			
	B9		P2		P4		P6	P7		P9													
	B11			P3	P4							P11											
Chr. 23, 13.44 Mb	B4				P4					P9		P11			P14	P15				P19			
	B9		P2		P4		P6	P7		P9													

Figure 5.14. Chromosomes and sites with more than one QTL for resistance to Psa in common between analyses. The parents contributing alleles to each QTL can be determined for some peaks by analysing the families that contribute to each bulk. The QTL, highlighted in green, from Chromosome 4 was likely from the parent P1. But the indication of parent P1 being the main contributor to QTL peaks on Chromosome 14 at 11.13 Mb and Chromosome 22 at 4.97 Mb may be misleading as the bulk B10 had a low contribution from parent P1. This may not have had a strong enough signal to present as a peak unless the loci were shared with another parent such as

P12. The QTL on Chromosome 2 at 5.35 Mb was likely from parent P12, and the peak on Chromosome 22 at 8.07 Mb was likely from parent P4. Other parents contributing to quantitative trait loci (QTL) in light green had two individuals that could have contributed to the QTL. Chromosome 11 had no parental contributors to bulks in common with the QTL at that position, despite four bulks presenting QTL at that site. It is possible that the parents, P1 and P18, contained the same alleles for Psa resistance on Chromosome 11 at 16.95 Mb.

		Parents contributing to bulks																				
		P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**	**P16**	**P17**	**P18**	**P19**	**P20**	**P21**	**P22**
Bulk name	**B1**	P1									P10		P12						P18			P21	P22
	B2		P2							P9			P12				P16	P17		P19			P22
	B3			P3								P11		P13					P18		P20		
	B4				P4					P9		P11			P14	P15				P19			
	B9		P2		P4		P6	P7		P9													
	B10	P1				P5			P8		P10												
	B11			P3	P4							P11											
	B12	P1	P2				P6						P12										

Figure 5.15. Bulks of *A. chinensis* with unique parent contributors to each bulk. Parents that were represented in a single pool are highlighted in green.

Where a single parent contributed to a QTL site, the parent that contributed to that QTL from the bulk should be unique to that bulk. Therefore, the QTL peaks on Chromosomes 11 at 0.36 Kb, 13 at 11.55 Mb, 26 at 9.12 Mb, 27 at 8.46 Mb and 29 at 10.68 Mb from bulk B9 were likely contributed by the parent P7. The remaining QTL sites from bulks B2, B3, B4 and B10 each had two unique parents that likely contributed to the detected QTL: namely, parents P16 or P17 contributed to Chromosomes 3 at 18.56 Mb, 6 at 12.65 Mb, 8 at 12.07 Mb, 19 at 23.56 Mb, 21 at 7.45 Kb, and 27 at 4.76 Kb, parents P13 or P17 to Chromosomes 11 at 7.97 Mb and 12 at 13.65 Mb, parents P13 or P20 to Chromosomes 7 at 16.92 Mb and 13 at 13.65 Mb from bulk B3, and parents P5 or P8 likely contributed to QTL on Chromosomes 5 at 10.95 Mb, 17 at 13.31 Mb and 24 at 9.12 Mb from bulk B10. It was reassuring that the B11 and B12 bulks, which had no parents unique to the bulk, had no unique QTL sites attributed to them.

Further information about parent contributors to QTL can be gained by identifying the parents in common among bulks that contributed to a QTL. This approach identified the likely parental contributors to three QTL on Chromosomes 2, 4, and 22 from parents P1, P4 and P12, respectively (Table 5.3).

Table 5.3. The commonality of *A. chinensis* parents contributing to quantitative trait loci (QTL) positions among bulks. The parents that contributed to specific QTL can be inferred where multiple bulks have QTL at the same site that shared parents among those bulks. Parent P1 was the likely contributor to QTL on Chromosome 4. Parent P4 was the likely contributor to QTL on Chromosome 22, and Parent 12 was the likely contributor to QTL on Chromosome 2.

Parent contributing to bulks	**P1**	**P4**	**P12**
Bulks containing parents	B1 B10 B12	B4 B9 B11	B1 B2 B12
QTL position	Chr. 4, 13.02 Mb	Chr. 22, 8.07 Mb	Chr. 2, 5.35 Mb

5.5.3 Effects of the selective sweep for Psa tolerance alleles

Within each sample bulk, the families that had more individuals surviving Psa infection contributed more DNA to the bulk compared to those with fewer surviving individuals included in the same bulk. When performing the BSA, bulks with a skewed family representation may have preferentially identified loci from the families with more individuals in the bulk. This is likely because a higher amount of DNA contributed to a site from a particular parent increases the read depth of a locus unique to that parent compared to the bulk of parents. This is what was expected for the resistant alleles, but the families that had fewer surviving individuals would be under-represented in the bulk and therefore the Gprime value may be lower for these loci. The lower Gprime value may be excluded as a locus of interest due to families with greater representation presenting higher Gprime values over a greater number of loci.

The individuals that contributed to bulks all survived the selective sweep caused by Psa infection. The selective sweep would have exerted strong selection for alleles linked to resistance loci, such as those at 16.95 Mb along Chromosome 11 in bulks 1, 3, 10, and 12. Conversely, the selective sweep would have significantly reduced the frequency of alternative alleles at those loci (Nielsen et al., 2005). With strong selection pressure for an allele from a parent contributing to the family, the other allele would be effectively eliminated from the population at that locus. However, changes in allele frequency can also be indirectly caused through genetic correlations from linkage disequilibrium (Barrett & Hoekstra, 2011; Kemppainen et al., 2017) and genetic drift (Conolly et al., 2008). The Gprime method of BSA was implemented to adjust for the effects of linkage disequilibrium (Magwene et al., 2011), but genetic drift could have skewed the results, particularly in families with poor representation in the bulk. This is because the families with poor representation in the bulk are also a poor representation of that family, which will predispose the alleles from these families to genetic drift (Magwene et al., 2011). Similar effects will have occurred at genomic regions linked to the QTL for Psa resistance (Robertson, 1970). However, because the effects of genetic drift are assumed to be random throughout the genome (Conolly et al., 2008), the effect of genetic drift on the results from this project are assumed to be minor.

Identifying the parents that contributed the Psa resistance loci to each bulk will help with family-based breeding strategies (Hinds et al., 2005). Although the families that contributed the largest number of individuals to bulks are likely to be those that are contributing the QTL detected for resistance in each bulk, these loci could be coming from one or many parents. An attempt was made to identify the parents contributing resistance to the sample bulks analysed, but a combination of the missing parental WGS data and the way the bulks were constructed limited the information available. This could be overcome by developing markers to target the loci with high Gprime values. The markers could then be used on DNA from parents to identify which parents contributed these loci and enable marker-assisted selection for Psa resistance in families related to these parents. Identifying parents that contributed causal loci to a bulk could also be done by identifying alleles of higher frequency from the BSA under the QTL that were private to a single parent (Hinds et al., 2005). However, this was not possible in this project because many of the parents did not have WGS data available (Figure 5.2). Identifying haplotypes for each parent

would also be beneficial by allowing the haplotype sequence of each parent to be matched with loci with high Gprime values under QTL. This would be informative for the parents of bulks B1, B4 and B11, but WGS data from parents P5, P6, P7, P16, P18, P19 and P22 would still be needed to generate haplotype sequences to identify the parents contributing to QTL in bulks B2, B3, B9, B10, and B12.

5.5.4 QTL detection accuracy of loci from parents that contributed a small or large percentage of DNA to bulks

Accurately detecting loci for resistance to Psa in the second part of the project was dependent on the percentage of alleles that each family contributed to the bulk. Families that contributed more than 20 % to a bulk and had strong allelic selection pressure on alleles unique to that family are likely to have those alleles present as a QTL in 72 % of analyses (Figure 5.4). But, families that had strong selection pressure on alleles unique to that family and contributed 10–20 % to a bulk were likely to present as a QTL in only 40 % of analyses. Families that contributed less than 10 % to a bulk were unlikely to present any QTL in the BSA, even with strong selection pressure on alleles unique to that family (Figure 5.4). Therefore, it is unlikely that families with poor representation in the sample bulk contributed to QTL in the second part of this project. However, these families with low contribution to bulks may contribute the same resistance loci as other parents included in the bulk, adding to the significance of those sites. The lack of representation from the families contributing less than 10 % to a bulk was due to the dilution created by other families that made up a bulk. For example, if an individual contributed 10 % to a bulk and 50 % of alleles were from one heterozygous parent under strong selection for Psa resistance, the resistance allele might be in all the individuals sampled from that family. Conversely, where resistance alleles are shared between families contributing to a bulk, their contribution to the sample bulk would stack, making their representation in the sample bulk 50 % higher than in the parental bulk.

5.5.5 Retrospective improvements to this project

Future projects could improve upon the methods used in the first part of the project. Sampling of individuals for the bulks containing different sex ratios should have been done on families where parental WGS data were available for all the parents contributing to the families used in each bulk.

If this were done when sampling in the first part of this project, the sample bulks with a differing number of males and females in each bulk could have been compared against the bioinformatically created bulk of parents to match the second part of this project. However, this was not done because some of the parents that contributed to these bulks did not have WGS data. Having an accurate test of the methods would alleviate the concern that the inferences made in the first part of this experiment reflect only the allelic variance included in the sampling and DNA extraction methodology and may not accurately reflect the influence of the BSA methods on identifying QTL for Psa resistance in the second part of the project.

If the second part of this project were repeated using the same material, families from bulks whose parents had no WGS data and the families that had a poor representation within bulks would not be included in bulks. The method of field sampling would also be altered by sampling each leaf into its own bag. In this project, sampling all leaves for a bulk into a single bag was much faster than labelling separate sample bags. However, upon storage in the -80°C freezer, some leaves shattered. This required reassembly of the leaf pieces to ensure that a single leaf disk from each leaf was sampled. In samples with more than one shattered leaf, incorrect reassembly may have led to sampling errors. Further, separating each leaf to ensure sample inclusion caused warming of leaves and may have led to poorer DNA quality and/or quantity where re-sampling was required.

Collecting leaf material from plants before being culled would create a better match to a typical BSA (Li & Xu, 2022) and allow the creation of a bulk of alleles that were being selected against instead of using a pool of parental WGS data as the comparison bulk. Also, collecting leaf samples from all plants before they were affected by Psa would enable the creation of an unselected bulk, a negatively selected bulk, and a positively selected bulk. Comparing the positively selected bulk of individuals resistant to Psa against the unselected bulk may allow the identification of alleles associated with resistance to Psa. Comparing the negatively selected bulk of individuals that were removed because of Psa against the unselected bulk may allow the identification of alleles associated with susceptibility to Psa. Integrating the crossing and sampling plans would increase the accuracy of analyses because the unselected bulk would be a better representation of the alleles within the family than that of bulks of WGS data from parents.

5.5.6 Conclusions

The two parts of this project showed that finding multiple QTL for resistance to Psa can be achieved using BSA of bulks containing multiple families while greatly simplifying field sampling, DNA extraction, and reducing sequencing costs. To my knowledge, this is the first time DNA has been extracted as a bulk for a BSA instead of quantifying DNA from each individual separately and bulking the resulting DNA. This is also the first time BSA has been applied to bulks of families, where the comparison bulk was made up of a bioinformatically generated bulk of parental WGS data to identify loci affecting the target trait. Utilising the QTL for Psa resistance found in this project as selection criteria in breeding programmes may enable faster breeding of cultivars with greater resistance to Psa than without marker-assisted selection, and provide an opportunity to stack resistance loci to create a more robust resistance to Psa in future cultivars.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the student and the student's main supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the student's contribution as indicated below in the Statement of Originality.

Student name:	Casey Flay
Name and title of main supervisor:	Vaughan Symonds
In which chapter is the manuscript/published work?	Chapter 5
What percentage of the manuscript/published work was contributed by the student?	100% unless otherwise stated in text

Describe the contribution that the student has made to the manuscript/published work:

Some resources were existing at Plant and Food Research which the student used to start their work.

Please select one of the following three options:

- ○ The manuscript/published work is published or in press
  Please provide the full reference of the research output:
- ○ The manuscript is currently under review for publication
  Please provide the name of the journal:
- ☑ It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Student's signature:	Casey Flay	Main supervisor's signature:	Vaughan Symonds

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

Doctoral Research Committee
July 2022

6 General discussion

Kiwifruit is a billion dollar industry in New Zealand that is affected by various biotic threats (MPI, 2022). This study endeavoured to help this industry by working to understand the genomic architecture of kiwifruit resistance to three different biotic threats: the armoured scale insect (*Hemiberlesia lataniae*), the brown-headed leafroller moth (*Ctenopseustis obliquana*), and the bacterial pathogen *Pseudomonas syringae* pv. *actinidiae* (Psa).

In this thesis, resistance of *Actinidia chinensis* var. *chinensis* to *H. lataniae* was investigated in greatest detail. It was found that a single region of low recombination on Chromosome 10 was strongly associated with resistance in the populations studied. While the region of low recombination was a hindrance to gene identification, the markers used for fine mapping that are associated with the region could be utilised for marker-assisted selection. Association mapping and marker-assisted selection have been used extensively in major crops to breed for beneficial traits (Jighly et al., 2019; Voss-Fels et al., 2019). Gathering phenotypic data from plants is expensive and time-consuming. In this project, the identification of variants in the region of low recombination and the apparent monogenic nature of resistance to *H. lataniae* were used to predict the presence of *H. lataniae* resistance loci in individuals that were genotyped but not phenotyped. Future work to phenotype these individuals, to confirm the association of the resistance marker with the resistance phenotype, would be a valuable step to prove the marker-phenotype association. Confirming the marker phenotype association would allow testing of breeding parents and populations for resistance to *H. lataniae* conferred by this allele. This would give breeders a cheaper and easier means of identifying individuals resistant to *H. lataniae* and allow the preferential use of these for breeding purposes.

Using the knowledge and markers from diploid families allowed the transferral of the locus associated with resistance to *H. lataniae* from a diploid level to the tetraploid level, where the majority of commercial *A. chinensis* var. *chinensis* breeding is done. Additionally, because the parent resistant to *H. lataniae* was susceptible to Psa, it needed to be crossed with parents from commercial breeding programmes that had greater resistance to Psa to ensure its survival in the field. To achieve these goals, unreduced gametes from the parent contributing resistance to *H.*

lataniae were used to breed tetraploid offspring with resistance to *H. lataniae* and possibly resistance to Psa. The markers targeting *H. lataniae* resistance were used to select individuals carrying the haplotype for resistance to *H. lataniae*, but no markers were available for resistance to Psa when this population was reared.

To identify QTL for Psa resistance, another approach was sought that could utilise existing populations exposed to natural selection pressure exerted by Psa in the field. Bulked segregant analysis (BSA) was identified as a method that, with a bit of modification, could utilise the existing population to identify multiple QTL for resistance to Psa in multiple families faster, easier, and at a lower cost than that of standard association mapping. For example, the two association maps for Psa resistance completed to date used specially created families from a resistant parent crossed with a susceptible parent (Tahir et al., 2020; Tahir et al., 2019). These two experiments each required more than 229 genotypes with at least five replicates per genotype to be in the field for at least two years. With DNA extraction and analysis costs, the association mapping approach was more costly than the BSA, yet the BSA approach had sufficient accuracy to identify the same QTL as that from the diploid association map for Psa completed by (Tahir et al., 2019).

Experiments that reveal the genetic architecture for resistance to pests and pathogens rely on accurate phenotypic data. The resistance phenotypic data used in this thesis characterise the interaction between plant, pest and pathogen. These interactions are highly affected by the interaction of the plant genotype with its environment, and the interaction between the pest or pathogen genotype and environment. The influence of this variance can be seen in all the experiments in this thesis, with the greatest negative influence preventing QTL detection for resistance to *C. obliquana* (Chapter 4). Identifying the architecture for resistance to *C. obliquana* was not achieved in this work, even though the same family and markers identified QTL for resistance to *H. lataniae*. The lack of QTL detection for resistance to *C. obliquana* was attributed mainly to the bioassay lacking the accuracy to enable the association of the phenotype with the genotype. The likely polygenic nature of resistance to *C. obliquana* would have made it more difficult to detect QTL because the genes associated with resistance would not have a single strong association with phenotype, as seen with the likely monogenic resistance revealed for *H. lataniae*. The insect and plant rearing procedures may have also affected the plant insect interaction

phenotype. The *C. obliquana* and *H. lataniae* experiments both had issues with the rearing of insects. The *C. obliquana* population had a component missing from their diet, while the cross of wild plants resistant to *H. lataniae* suffered from contamination with the armoured scale *Aspidiotus nerii* and the parasitoid *Encarsia citrina* in the rearing colony. However, the likely monogenic nature of *H. lataniae* resistance allowed the identification of the erroneous phenotype where families failed to segregate in an expected pattern before extensive phenotyping had been completed. Conversely, the likely polygenic nature of *C. obliquana* resistance and its phenotypic presentation as a quantitative scale did not facilitate an early identification of issues with the phenotype. Before QTL can be discovered for resistance to *C. obliquana*, additional work will be needed to reduce the variance associated with the bioassay.

The accurate identification of genotype by environment effects on the phenotype was less of an issue when identifying QTL for resistance to Psa by BSA. This was because the experiment used the natural selection caused by Psa in the field as the cause of allelic segregation. Moreover, because the methods used for sampling significantly reduced the time and cost of sampling and reduced the time, cost and complexity of DNA extraction, the chance of genotyping error was decreased. However, after identifying QTL in this experiment, it was noted that the sampling design could have been improved to allow the identification of parents that contributed QTL to each bulk for breeding purposes.

Breeding using marker-assisted selection for resistance to *H. lataniae* will be much easier than breeding for resistance to Psa. One reason for this is because the haplotype region of resistance to *H. lataniae* has been well established in this work. In addition, the region of low recombination associated with resistance and the parthenogenic nature of *H. lataniae* in New Zealand (Hill et al., 2011), make it unlikely that the monogenic resistance identified in this haplotype will be broken. Conversely, resistance to Psa was shown to be a polygenic trait mediated by more than one QTL dispersed throughout the genome. The QTL discovered for Psa resistance in this thesis will need further work to identify individuals carrying resistance QTL.

Future breeding work for Psa resistance would benefit by filling in the missing parental WGS data from Chapter 5 that is available. Obtaining a more representative set of parental data would allow

the incorporation of these data into the bulks of parents for each bulk segregant analysis. This would allow the full parental contribution to each bulk to be analysed, removing error associated with a shift in allele frequency caused by comparing bulks of parents missing some parent data with the sample bulks. In addition, analysing a more representative set of the bulk parent WGS data will also allow the identification of parents that contributed alleles for resistance to Psa within each bulk via a PCA. Once the parents that contributed to the loci of interest are identified, markers can be made to identify these sites using populations heterozygous at the marker site. Future kiwifruit breeding projects will benefit from identifying whether the pedigree of parental individuals matched those used for the bulks in Chapter 5. If common pedigree were identified in the bulks, it would be worth genotyping F_1 families from these pedigree with the Psa resistance markers and phenotyping them for Psa resistance. This would indicate whether the markers for Psa resistance can confer greater resistance to individuals possessing them. A genomic selection approach could then be performed to select individuals from the F_1 families with greater Psa resistance. This approach effectively stacks resistance loci to impart greater resistance to Psa attack (McDonald & Linde, 2002; Mundt, 2014). The plants with greater resistance to Psa could then be bred with plants resistant to *H. lataniae* which also have elite fruit characteristics to make pest and disease resistant populations with high fruit quality.

Beyond kiwifruit breeding, developing resistance markers for use in any breeding program could be accelerated using the insights revealed by the work described in this thesis. An ideal approach to characterise traits for breeding purposes and have WGS data available for marker development would be the use of BSA. If a phenotype was segregating in a population, a small subset of the population could be used to make two bulks of DNA from the most resistant and least resistant phenotypic extremes. WGS data collected from each bulk should identify alleles segregating within the population between the two DNA bulks. The segregating sites should be associated with the trait of interest where other factors such as linkage drag or genetic drift do not significantly affect allele frequency. The WGS data analysed in this project will allow detailed information for the generation of accurate markers to target the trait of interest. The markers would then need to be validated in a wider set of phenotyped plants to verify their association with the trait of interest.

Utilising the information gathered in this thesis will also allow faster breeding for resistance to *H. lataniae* and resistance to Psa with marker-assisted selection. Marker-assisted selection was used here to select individuals with resistance to *H. lataniae*. Cultivars with resistance to *H. lataniae* and Psa would be greatly beneficial to organic growers and conventional growers alike. These cultivars would benefit the entire kiwifruit industry and save the industry in excess of NZD $154 million per annum in pest control and associated costs. The benefits of reduced chemical control for these pests and pathogen would also promote the sustainability of the kiwifruit industry in a market where these features are becoming more valued.

Analysing the genetics of families segregating for resistance with association mapping is an efficient way to identify genomic regions that contribute to these traits. This thesis focused on populations that segregate for resistance to biotic threats, but the methods applied here can also be applied to other diploid species to investigate any qualitative or quantitative traits of interest to breeders. In particular, the methods modified here to analyse bulks of multiple families that segregated for Psa resistance would interest breeders looking to incorporate traits from distantly related species. This is because families that segregate for a trait of interest, but which are too small to make an association map, can be bulked to identify QTL of interest. This method would be a significant advantage to future breeding programmes, particularly because the whole process can be done significantly more quickly and cheaply than standard association mapping.

7 Appendix

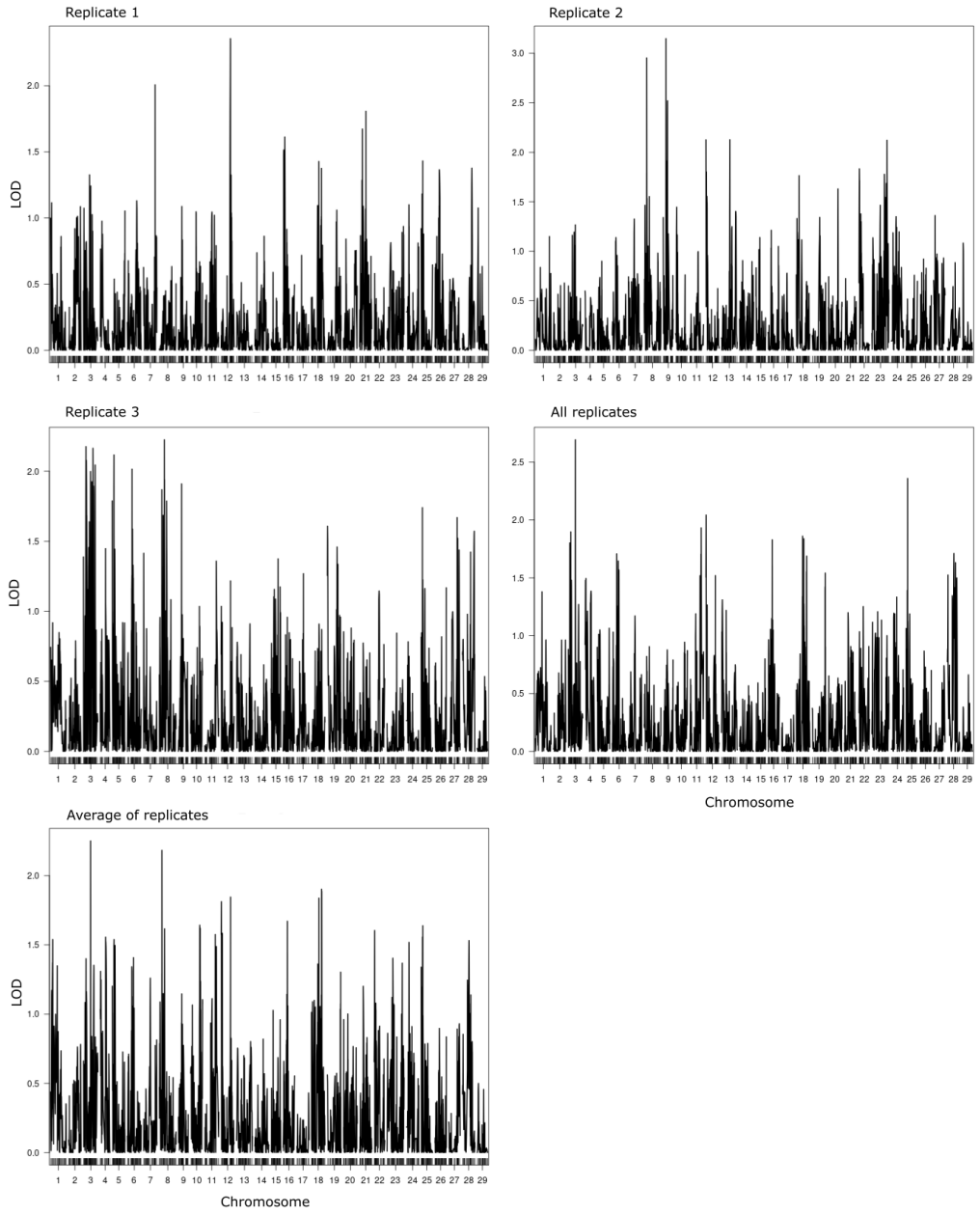


Figure 7.1. Survival of *C. obliquana* larvae on *A. chinensis* var. *chinensis* was measured seven days after being introduced to the cut leaf bioassay. No quantitative trait loci (QTL) above the threshold of significance were detected. Five separate QTL maps were made, including one for each of the three replicates, one for a map including all replicates together (All replicates), and one for a map that averaged phenotypes for each individual between replicates (Average of replicates).

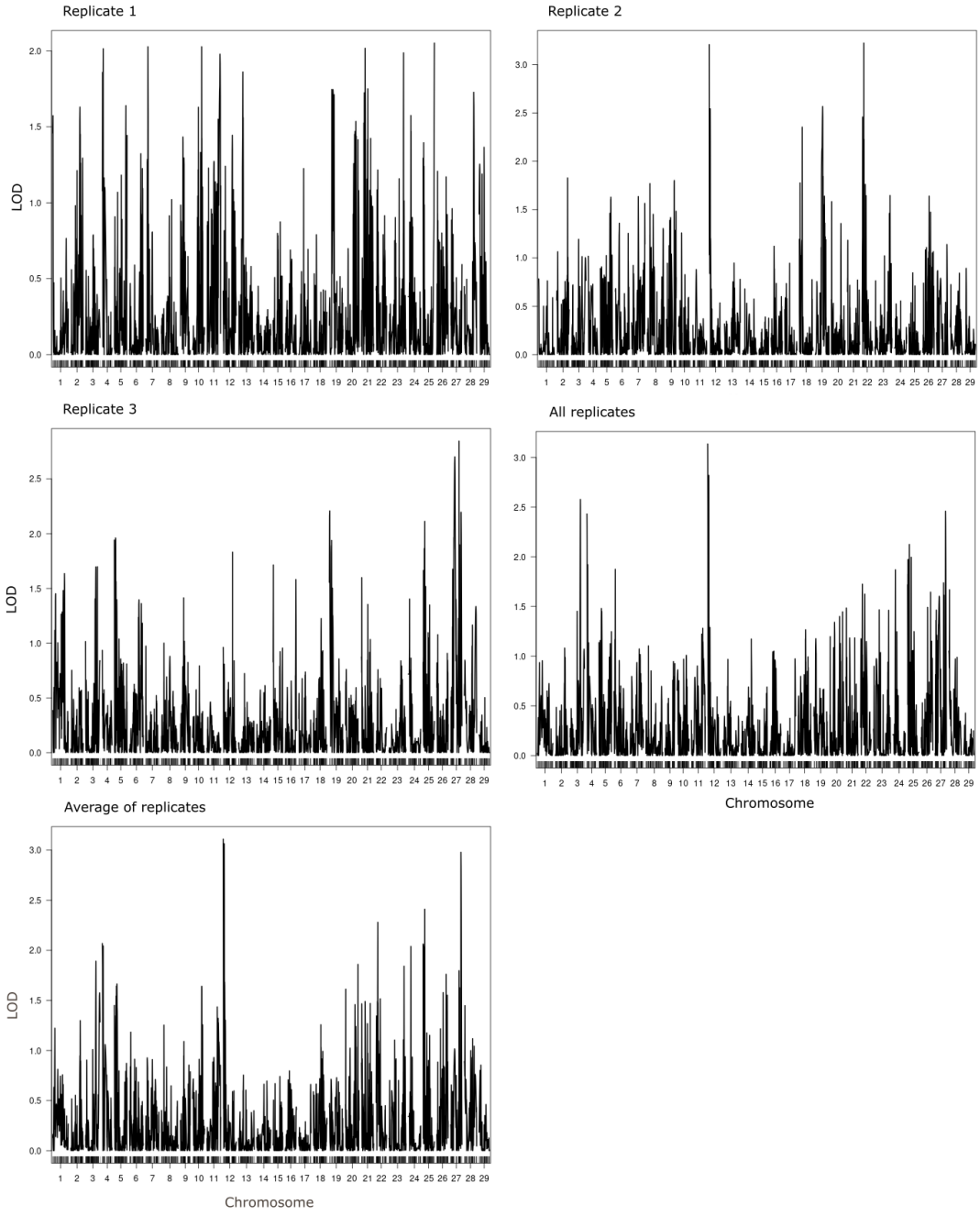


Figure 7.2. Survival of *C. obliquana* larvae on *A. chinensis* var. *chinensis* was measured 14 days after being introduced to the cut leaf bioassay. No quantitative trait loci (QTL) above the threshold of significance were detected. Five separate QTL maps were made, including one for each of the three replicates, one for a map including all replicates together (All replicates), and one for a map that averaged phenotypes for each individual between replicates (Average of replicates).

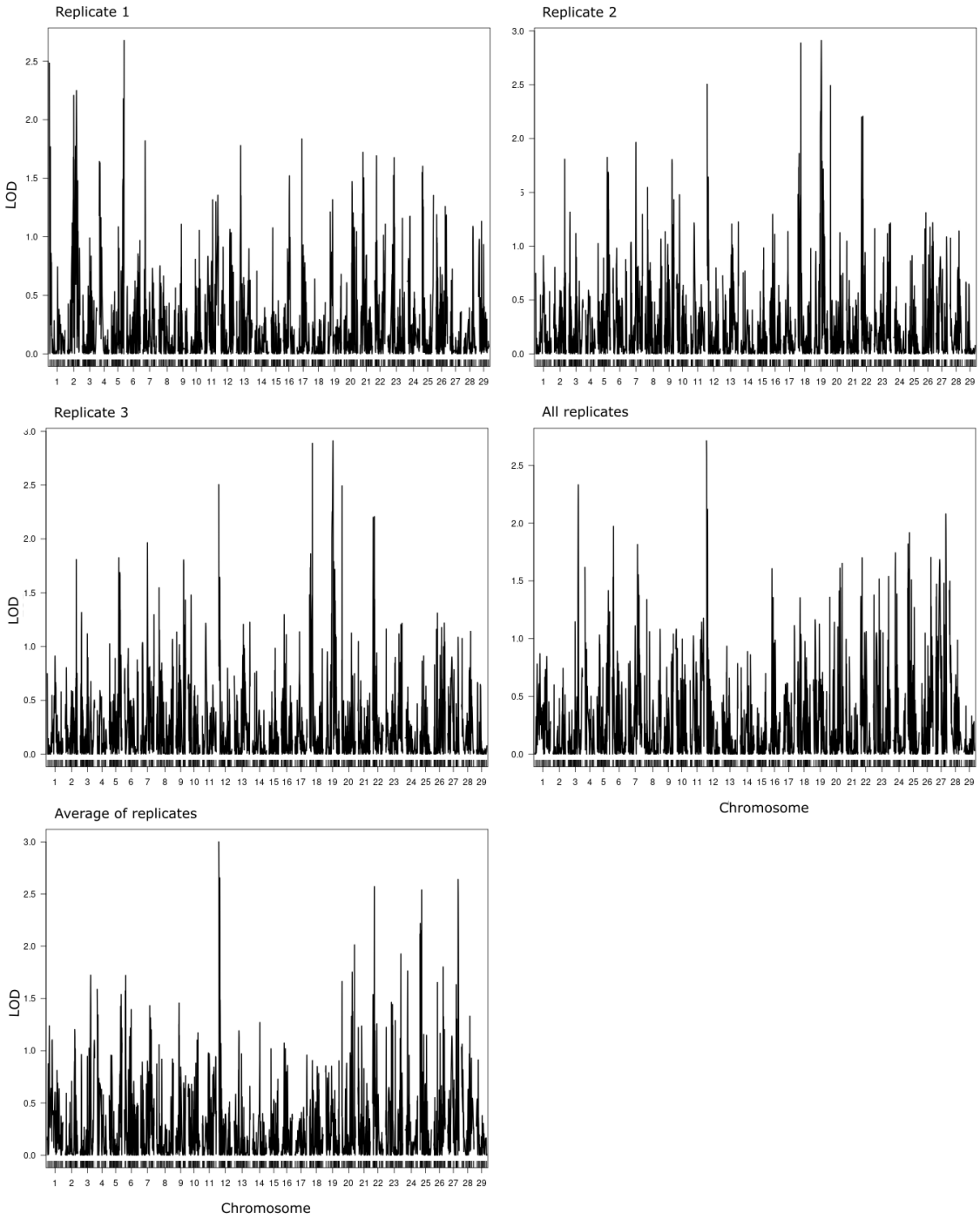


Figure 7.3. Survival of *C. obliquana* larvae on *A. chinensis* var. *chinensis* was measured 21 days after being introduced to the cut leaf bioassay. No quantitative trait loci (QTL) above the threshold of significance were detected. Five separate QTL maps were made, including one for each of the three replicates, one for a map including all replicates together (All replicates), and one for a map that averaged phenotypes for each individual between replicates (Average of replicates).

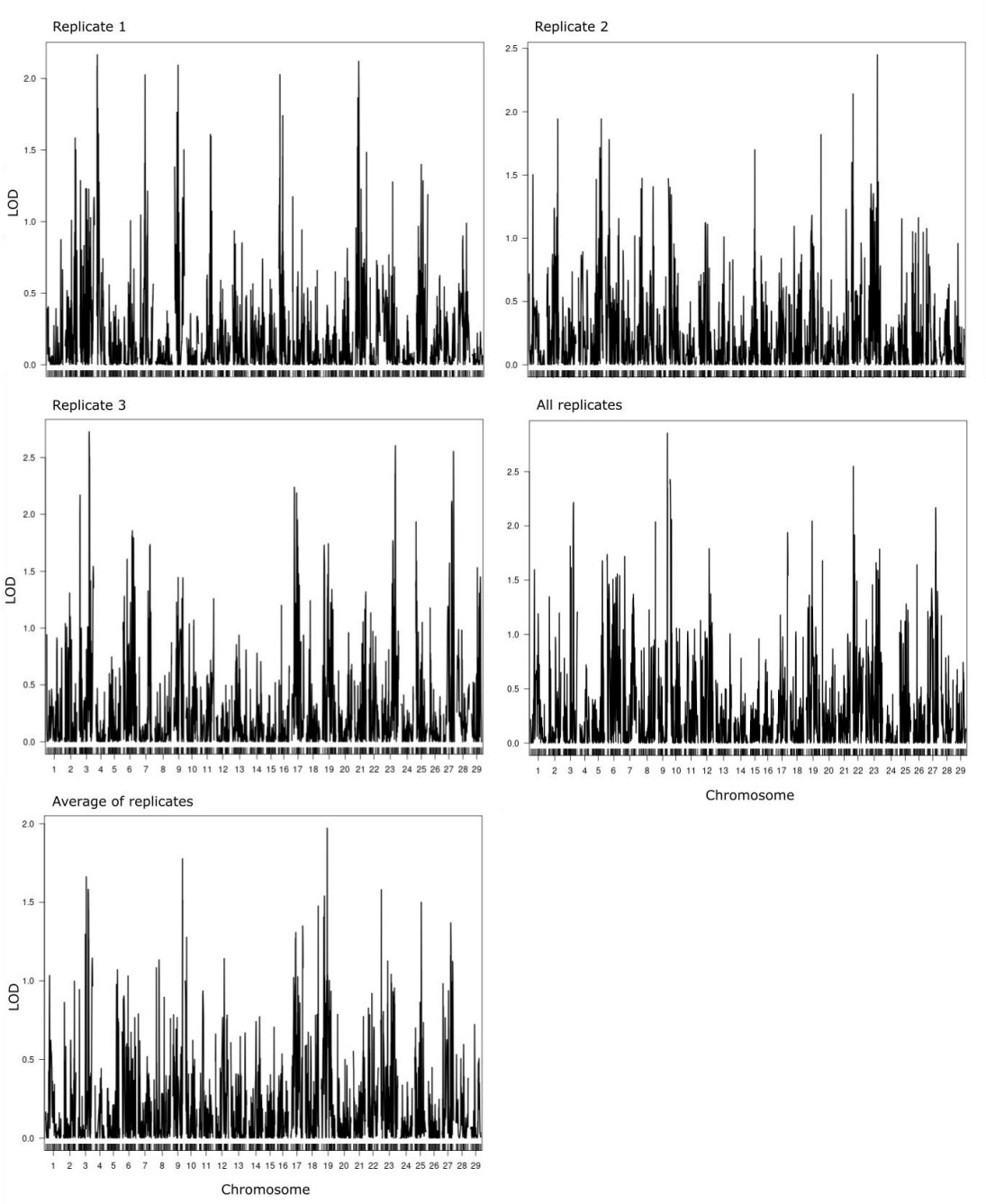


Figure 7.4. The average weight of surviving *C. obliquana* larvae on *A. chinensis* var. *chinensis* was measured 21 days after being introduced to the cut leaf bioassay. No quantitative trait loci (QTL) above the threshold of significance were detected. Five separate QTL maps were made, including one for each of the three replicates, one for a map including all replicates together (All replicates), and one for a map that averaged phenotypes for each individual between replicates (Average of replicates).

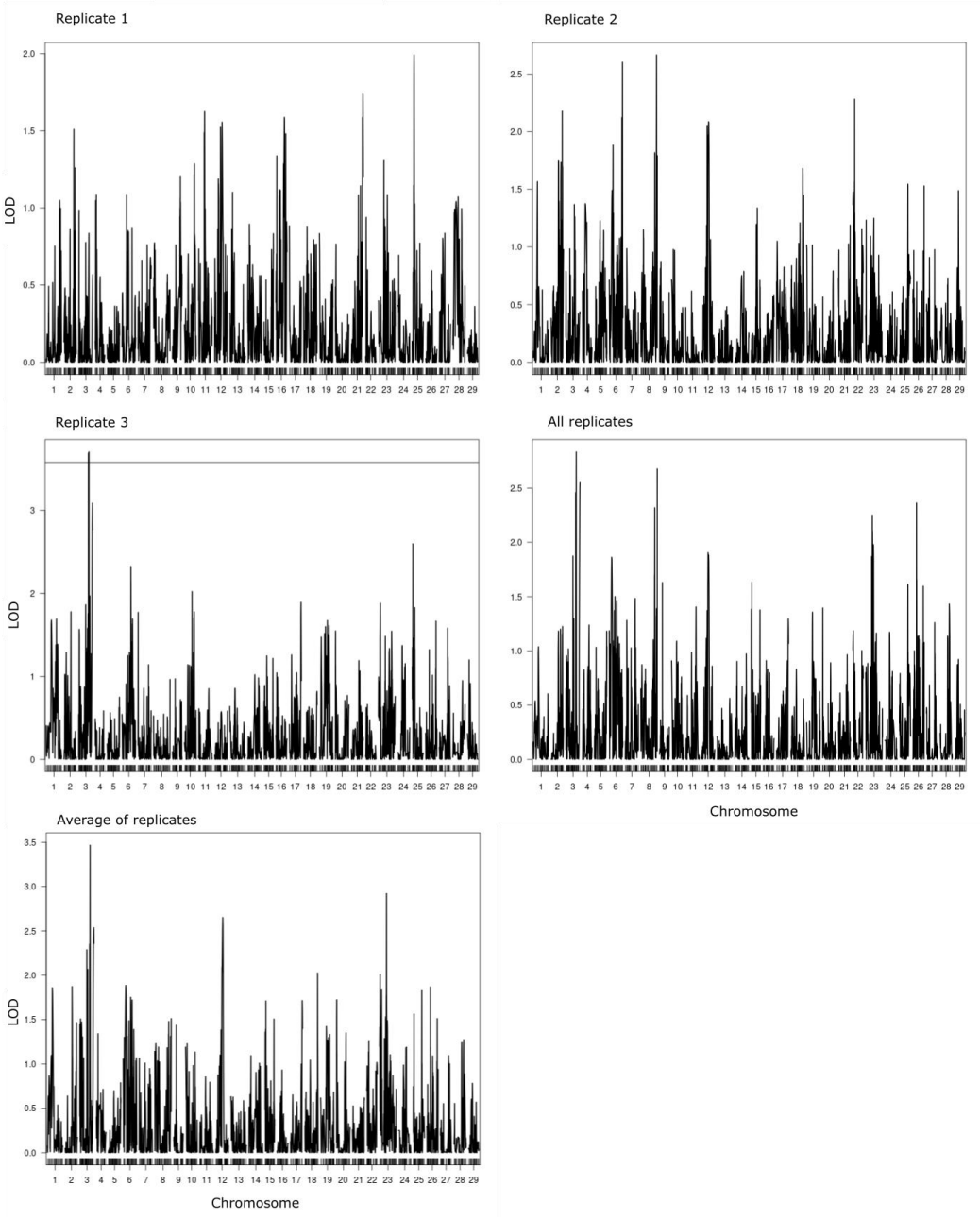


Figure 7.5. The average head width of surviving *C. obliquana* larvae on *A. chinensis* var. *chinensis* was measured 21 days after being introduced to the cut leaf bioassay. A single quantitative trait locus (QTL) was detected in replicate three on chromosome 3 above the threshold of significance. No other QTL were found above the threshold of significance. Five separate QTL maps were made, including one for each of the three replicates, one for a map including all replicates together (All replicates), and one for a map that averaged phenotypes for each individual between replicates (Average of replicates).

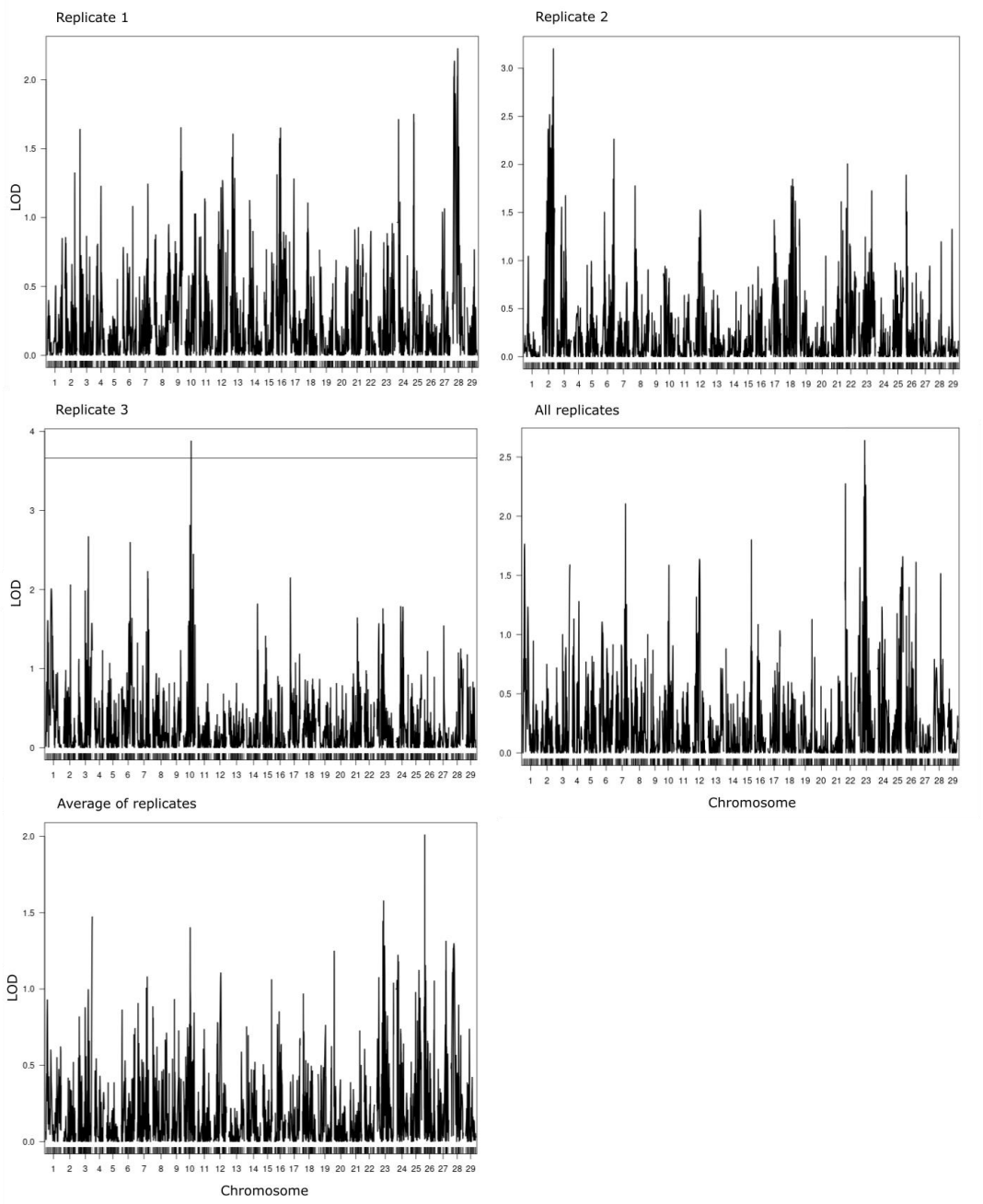


Figure 7.6. The average instar of surviving *C. obliquana* larvae on *A. chinensis* var. *chinensis* was measured 21 days after being introduced to the cut leaf bioassay. No quantitative trait loci (QTL) above the threshold of significance were detected. Five separate QTL maps were made, including one for each of the three replicates, one for a map including all replicates together (All replicates), and one for a map that averaged phenotypes for each individual between replicates (Average of replicates).

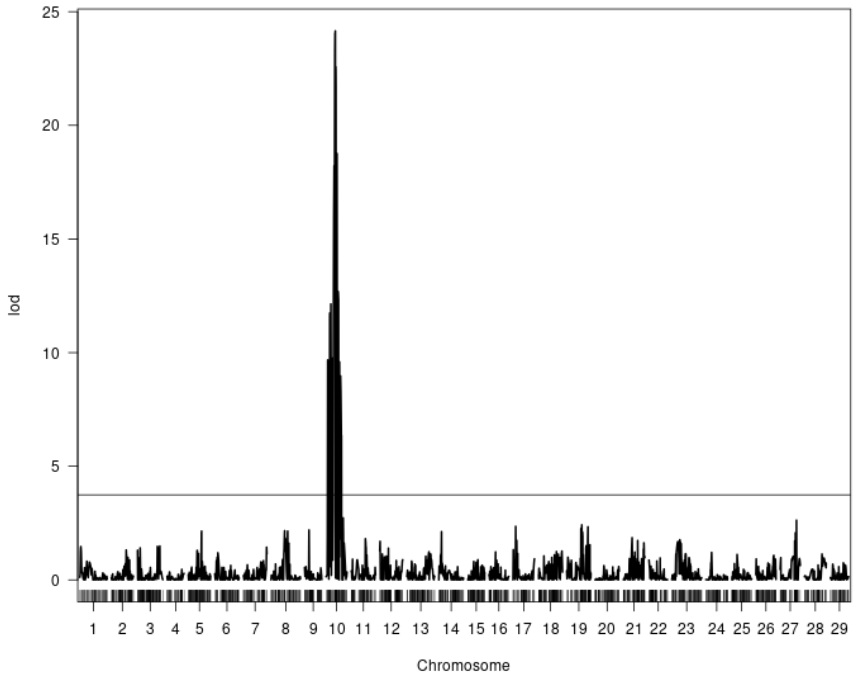


Figure 7.7. A genome wide Quantitative trait loci (QTL) map of the monogenic *H. lataniae* resistance phenotype on the B1 population of *A. chinensis* var. *chinensis*. A QTL was detected above the significance threshold on chromosome 10 with a LOD of 24.1.

8 References

Agnew, P., Koella, J. C., & Michalakis, Y. (2000). Host life history responses to parasitism. *Microbes and Infection*, *2*(8), 891-896.

Akagi, T., Henry, I. M., Ohtani, H., Morimoto, T., Beppu, K., Kataoka, I., & Tao, R. (2018). A Y-encoded suppressor of feminization arose via lineage-specific duplication of a cytokinin response regulator in kiwifruit. *The Plant Cell*, *30*(4), 780-795.

Aljbory, Z., Aikins, M. J., Park, Y., Reeck, G. R., & Chen, M. S. (2020). Differential localization of hessian fly candidate effectors in resistant and susceptible wheat plants. *Plant direct*, *4*(8), e00246.

Andermann, T., Torres Jiménez, M. F., Matos-Maraví, P., Batista, R., Blanco-Pastor, J. L., Gustafsson, A. L. S., Kistler, L., Liberal, I. M., Oxelman, B., & Bacon, C. D. (2020). A guide to carrying out a phylogenomic target sequence capture project. *Frontiers in Genetics*, *10*, 1407.

Aronesty, E. (2013). Comparison of sequencing utility programs. *The open bioinformatics journal*, *7*(1).

Balestra, G., Renzi, M., & Mazzaglia, A. (2010). First report of bacterial canker of *Actinidia deliciosa* caused by *Pseudomonas syringae pv. actinidiae* in Portugal. *New Disease Reports*, *22*(10), 2044-2058.

Balint-Kurti, P. (2019). The plant hypersensitive response: concepts, control and consequences. *Molecular Plant Pathology*, *20*(8), 1163-1178.

Barrett, R. D., & Hoekstra, H. E. (2011). Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, *12*(11), 767.

Bartkiewicz, A. M., Chilla, F., Terefe-Ayana, D., Lübeck, J., Strahwald, J., Tacke, E., Hofferbert, H.-R., Linde, M., & Debener, T. (2018). Maximization of markers linked in coupling for tetraploid potatoes via monoparental haploids. *Frontiers in plant science*, *9*, 620.

Bastiaanssen, H. J. M., Van Den Berg, P. M. M. M., Lindhout, P., Jacobsen, E., Ramanna, M. S. . (1998). Postmeiotic restitution in 2n-egg formation of diploid potato. *Heredity*, *81*(1), 20-27.

Beardsley Jr, J. W., & Gonzalez, R. H. (1975). The biology and ecology of armored scales. *Annual review of entomology*, *20*(1), 47-73.

Beavis, W. D. (1998). QTL analyses: power, precision, and accuracy. *Molecular dissection of complex traits*, *1998*, 145-162.

Belkhadir, Y., Subramaniam, R., & Dangl, J. L. (2004). Plant disease resistance protein signaling: NBS--LRR proteins and their partners. *Current opinion in plant biology*, *7*(4), 391-399.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300.

Bent, A. F. (1996). Plant disease resistance genes: function meets structure. *The Plant Cell*, *8*(10), 1757.

Berry, J. A., Morales, C. F., Hill, M. G., Lofroth, B. J., Allan, D. J., & others. (1989, August 8-10). The incidence of three diaspid scales on kiwifruit in New Zealand. Proceedings of the Forty

Second New Zealand Weed and Pest Control Conference, Taranki Country Lodge, New Plymouth.
Bigeard, J., Colcombet, J., & Hirt, H. (2015). Signaling mechanisms in pattern-triggered immunity (PTI). *Molecular plant*, *8*(4), 521-539.
Birchler, J. A., & Veitia, R. A. (2014). The gene balance hypothesis: dosage effects in plants. *Plant Epigenetics and Epigenomics*, 25-32.
Blank, R. H., Olson, M. H., & Bell, D. S. (1987). Invasion of greedy scale crawlers (*Hemiberlesia rapax*) onto kiwifruit from taraire trees. *New Zealand Entomologist*, *10*(1), 127-130.
Blank, R. H., Olson, M. H., & Gill, G. S. C. (1992). Armoured scale, *Hemiberlesia lataniae* and *H. rapax* (Hemiptera: Diaspididae), infestation of kiwifruit rejected for export at two packhouses from 1987 to 1991. *New Zealand Journal of Crop and Horticultural Science*, *20*(4), 397-405.
Blank, R. H., Olson, M. H., & Lo, P. L. (1990). Armoured scale (Hemiptera: Diaspididae) aerial invasion into kiwifruit orchards from adjacent host plants. *New Zealand Journal of Crop and Horticultural Science*, *18*(2-3), 81-87.
Bortesi, L., & Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology advances*, *33*(1), 41-52.
Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*(7), 889-890.
Brown, J. K., & Rant, J. C. (2013). Fitness costs and trade-offs of disease resistance and their consequences for breeding arable crops. *Plant Pathology*, *62*(S1), 83-95.
Brownfield, L., & Köhler, C. (2011). Unreduced gamete formation in plants: mechanisms and prospects. *Journal of experimental botany*, *62*(5), 1659-1668.
Bruce Alberts, A. J., Julian Lewis, Martin Raff, Keith Roberts, Peter Walter. (2002). General Recombination. In *Molecular Biology of the Cell* (4th edition ed.). Garland Science.
Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., Goodstein, D. M., Elsik, C. G., Lewis, S. E., & Stein, L. (2016). JBrowse: a dynamic web platform for genome visualization and analysis. *Genome biology*, *17*(1), 1-12.
Campenhout, C. V., Cabochette, P., Veillard, A.-C., Laczik, M., Zelisko-Schmidt, A., Sabatel, C., Dhainaut, M., Vanhollebeke, B., Gueydan, C., & Kruys, V. (2019). Guidelines for optimized gene knockout using CRISPR/Cas9. *BioTechniques*, *66*(6), 295-302.
Cao, H., Bowling, S. A., Gordon, A. S., & Dong, X. (1994). Characterization of an Arabidopsis mutant that is nonresponsive to inducers of systemic acquired resistance. *The Plant Cell*, *6*(11), 1583-1592.
Cellini, A., Fiorentini, L., Buriani, G., Yu, J., Donati, I., Cornish, D., Novak, B., Costa, G., Vanneste, J., & Spinelli, F. (2014). Elicitors of the salicylic acid pathway reduce incidence of bacterial canker of kiwifruit caused by *Pseudomonas syringae* pv. *actinidae*. *Annals of applied biology*, *165*(3), 441-453.
Chen, K., & Gao, C. (2013). TALENs: customizable molecular DNA scissors for genome engineering of plants. *Journal of Genetics and Genomics*, *40*(6), 271-279.
Chen, N., Juric, I., Cosgrove, E. J., Bowman, R., Fitzpatrick, J. W., Schoech, S. J., Clark, A. G., & Coop, G. (2019). Allele frequency dynamics in a pedigreed natural population. *Proceedings of the National Academy of Sciences*, *116*(6), 2158-2164.

Chen, Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual review of plant biology*, *58*, 377.

Cheng, C.-H. (2014). Inheritance of resistance to *Pseudomonas syringae* pv. *actinidiae* and genetic correlations with fruit characters in a diploid *Actinidia chinensis* (kiwifruit) population. *Euphytica*, *198*(2), 305-315.

Cheng, C., Crowhurst, R., Hilario, E., Datson, P., Barron, L., Manako, K., Deng, C., De Silva, N., & Bomert, M. (2015). The application of genome selection to kiwifruit breeding. XIV EUCARPIA Symposium on Fruit Breeding and Genetics 1172.

Cheng, C., Seal, A., Murphy, S., & Lowe, R. (2006). Variability and inheritance of flowering time and duration in *Actinidia chinensis* (kiwifruit). *Euphytica*, *147*, 395-402.

Choi, H. W., & Klessig, D. F. (2016). DAMPs, MAMPs, and NAMPs in plant innate immunity. *BMC plant biology*, *16*(1), 232.

Churchill, G. A., & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, *138*(3), 963-971.

Clarke, D. (1986). Tolerance of parasites and disease in plants and its significance in host-parasite interactions.

Collard, B., Jahufer, M., Brouwer, J., & Pang, E. (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.

Colombi, E., Straub, C., Künzel, S., Templeton, M. D., McCann, H. C., & Rainey, P. B. (2017). Evolution of copper resistance in the kiwifruit pathogen *Pseudomonas syringae* pv. *actinidiae* through acquisition of integrative conjugative elements and plasmids. *Environmental microbiology*, *19*(2), 819-832.

Conolly, J., Colledge, S., & Shennan, S. (2008). Founder effect, drift, and adaptive change in domestic crop use in early Neolithic Europe. *Journal of Archaeological Science*, *35*(10), 2797-2804.

Cornish-Bowden, A. (1985). Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic acids research*, *13*(9), 3021.

Costa, G., Ferguson, R., Huang, H., & Testolin, R. (2017). Main changes in the kiwifruit industry since its introduction: present situation and future. IX International Symposium on Kiwifruit 1218.

Costarelli, A., Bianchet, C., Ederli, L., Salerno, G., Piersanti, S., Rebora, M., & Pasqualini, S. (2020). Salicylic acid induced by herbivore feeding antagonizes jasmonic acid mediated plant defenses against insect attack. *Plant Signaling & Behavior*, *15*(1), 1704517.

Crowhurst, R., Davy, M., & Deng, C. (2006). BioView-an enterprise bioinformatics system for automated analysis and annotation of non-genomic DNA sequence. *Napier, New Zealand*.

Crowhurst, R., Lints, R., Atkinson, R., & Gardner, R. (1990). Restriction fragment length polymorphisms in the genus *Actinidia* (Actinidiaceae). *Plant systematics and evolution*, *172*(1-4), 193-203.

Crowhurst, R. N., Gleave, A. P., MacRae, E. A., Ampomah-Dwamena, C., Atkinson, R. G., Beuning, L. L., Bulley, S. M., Chagne, D., Marsh, K. B., & Matich, A. J. (2008). Analysis of expressed sequence tags from *Actinidia*: applications of a cross species EST database for gene discovery in the areas of flavor, health, color and ripening. *BMC genomics*, *9*(1), 351.

Dai, P., Kong, J., Wang, S., Lu, X., Luo, K., Cao, B., Meng, X., & Luan, S. (2018). Identification of SNPs associated with residual feed intake from the muscle of *Litopenaeus vannamei* using bulk segregant RNA-seq. *Aquaculture*, *497*, 56-63.

Datson, P., & Ferguson, A. (2011). *Actinidia*. In *Wild crop relatives: genomic and breeding resources* (pp. 1-20). Springer.

Datson, P., Nardozza, S., Manako, K., Herrick, J., Martinez-Sanchez, M., Curtis, C., & Montefiori, M. (2013). Monitoring the *Actinidia* germplasm for resistance to *Pseudomonas syringae* pv. *actinidiae*. I International Symposium on Bacterial Canker of Kiwifruit 1095.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), 499.

De Silva, N. H., Gea, L., & Lowe, R. (2014). Genetic analysis of resistance to *Pseudomonas syringae* pv. *actinidiae* (Psa) in a kiwifruit progeny test: an application of generalised linear mixed models (GLMMs). *SpringerPlus*, *3*(1), 1-12.

Deepak, S., Shailasree, S., Kini, R. K., Muck, A., Mithöfer, A., & Shetty, S. H. (2010). Hydroxyproline-rich glycoproteins and plant defence. *Journal of Phytopathology*, *158*(9), 585-593.

Denwood, M., Mather, A., Haydon, D., Matthews, L., Mellor, D., & Reid, S. (2011). From phenotype to genotype: a Bayesian solution. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1710), 1434-1440.

Deokar, A., Sagi, M., Daba, K., & Tar'an, B. (2019). QTL sequencing strategy to map genomic regions associated with resistance to ascochyta blight in chickpea. *Plant biotechnology journal*, *17*(1), 275-288.

Ding, M., Jiang, J., Yang, F., Zheng, F., Fang, J., Wang, Q., Wang, J., Yao, W., Liu, X., & Gao, X. (2019). Holliday junction recognition protein interacts with and specifies the centromeric assembly of CENP-T. *Journal of Biological Chemistry*, *294*(3), 968-980.

Dixit, G., Srivastava, A., Rai, K. M., Dubey, R. S., Srivastava, R., & Verma, P. C. (2020). Distinct defensive activity of phenolics and phenylpropanoid pathway genes in different cotton varieties toward chewing pests. *Plant Signaling & Behavior*, *15*(5), 1747689.

Doerge, R. W., & Churchill, G. A. (1996). Permutation tests for multiple loci affecting a quantitative character. *Genetics*, *142*(1), 285-294.

Donati, I., Buriani, G., Cellini, A., Mauri, S., Costa, G., & Spinelli, F. (2014). New insights on the bacterial canker of kiwifruit (*Pseudomonas syringae* pv. *actinidiae*). *Journal of Berry Research*, *4*(2), 53-67.

Dong, W., Wu, D., Li, G., Wu, D., & Wang, Z. (2018). Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. *Scientific reports*, *8*(1), 1-7.

Dugdale, J. S. (1990). Reassessment of *Ctenopseustis Meyrick* and *Planotortrix Dugdale* with descriptions of two new genera (Lepidoptera: Tortricidae). *New Zealand journal of zoology*, *17*(3), 437-465.

Dutta, S. K., & Devaiah, M. C. (1992). Response of certain sugarcane varieties to scale insect and association of plant character with infestation. *Cooperative Sugar*, *23*(5), 337-342.

Dwiartama, A. (2017). Resilience and transformation of the New Zealand kiwifruit industry in the face of Psa-V disease. *Journal of Rural Studies*, *52*, 118-126.

Edwards, D., Forster, J. W., Chagné, D., & Batley, J. (2007). What Are SNPs? In *Association mapping in plants* (pp. 41-52). Springer.

Edwards, R., Carraher, C., Poulton, J., Sandanayaka, M., Todd, J. H., Dobson, S., Mauchline, N., Hill, G., McKenna, C., & Newcomb, R. D. (2008). DNA diagnostics of three armored scale species on kiwifruit in New Zealand. *Journal of economic entomology*, *101*(6), 1944-1949.

Egesi, C. N., Odu, B. O., Ogunyemi, S., Asiedu, R., & Hughes, J. (2007). Evaluation of water yam (*Dioscorea alata* L.) germplasm for reaction to yam anthracnose and virus diseases and their effect on yield. *Journal of Phytopathology*, *155*(9), 536-543.

El Bali, L., Diman, A., Bernard, A., Roosens, N. H., & De Keersmaecker, S. C. (2014). Comparative study of seven commercial kits for human DNA extraction from urine samples suitable for DNA biomarker-based public health studies. *Journal of biomolecular techniques: JBT*, *25*(4), 96.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, *6*(5), e19379.

Everett, K. R., Taylor, R. K., Romberg, M. K., Rees-George, J., Fullerton, R. A., Vanneste, J. L., & Manning, M. A. (2011). First report of *Pseudomonas syringae* pv. *actinidiae* causing kiwifruit bacterial canker in New Zealand. *Australasian Plant Disease Notes*, *6*(1), 67-71.

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, *103*(4), 285-298.

Fagundes, D., Bohn, B., Cabreira, C., Leipelt, F., Dias, N., Bodanese-Zanettini, M. H., & Cagliari, A. (2015). Caspases in plants: metacaspase gene family in plant stress responses. *Functional & integrative genomics*, *15*(6), 639-649.

Fairweather, J. R. (1999). Understanding how farmers choose between organic and conventional production: Results from New Zealand and policy implications. *Agriculture and human values*, *16*(1), 51-63.

Falcon, K. M. a. H. W. a. D. A. J. a. S. (2020). *RSQLite: 'SQLite' interface for R*. In https://CRAN.R-project.org/package=RSQLite

Falconer, D. S., Mackay, T. F., & Frankham, R. (1996). Introduction to quantitative genetics (4th edn). *Trends in Genetics*, *12*(7), 280.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.-L., Wei, P., Cao, F., Zhu, S., Zhang, F., & Mao, Y. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell research*, *23*(10), 1229-1232.

Ferguson, A. R., & Huang, H. W. (2007). Genetic resources of kiwifruit: Domestication and breeding. *Horticultural reviews*, *33*, 1-121.

Ferguson, A. R., & Seal, A. G. (2008). Kiwifruit. In *Temperate fruit crop breeding* (pp. 235-264). Springer.

Ferrante, P., & Scortichini, M. (2010). Molecular and phenotypic features of *Pseudomonas syringae* pv. *actinidiae* isolated during recent epidemics of bacterial canker on yellow kiwifruit (*Actinidia chinensis*) in central Italy. *Plant Pathology*, *59*(5), 954-962.

Filipe, J. A., & Kyriazakis, I. (2019). Bayesian, likelihood-free modelling of phenotypic plasticity and variability in individuals and populations. *Frontiers in Genetics*, 727.

Flint-Garcia, S. A., Thornsberry, J. M., & Buckler IV, E. S. (2003). Structure of linkage disequilibrium in plants.

Foss, E., Lande, R., Stahl, F., & Steinberg, C. (1993). Chiasma interference as a function of genetic distance. *Genetics*, *133*(3), 681-691.

Fourment, M., & Gillings, M. R. (2008). A comparison of common programming languages used in bioinformatics. *BMC bioinformatics*, *9*(1), 1-9.

Fowler, J. H., Narváez-Vásquez, J., Aromdee, D. N., Pautot, V., Holzer, F. M., & Walling, L. L. (2009). Leucine aminopeptidase regulates defense and wound signaling in tomato downstream of jasmonic acid. *The Plant Cell*, *21*(4), 1239-1251.

Francois, H. W. a. J. H. a. R. (2018). *Readr: read rectangular text data*. In https://CRAN.R-project.org/package=readr

Fraser, L., Harvey, C., Crowhurst, R., & De Silva, H. (2004). EST-derived microsatellites from *Actinidia* species and their potential for mapping. *Theoretical and Applied Genetics*, *108*(6), 1010-1016.

Fraser, L. G., Tsang, G. K., Datson, P. M., De Silva, H. N., Harvey, C. F., Gill, G. P., Crowhurst, R. N., & McNeilage, M. A. (2009). A gene-rich linkage map in the dioecious species *Actinidia chinensis* (kiwifruit) reveals putative X/Y sex-determining chromosomes. *BMC genomics*, *10*(1), 1-15.

Furbank, R. T., & Tester, M. (2011). Phenomics–technologies to relieve the phenotyping bottleneck. *Trends in plant science*, *16*(12), 635-644.

Gaj, T., Gersbach, C. A., & Barbas III, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology*, *31*(7), 397-405.

Gaut, B. S., & Long, A. D. (2003). The lowdown on linkage disequilibrium. *The Plant Cell*, *15*(7), 1502-1506.

Gea, L. (2011). Sustainable population management of *Actinidia* (Kiwifruit) for multiple products. In *Fruit, vegetable and cereal science and biotechnonogy* (pp. 68-71). Global science books.

Gentleman, R. (2008). *R programming for bioinformatics*. Chapman and Hall/CRC.

Gheysen, G., & Vanholme, B. (2007). RNAi from plants to nematodes. *Trends in biotechnology*, *25*(3), 89-92.

Gibbs, R. A. (2020). The human genome project changed everything. *Nature Reviews Genetics*, *21*(10), 575-576.

Gibbs, R. A., Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F., Yang, H., Ch'ang, L.-Y., Huang, W., Liu, B., & Shen, Y. (2003). The international HapMap project.

Glazier, A. M., Nadeau, J. H., & Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science*, *298*(5602), 2345-2349.

Grant, M. R., & Jones, J. D. G. (2009). Hormone (dis) harmony moulds plant health and disease. *Science*, *324*(5928), 750-752.

Grønvold, L. (2021). *RLinuxModules: Interface to the linux environment modules*. In https://CRAN.R-project.org/package=data.table

Gu, N., Zhang, X., Gu, X., Zhao, L., Godana, E. A., Xu, M., & Zhang, H. (2021). Transcriptomic and proteomic analysis of the mechanisms involved in enhanced disease resistance of strawberries induced by *Rhodotorula mucilaginosa* cultured with chitosan. *Postharvest Biology and Technology*, *172*, 111355.

Guan, L., Fan, P., Li, S.-H., Liang, Z., & Wu, B.-H. (2019). Inheritance patterns of anthocyanins in berry skin and flesh of the interspecific population derived from teinturier grape. *Euphytica*, *215*(4), 1-14.
Gupta, P. K., Kulwal, P. L., & Jaiswal, V. (2019). Association mapping in plants in the post-GWAS genomics era. *Advances in genetics*, *104*, 75-154.
Hackett, C. A. (2002). Statistical methods for QTL mapping in cereals. *Plant molecular biology*, *48*(5-6), 585-599.
Hale, H., Gardner, E. M., Viruel, J., Pokorny, L., & Johnson, M. G. (2020). Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. *Applications in plant sciences*, *8*(4), e11337.
Hale, I., Melo, A., & Gustafson, H. (2018). Sex-linked molecular markers for two cold-hardy kiwifruit species, *Actinidia arguta* and *A. kolomikta*. *European Journal of Horticultural Science*, *83*(4), 236-246.
Hancock, J. M. (1999). Microsatellites and other simple sequences: genomic context and mutational mechanisms. *Microsatellites: Evolution and applications*.
Hare, J. D., Yu, D. S., & Luck, R. F. (1990). Variation in life history parameters of California red scale on different citrus cultivars. *Ecology*, *71*(4), 1451-1460.
Hartl, D. L., Clark, A. G., & Clark, A. G. (1997). *Principles of population genetics* (Vol. 116). Sinauer associates Sunderland.
Harvey, C. F., Fraser, L. G., & Kent, J. (1991). *Actinidia* seed development in interspecific crosses. II International Symposium on Kiwifruit 297.
Henderson, R. (2011). *MAF plant health & environment laboratory*. PaDIL. Retrieved 20 Jan from http://www.padil.gov.au
Hester, J. (2020). *glue: Interpreted string literals*. In https://CRAN.R-project.org/package=glue
Hilario, E., Barron, L., Deng, C. H., Datson, P. M., De Silva, N., Davy, M. W., & Storey, R. D. (2015). Random Tagging Genotyping by Sequencing (rtGBS), an unbiased approach to locate restriction enzyme sites across the target genome. *PloS one*, *10*(12), e0143193.
Hill, M., Mauchline, N., Cheng, C., & Connolly, P. (2006). Measuring the resistance of Actinidia chinensis to armoured scale insects. VI International Symposium on Kiwifruit 753. https://www.actahort.org/members/showpdf?session=8238
Hill, M. G. (2013). Does Psa affect kiwifruit susceptibility to leafrollers? *New Zealand Plant Protection*, *66*, 162-169.
Hill, M. G., & Holmes, T. (2009). An analysis of latania scale (*Hemiberlesia lataniae*) crawler settlement behaviour on kiwifruit leaves and bark. *New Zealand Plant Protection*, *62*, 55-62.
Hill, M. G., Mauchline, N. A., Cate, L. R., Connolly, P. G., & others. (2005). A technique for measuring growth rate and survival of armoured scale insects. *New Zealand Plant Protection*, *58*, 288-288.
Hill, M. G., Mauchline, N. A., Cheng, C. H., & Connolly, P. G. (2006). Measuring the resistance of *Actinidia chinensis* to armoured scale insects. VI International Symposium on Kiwifruit 753.
Hill, M. G., Mauchline, N. A., Hall, A. J., & Stannard, K. A. (2009). Life table parameters of two armoured scale insect (Hemiptera: Diaspididae) species on resistant and susceptible

kiwifruit (*Actinidia* spp.) germplasm. *New Zealand Journal of Crop and Horticultural Science*, *37*(4), 335-343.

Hill, M. G., Mauchline, N. A., Jones, M. K., & Sutherland, P. W. (2011). The response of resistant kiwifruit (*Actinidia chinensis*) to armoured scale insect (Diaspididae) feeding. *Arthropod-Plant Interactions*, *5*(2), 149-161.

Hill, M. G., Wurms, K. V., Davy, M. W., Gould, E., Allan, A., Mauchline, N. A., Luo, Z., Chee, A. A., Stannard, K., & Storey, R. D. (2015). Transcriptome analysis of kiwifruit (*Actinidia chinensis*) bark in response to armoured scale insect (*Hemiberlesia lataniae*) feeding. *PloS one*, *10*(11), e0141664.

Hill, M. G., Wurms, K. V., Davy, M. W., Gould, E., Allan, A., Mauchline, N. A., Luo, Z., Chee, A. A., Stannard, K., & Storey, R. D. (2016). Microarray analysis of kiwifruit (*Actinidia chinensis*) bark following challenge by the sucking insect *Hemiberlesia lataniae* (Hemiptera: Diaspididae). *Genomics data*, *7*, 281-283.

Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D. G., Frazer, K. A., & Cox, D. R. (2005). Whole-genome patterns of common DNA variation in three human populations. *Science*, *307*(5712), 1072-1079.

Hirsch, A., Testolin, R., Brown, S., Chat, J., Fortune, D., Bureau, J., & De Nay, D. (2001). Embryo rescue from interspecific crosses in the genus *Actinidia* (kiwifruit). *Plant Cell Reports*, *20*(6), 508-516.

Homer, I. J. (1991). Epidemiology of Armillaria root-rot of kiwifruit. II International Symposium on Kiwifruit 297.

Huang, H.-W., & Ferguson, A. (2006). Actinidia in China: Natural diversity, phylogeographical evolution, interspecific gene flow and kiwifruit cultivar improvement. VI International Symposium on Kiwifruit 753.

Huang, H., Li, Z., Li, J., Kubisiak, T. L., & Layne, D. R. (2002). Phylogenetic relationships in *Actinidia* as revealed by RAPD analysis. *Journal of the American Society for Horticultural Science*, *127*(5), 759-766.

Huang, H. W., Gong, J. J., Wang, S. M., He, Z.-C., Zhang, Z.-H., Li, J.-Q., & Jiang, Z. W. (2000). Genetic diversity in the genus *Actinidia*. *Chinese Biodiversity*, *8*(1), 1-12.

Huang, S., Ding, J., Deng, D., Tang, W., Sun, H., Liu, D., Zhang, L., Niu, X., Zhang, X., & Meng, M. (2013). Draft genome of the kiwifruit *Actinidia chinensis*. *Nature communications*, *4*, 2640.

Huang, W.-G., Cipriani, G., Morgante, M., & Testolin, R. (1998). Microsatellite DNA in *Actinidia chinensis*: isolation, characterisation, and homology in related species. *Theoretical and Applied Genetics*, *97*(8), 1269-1278.

Jacobs, A. K., Lipka, V., Burton, R. A., Panstruga, R., Strizhov, N., Schulze-Lefert, P., & Fincher, G. B. (2003). An Arabidopsis callose synthase, GSL5, is required for wound and papillary callose formation. *The Plant Cell*, *15*(11), 2503-2513.

James, J. (1970). The founder effect and response to artificial selection. *Genetics Research*, *16*(3), 241-250.

Jansen, R. C. (1996). Complex plant traits: time for polygenic analysis. *Trends in plant science*, *1*(3), 89-94.

Jighly, A., Lin, Z., Pembleton, L. W., Cogan, N. O., Spangenberg, G. C., Hayes, B. J., & Daetwyler, H. D. (2019). Boosting genetic gain in allogamous crops via speed breeding and genomic selection. *Frontiers in plant science*, *10*, 1364.

Johnsson, M. (2018). Integrating selection mapping with genetic mapping and functional genomics. *Frontiers in Genetics*, *9*, 603.

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323-329.

Kamphuis, L. G., Klingler, J. P., Jacques, S., Gao, L.-L., Edwards, O. R., & Singh, K. B. (2019). Additive and epistatic interactions between AKR and AIN loci conferring bluegreen aphid resistance and hypersensitivity in *Medicago truncatula*. *Journal of experimental botany*, *70*(18), 4887-4902.

Kang, H. (2013). The prevention and handling of the missing data. *Korean journal of anesthesiology*, *64*(5), 402.

Karger, B. (2011). *DNA sequencing using capillary electrophoresis*.

Kefelegn, N., Haile, G., & Palanivel, H. (2021). Effect of plant tissues on DNA quantity and quality of barley (*Hordeum vulgare*) validating through PCR technique. *bioRxiv*(2021-10). https://doi.org/https://doi.org/10.1101/2021.10.28.466350

Keller, B., Bieri, S., Bossolini, E., & Yahiaoui, N. (2007). Cloning genes and QTLs for disease resistance in cereals. In *Genomics-assisted crop improvement* (pp. 103-127). Springer.

Kellerhals, M. (2009). Introduction to apple (Malus× domestica). In *Genetics and Genomics of Rosaceae* (pp. 73-84). Springer.

Kemppainen, P., Rønning, B., Kvalnes, T., Hagen, I. J., Ringsby, T. H., Billing, A. M., Pärn, H., Lien, S., Husby, A., & Sæther, B. E. (2017). Controlling for P-value inflation in allele frequency change in experimental evolution and artificial selection experiments. *Molecular ecology resources*, *17*(4), 770-782.

Klingler, J. P., Nair, R. M., Edwards, O. R., & Singh, K. B. (2009). A single gene, *AIN*, in *Medicago truncatula* mediates a hypersensitive response to both bluegreen aphid and pea aphid, but confers resistance only to bluegreen aphid. *Journal of experimental botany*, *60*(14), 4115-4127.

Koh, Y., Chung, H., Cha, B., & Lee, D. (1994). Outbreak and spread of bacterial canker in kiwifruit. *Korean Journal of Plant Pathology (Korea Republic)*, *10*, 68-72.

Konishi, T., Matsukuma, S., Fuji, H., Nakamura, D., Satou, N., & Okano, K. (2019). Principal component analysis applied directly to sequence matrix. *Scientific reports*, *9*(1), 1-13.

Konno, K., Inoue, T. A., & Nakamura, M. (2014). Synergistic defensive function of raphides and protease through the needle effect. *PloS one*, *9*(3), 91341-91341.

Korkovelos, A., Mavromatis, A., Huang, W., Hagidimitriou, M., Giakoundis, A., & Goulas, C. (2008). Effectiveness of SSR molecular markers in evaluating the phylogenetic relationships among eight *Actinidia* species. *Scientia horticulturae*, *116*(3), 305-310.

Koshland, D. E. (1989). Sequences and consequences of the human genome. *Science*, *246*(4927), 189-189.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, *47*(260), 583-621.

KVH. (2016). *Kiwifruit vine health*. Retrieved 05 March from http://www.kvh.org.nz/vdb/document/99225

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS computational biology*, *9*(8), e1003118.

Lefebvre, V., & Palloix, A. (1996). Both epistatic and additive effects of QTLs are involved in polygenic induced resistance to disease: a case study, the interaction pepper—*Phytophthora capsic*i Leonian. *Theoretical and Applied Genetics*, *93*(4), 503-511.

Li, D., Zhong, C., Liu, Y., & Huang, H. (2010). Correlation between ploidy level and fruit characters of the main kiwifruit cultivars in China: implication for selection and improvement. *New Zealand Journal of Crop and Horticultural Science*, *38*(2), 137-145.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987-2993.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Li, H., Hearne, S., Bänziger, M., Li, Z., & Wang, J. (2010). Statistical properties of QTL linkage mapping in biparental genetic populations. *Heredity*, *105*(3), 257.

Li, Q., Von Lanken, C., Yang, J., Lawrence, C. B., & Hunt, A. G. (2000). The yeast polyadenylate-binding protein (PAB1) gene acts as a disease lesion mimic gene when expressed in plants. *Plant molecular biology*, *42*(2), 335-344.

Li, Y. C., Korol, A. B., Fahima, T., Beiles, A., & Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, *11*(12), 2453-2465.

Li, Z., & Xu, Y. (2022). Bulk segregation analysis in the NGS era: a review of its teenage years. *The Plant Journal*, *109*(6), 1355-1374.

Little, T. J., Shuker, D. M., Colegrave, N., Day, T., & Graham, A. L. (2010). The coevolution of virulence: tolerance in perspective. *PLoS pathogens*, *6*(9), e1001006.

Lo, P., Walker, J., & Suckling, D. (2000). Insecticide resistance management of leafrollers (Lepidoptera: Tortricidae) in New Zealand. *New Zealand Plant Protection*, *53*, 163-167.

Lo, P. L., & Blank, R. H. (1989). A survey of armoured scale species (Hemiptera: Diaspididae) in kiwifruit orchards. *New Zealand Entomologist*, *12*(1), 1-4.

Lowe, R. G., Marsh, H. D., & McNeilage, M. A. (1999). *Kiwi plant named 'Hort16A'*. https://www.google.com/patents/USPP11066

Magwene, P. M., Willis, J. H., & Kelly, J. K. (2011). The statistics of bulk segregant analysis using next generation sequencing. *PLoS computational biology*, *7*(11), e1002255.

Mansfeld, B. N., & Grumet, R. (2018). QTLseqr: an R package for bulk segregant analysis with next-generation sequencing. *The plant genome*, *11*(2), 180006.

Marone, D., Russo, M. A., Laidò, G., De Vita, P., Papa, R., Blanco, A., Gadaleta, A., Rubiales, D., & Mastrangelo, A. M. (2013). Genetic basis of qualitative and quantitative resistance to powdery mildew in wheat: from consensus regions to candidate genes. *BMC genomics*, *14*(1), 562.

Marsal, G., Boronat, N., Canals, J. M., Zamora, F., & Fort, F. (2013). Comparison of the efficiency of some of the most usual DNA extraction methods for woody plants in different tissues of *Vitis vinifera* L. *OENO One*, *47*(4), 227-237.

Martin, R., & Luxton, P. (2004). The Successful Commercialisation of ZESPRI™ Gold Kiwifruit. International Symposium on Harnessing the Potential of Horticulture in the Asian-Pacific Region 694.
Martinez Henao, J., Demers, L. E., Grosser, K., Schedl, A., Van Dam, N. M., & Bede, J. C. (2020). Fertilizer rate-associated increase in foliar jasmonate burst observed in wounded *Arabidopsis thaliana* leaves is attenuated at eCO_2. *Frontiers in plant science*, 1636.
Matsumoto, Y., Goto, T., Nishino, J., Nakaoka, H., Tanave, A., Takano-Shimizu, T., Mott, R. F., & Koide, T. (2017). Selective breeding and selection mapping using a novel wild-derived heterogeneous stock of mice revealed two closely-linked loci for tameness. *Scientific reports*, *7*(1), 4607.
Matt Dowle, A. S. (2020). *Data.table: Extension of `data.frame`*. In https://CRAN.R-project.org/package=data.table
McCallum, J., Laing, W., Bulley, S., Thomson, S., Catanach, A., Shaw, M., Knaebel, M., Tahir, J., Deroles, S., & Timmerman-Vaughan, G. (2019). Molecular characterisation of a supergene conditioning super-high vitamin c in kiwifruit hybrids. *Plants*, *8*(7), 237.
McCann, H. C., Rikkerink, E. H., Bertels, F., Fiers, M., Lu, A., Rees-George, J., Andersen, M. T., Gleave, A. P., Haubold, B., & Wohlers, M. W. (2013). Genomic analysis of the kiwifruit pathogen *Pseudomonas syringae* pv. *actinidiae* provides insight into the origins of an emergent plant disease. *PLoS Pathog*, *9*(7), e1003503.
McConn, M., Creelman, R. A., Bell, E., Mullet, J. E., & Browse, J. (1997). Jasmonate is essential for insect defense in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, *94*(10), 5473-5477.
McDonald, B. A., & Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annual review of phytopathology*, *40*(1), 349-379.
McHale, L., Tan, X., Koehl, P., & Michelmore, R. W. (2006). Plant *NBS-LRR* proteins: adaptable guards. *Genome biology*, *7*(4), 1-11.
McKenna, C., Dobson, S., & Phare, J. (2009). The insect pest complex of *Actinidia arguta* kiwifruit. *New Zealand Plant Protection*, *62*, 262-267.
McKenna, C., & Stevens, P. (2007). A comparison of lepidopteran damage to Hort16A and Hayward kiwifruit. *New Zealand Plant Protection*, *60*, 254-258.
McNeilage, M. (2014). Genetics and breeding of kiwifruit. VIII International Symposium on Kiwifruit 1096.
McNeilage, M., & Considine, J. (1989). Chromosome studies in some *Actinidia* taxa and implications for breeding. *New Zealand journal of botany*, *27*(1), 71-81.
Mercer, T. R., Gerhardt, D. J., Dinger, M. E., Crawford, J., Trapnell, C., Jeddeloh, J. A., Mattick, J. S., & Rinn, J. L. (2012). Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nature biotechnology*, *30*(1), 99-104.
Mertten, D., Tsang, G., Manako, K., McNeilage, M., & Datson, P. (2012). Meiotic chromosome pairing in *Actinidia chinensis* var. *deliciosa*. *Genetica*, *140*(10), 455-462.
Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature Reviews Genetics*, *11*(1), 31.
Michelmore, R. W., Paran, I., & Kesseli, R. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in

specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences*, *88*(21), 9828-9832.
Miles, C., & Wayne, M. (2008). Quantitative trait locus (QTL) analysis. *Nature Education*, *1*(1), 208.
Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome research*, *17*(2), 240-248.
Miller, R. N. G., Costa Alves, G. S., & Van Sluys, M.-A. (2017). Plant immunity: unravelling the complexity of plant responses to biotic stresses. *Annals of botany*, *119*(5), 681-687.
Minervini, M., Scharr, H., & Tsaftaris, S. A. (2015). Image analysis: the new bottleneck in plant phenotyping. *IEEE signal processing magazine*, *32*(4), 126-131.
Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., & Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Science*, *282*, 60-72.
Mithöfer, A., & Boland, W. (2008). Recognition of herbivory-associated molecular patterns. *Plant Physiology*, *146*(3), 825-831.
Mohan, M., Nair, S., Bhagwat, A., Krishna, T., Yano, M., Bhatia, C., & Sasaki, T. (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular breeding*, *3*(2), 87-103.
Mondragón-Palomino, M., Meyers, B. C., Michelmore, R. W., & Gaut, B. S. (2002). Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome research*, *12*(9), 1305-1315.
Morales, C. (1988). The occurrence of latania scale, *Hemiberlesia lataniae* (Signoret)(Hemiptera: Diaspididae), in New Zealand. *New Zealand journal of experimental agriculture*, *16*(1), 77-82.
Morgan, D., Warrington, I., & Halligan, E. (1985). Effect of temperature and photosynthetic photon flux density on vegetative growth of kiwifruit (*Actinidia chinensis*). *New Zealand Journal of Agricultural Research*, *28*(1), 109-116.
Morozova, O., & Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, *92*(5), 255-264.
Mott, G. A., Guttman, D. S., & Desveaux, D. (2017). The study of pattern-triggered immunity in Arabidopsis. *Canadian Journal of Plant Pathology*, *39*(3), 275-281.
MPI. (2022). *Situation and outlook for primary industries (SOPI) reports.* M. f. P. Industries. https://www.mpi.govt.nz
Müller, K. (2017). *Here: A aimpler way to find your files*. In https://CRAN.R-project.org/package=here
Mundt, C. C. (2014). Durable resistance: a key to sustainable management of pathogens and pests. *Infection, Genetics and Evolution*, *27*, 446-455.
Munjal, G., Hao, J., Teuber, L. R., & Brummer, E. C. (2018). Selection mapping identifies loci underpinning autumn dormancy in alfalfa (*Medicago sativa*). *G3: Genes, Genomes, Genetics*, *8*(2), 461-468.
Mur, L. A. J., Kenton, P., Atzorn, R., Miersch, O., & Wasternack, C. (2006). The outcomes of concentration-specific interactions between salicylate and jasmonate signaling include synergy, antagonism, and oxidative stress leading to cell death. *Plant Physiology*, *140*(1), 249-262.

Narum, S. R., & Hess, J. E. (2011). Comparison of FST outlier tests for SNP loci under selection. *Molecular ecology resources*, *11*, 184-194.
Neethling, C., & De Lange, H. (1995). Susceptibility of three passion fruit types to the white peach scale. *Fruits*, *50*(4), 305-309.
Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome research*, *15*(11), 1566-1575.
Novo, M., Romo, S., Rey, M., Prado, M. J., & González, M. V. (2010). Identification and sequence characterisation of molecular markers polymorphic between male kiwifruit (*Actinidia chinensis* var. *deliciosa* (A. Chev.) accessions exhibiting different flowering time. *Euphytica*, *175*(1), 109-121.
Nyholt, D. R. (2000). All LODs are not created equal. *American journal of human genetics*, *67*(2), 282.
Onorato, R., Yu, J., Cornish, D., Spinelli, F., Max, S., Vanneste, J., & Kay, C. (2010). Recent advances in the characterisation and control of *Pseudomonas syringae* pv. *actinidiae*, the causal agent of bacterial canker on kiwifruit. VII International Symposium on Kiwifruit 913.
Oren, M., Hudgell, M. A. B., Golconda, P., Lun, C. M., & Smith, L. C. (2016). Genomic instability and shared mechanisms for gene diversification in two distant immune gene families: The plant *NBS-LRR* genes and the *echinoid 185/333* genes. In *The Evolution of the Immune System* (pp. 295-310). Elsevier.
Pagán, I., & García-Arenal, F. (2018). Tolerance to plant pathogens: theory and experimental evidence. *International journal of molecular sciences*, *19*(3), 810.
Pandey, S. P., & Somssich, I. E. (2009). The role of WRKY transcription factors in plant immunity. *Plant Physiology*, *150*(4), 1648-1655.
Peñalba, J. V., & Wolf, J. B. (2020). From molecules to populations: appreciating and estimating recombination rate variation. *Nature Reviews Genetics*, *21*(8), 476-492.
Pennycook, S. (1985). Fungal fruit rots of *Actinidia deliciosa* (kiwifruit). *New Zealand journal of experimental agriculture*, *13*(4), 289-299.
Pieterse, C. M., Van der Does, D., Zamioudis, C., Leon-Reyes, A., & Van Wees, S. C. (2012). Hormonal modulation of plant immunity. *Annual review of cell and developmental biology*, *28*, 489-521.
Piisilä, M., Keceli, M. A., Brader, G., Jakobson, L., Jõesaar, I., Sipari, N., Kollist, H., Palva, E. T., & Kariola, T. (2015). The F-box protein MAX2 contributes to resistance to bacterial phytopathogens in *Arabidopsis thaliana*. *BMC plant biology*, *15*(1), 1-17.
Pilkington, S. M., Crowhurst, R., Hilario, E., Nardozza, S., Fraser, L., Peng, Y., Gunaseelan, K., Simpson, R., Tahir, J., & Deroles, S. C. (2018). A manually annotated *Actinidia chinensis* var. *chinensis* (kiwifruit) genome highlights the challenges associated with draft genomes and gene prediction in plants. *BMC genomics*, *19*(1), 257.
Pointeau, S., Ameline, A., Laurans, F., Sallé, A., Rahbé, Y., Bankhead-Dronnet, S., & Lieutier, F. (2012). Exceptional plant penetration and feeding upon cortical parenchyma cells by the woolly poplar aphid. *Journal of insect physiology*, *58*(6), 857-866.
Poland, J. A., Balint-Kurti, P. J., Wisser, R. J., Pratt, R. C., & Nelson, R. J. (2009). Shades of gray: the world of quantitative disease resistance. *Trends in plant science*, *14*(1), 21-29.

Powers, M., Mannthey, K., Sebastian, P., Adsule, S., Kiernan, E., Smith, J. T., Way, J., Shifaw, B., Roazen, D., & Narvaez, P. (2022). Deploying genomics workflows on high performance computing (HPC) platforms: storage, memory, and compute considerations. *bioRxiv*. https://doi.org/https://doi.org/10.1101/2022.04.05.485833

Preston, J., VanZeeland, A., & Peiffer, D. (2021). Innovation at Illumina: The road to the $600 human genome. *Nat. Portf*.

QIAGEN. (2016). *DNeasy® plant mini kit*. Retrieved June 2022 from https://www.qiagen.com/

Quenouille, J., Montarry, J., Palloix, A., & Moury, B. (2013). Farther, slower, stronger: how the plant genetic background protects a major resistance gene from breakdown. *Molecular Plant Pathology*, *14*(2), 109-118.

Råberg, L. (2014). How to live with the enemy: understanding tolerance to parasites. *PLoS Biology*, *12*(11), e1001989.

Rassam, M., & Laing, W. (2005). Variation in ascorbic acid and oxalate levels in the fruit of *Actinidia chinensis* tissues and genotypes. *Journal of agricultural and food chemistry*, *53*(6), 2322-2326.

Reglinski, T., Vanneste, J., Wurms, K., Gould, E., Spinelli, F., & Rikkerink, E. (2013). Using fundamental knowledge of induced resistance to develop control strategies for bacterial canker of kiwifruit caused by *Pseudomonas syringae* pv. *actinidiae*. *Frontiers in plant science*, *4*, 24.

Reglinski, T., Whitaker, G., Cooney, J. M., Taylor, J. T., Poole, P. R., Roberts, P. B., & Kim, K. K. (2001). Systemic acquired resistance to *Sclerotinia sclerotiorum* in kiwifruit vines. *Physiological and molecular plant pathology*, *58*(3), 111-118.

Reymond, P. (2021). Receptor kinases in plant responses to herbivory. *Current Opinion in Biotechnology*, *70*, 143-150.

Rivas, S., Rougon-Cardoso, A., Smoker, M., Schauser, L., Yoshioka, H., & Jones, J. D. (2004). Retraction: *CITRX thioredoxin* interacts with the tomato *Cf-9* resistance protein and negatively regulates defence. *The EMBO journal*, *23*(10), 2156-2165.

Robertson, A. (1970). A theory of limits in artificial selection with many linked loci. In *Mathematical topics in population genetics* (pp. 246-288). Springer.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, *29*(1), 24-26.

Rossi, M., Goggin, F. L., Milligan, S. B., Kaloshian, I., Ullman, D. E., & Williamson, V. M. (1998). The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. *Proceedings of the National Academy of Sciences*, *95*(17), 9750-9754.

Saint Pierre, C., Crossa, J. L., Bonnett, D., Yamaguchi-Shinozaki, K., & Reynolds, M. P. (2012). Phenotyping transgenic wheat for drought resistance. *Journal of experimental botany*, *63*(5), 1799-1808.

Samtools. (2020). *BCFtools*. Retrieved October from http://samtools.github.io/bcftools/

Sánchez-Pérez, R., Ballester, J., Dicenta, F., Arús, P., & Martínez-Gómez, P. (2006). Comparison of SSR polymorphisms using automated capillary sequencers, and polyacrylamide and agarose gel electrophoresis: implications for the assessment of genetic diversity and relatedness in almond. *Scientia horticulturae*, *108*(3), 310-316.

Sawicki, M. P., Samara, G., Hurwitz, M., & Passaro, E. (1993). Human Genome Project. *The American Journal of Surgery*, *165*(2), 258-264.

Schoen, D. J., & Schultz, S. T. (2019). Somatic mutation and evolution in plants. *Annual Review of Ecology, Evolution, and Systematics*, *50*, 49-73.
Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature biotechnology*, *18*(2), 233.
Scortichini, M. (1994). Occurrence of *Pseudomonas syringae* pv. *actinidiae* on kiwifruit in Italy. *Plant Pathology*, *43*(6), 1035-1038.
Seal, A., Dunn, J., & Jia, Y. (2013). Pollen parent effects on fruit attributes of diploid *Actinidia chinensis* 'Hort16A'kiwifruit. *New Zealand Journal of Crop and Horticultural Science*, *41*(4), 219-229.
Selig, P., Keough, S., Nalam, V. J., & Nachappa, P. (2016). Jasmonate-dependent plant defenses mediate soybean thrips and soybean aphid performance on soybean. *Arthropod-Plant Interactions*, *10*(4), 273-282.
Shen, R., & Messer, P. W. (2022). Predicting the genomic resolution of bulk segregant analysis. *G3*, *12*(3), jkac012.
Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature biotechnology*, *26*(10), 1135.
Shi, L., & Wang, Z. (2019). Computational strategies for scalable genomics analysis. *Genes*, *10*(12), 1017.
Small, I. (2007). RNAi for revealing and engineering plant gene functions. *Current Opinion in Biotechnology*, *18*(2), 148-153.
Smirle, M. J., Thomas Lowery, D., & Zurowski, C. L. (2002). Resistance and cross-resistance to four insecticides in populations of obliquebanded leafroller (Lepidoptera: Tortricidae). *Journal of economic entomology*, *95*(4), 820-825.
Sōgawa, K. (1982). The rice brown planthopper: feeding physiology and host plant interactions. *Annual review of entomology*, *27*(1), 49-73.
Song, J., Li, Z., Liu, Z., Guo, Y., & Qiu, L.-J. (2017). Next-generation sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in soybean. *Frontiers in plant science*, *8*, 919.
Song, P., Wang, J., Guo, X., Yang, W., & Zhao, C. (2021). High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. *The Crop Journal*, *9*(3), 633-645.
Soto-Cerda, B. J., & Cloutier, S. (2012). Association mapping in plant genomes. In *Genetic diversity in plants*. InTech.
Spoel, S. H., Johnson, J. S., & Dong, X. (2007). Regulation of tradeoffs between plant defenses against pathogens with different lifestyles. *Proceedings of the National Academy of Sciences*, *104*(47), 18842-18847.
Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: Join Map. *The Plant Journal*, *3*(5), 739-744.
Steven, D. (1991). Pest lepidoptera of kiwifruit in New Zealand. II International Symposium on Kiwifruit 297.
Steven, D. (1999). Integrated and organic production of kiwifruit. IV International Symposium on Kiwifruit 498.
Steven, D., Barnett, S., Stevens, P., & McKenna, C. (1995). Changing pest control on New Zealand kiwifruit. III International Symposium on Kiwifruit 444.

Steven, D., Tomkins, A. R., Blank, R. H., & Charles, J. G. (1994). A first-stage integrated pest management system for kiwifruit.

Strauss, S. Y., & Agrawal, A. A. (1999). The ecology and evolution of plant tolerance to herbivory. *Trends in ecology & evolution*, *14*(5), 179-185.

Sun, J., Yang, L., Wang, J., Liu, H., Zheng, H., Xie, D., Zhang, M., Feng, M., Jia, Y., & Zhao, H. (2018). Identification of a cold-tolerant locus in rice (*Oryza sativa* L.) using bulked segregant analysis with a next-generation sequencing strategy. *Rice*, *11*(1), 1-12.

Syvänen, A.-C. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, *2*(12), 930.

Tahir, J., Brendolise, C., Hoyte, S., Lucas, M., Thomson, S., Hoeata, K., McKenzie, C., Wotton, A., Funnell, K., & Morgan, E. (2020). Qtl mapping for resistance to cankers induced by *Pseudomonas syringae* pv. *actinidiae* (psa) in a tetraploid *Actinidia chinensis* kiwifruit population. *Pathogens*, *9*(11), 967.

Tahir, J., Crowhurst, R., Deroles, S., Hilario, E., Schaffer, R., Le Lievre, L., Brendolise, C., Chagne, D., Gardiner, S. E., & Knaebel, M. (2022). First chromosome-scale assembly and deep floral-bud transcriptome of a male kiwifruit. *Frontiers in Genetics*, 961.

Tahir, J., Gardiner, S., Bassett, H., Chagné, D., Deng, C., & Gea, L. (2017). Tolerance to *Pseudomonas syringae* pv. *actinidiae* in a kiwifruit breeding parent is conferred by multiple loci. IV International Symposium on Molecular Markers in Horticulture 1203.

Tahir, J., Hoyte, S., Bassett, H., Brendolise, C., Chatterjee, A., Templeton, K., Deng, C., Crowhurst, R., Montefiori, M., & Morgan, E. (2019). Multiple quantitative trait loci contribute to resistance to bacterial canker incited by *Pseudomonas syringae* pv. *actinidiae* in kiwifruit (*Actinidia chinensis*). *Horticulture research*, *6*(1), 1-18.

Takikawa, Y., Serizawa, S., Ichikawa, T., Tsuyumu, S., & Goto, M. (1989). *Pseudomonas syringae* pv. *actinidiae* pv. nov. the causal bacterium of canker of kiwifruit in Japan. *Japanese Journal of Phytopathology*, *55*(4), 437-444.

Testolin, R., Huang, H.-W., & Ferguson, A. R. (2016). *The kiwifruit genome*. Springer.

Thomma, B. P. H. J., Eggermont, K., Penninckx, I. A. M. A., Mauch-Mani, B., Vogelsang, R., Cammue, B. P. A., & Broekaert, W. F. (1998). Separate jasmonate-dependent and salicylate-dependent defense-response pathways in Arabidopsis are essential for resistance to distinct microbial pathogens. *Proceedings of the National Academy of Sciences*, *95*(25), 15107-15111.

Tian, D., Wang, P., Tang, B., Teng, X., Li, C., Liu, X., Zou, D., Song, S., & Zhang, Z. (2020). GWAS Atlas: a curated resource of genome-wide variant-trait associations in plants and animals. *Nucleic acids research*, *48*(D1), D927-D932.

Tsang, E., & Maciocci, G. (2020). Welcome to a new ERA of reproducible publishing. *eLIFE Labs*.

Tsang, G. K., Fraser, L. G., McNeilage, M. A., De Silva, H. N., & MacRae, E. A. (2006). Genetic markers: their transferability and usefulness across *Actinidia*. VI International Symposium on Kiwifruit 753.

Udagawa, H., Koga, K., Shinjo, A., Kitashiba, H., & Takakura, Y. (2020). Reduced susceptibility to a tobacco bushy top virus Malawi isolate by loss of function in host eIF (iso) 4E genes. *Breeding science*, 19135.

Van Nimwegen, K. J. M., Van Soest, R. A., Veltman, J. A., Nelen, M. R., Van der Wilt, G. J., Vissers, L. E. L. M., & Grutters, J. P. C. (2016). Is the $1000 genome as near as we think? A cost analysis of next-generation sequencing. *Clinical Chemistry*, *62*(11), 1458-1464.

Van Ooijen, J. (2004). *MapQTL 5, Software for the mapping of quantitative trait loci in experimental populations.* . In

Vanneste, J. L., Yu, J., Cornish, D. A., Tanner, D. J., Windner, R., Chapman, J. R., Taylor, R. K., Mackay, J. F., & Dowlut, S. (2013). Identification, virulence, and distribution of two biovars of *Pseudomonas syringae* pv. *actinidiae* in New Zealand. *Plant Disease*, *97*(6), 708-719.

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. d. F. (2016). Microsatellite markers: what they mean and why they are so useful. *Genetics and molecular biology*, *39*(3), 312-328.

Vo, K. T. X., Kim, C.-Y., Chandran, A. K. N., Jung, K.-H., An, G., & Jeon, J.-S. (2015). Molecular insights into the function of ankyrin proteins in plants. *Journal of plant biology*, *58*(5), 271-284.

Voss-Fels, K. P., Cooper, M., & Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. *Theoretical and Applied Genetics*, *132*(3), 669-686.

Wang, T., & Gleave, A. P. (2012). Applications of biotechnology in kiwifruit (*Actinidia*). In *Innovations in Biotechnology*. InTech.

Wang, T., Wang, G., Jia, Z.-H., Pan, D.-L., Zhang, J.-Y., & Guo, Z.-R. (2018). Transcriptome analysis of kiwifruit in response to *Pseudomonas syringae* pv. *actinidiae* infection. *International journal of molecular sciences*, *19*(2), 373.

Wang, Z., Yu, A., Li, F., Xu, W., Han, B., Cheng, X., & Liu, A. (2021). Bulked segregant analysis reveals candidate genes responsible for dwarf formation in woody oilseed crop castor bean. *Scientific reports*, *11*(1), 1-15.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 359-372.

Wearing, C., Colhoun, K., Attfield, B., Marshall, R., & McLaren, G. (2003). Screening for resistance in apple cultivars to lightbrown apple moth, *Epiphyas postvittana*, and greenheaded leafroller, *Planotortrix octo*, and its relationship to field damage. *Entomologia experimentalis et applicata*, *109*(1), 39-53.

Wickham, H. (2019a). *magrittr: A forward-pipe operator for R*. In https://CRAN.R-project.org/package=stringr

Wickham, H. (2019b). *stringr: Simple, consistent wrappers for common string operations*. In

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., & Hester, J. (2019). Welcome to the Tidyverse. *Journal of open source software*, *4*(43), 1686.

Wijnker, E., & de Jong, H. (2008). Managing meiotic recombination in plant breeding. *Trends in plant science*, *13*(12), 640-646.

Wildermuth, M. C., Dewdney, J., Wu, G., & Ausubel, F. M. (2001). Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature*, *414*(6863), 562-565.

Willocquet, L., Savary, S., & Yuen, J. (2017). Multiscale phenotyping and decision strategies in breeding for resistance. *Trends in plant science*, *22*(5), 420-432.

Wisser, R. J., Murray, S. C., Kolkman, J. M., Ceballos, H., & Nelson, R. J. (2008). Selection mapping of loci for quantitative disease resistance in a diverse maize population. *Genetics*, *180*(1), 583-599.

Wu, J.-H., Datson, P. M., Manako, K. I., & Murray, B. G. (2014). Meiotic chromosome pairing behaviour of natural tetraploids and induced autotetraploids of *Actinidia chinensis*. *Theoretical and Applied Genetics*, *127*(3), 549-557.

Wu, J.-H., Ferguson, A. R., Murray, B. G., Jia, Y., Datson, P. M., & Zhang, J. (2012). Induced polyploidy dramatically increases the size and alters the shape of fruit in *Actinidia chinensis*. *Annals of botany*, *109*(1), 169-179.

Wu, K., Zhang, J., Zhang, Q., Zhu, S., Shao, Q., Clark, K. D., Liu, Y., & Ling, E. (2015). Plant phenolics are detoxified by prophenoloxidase in the insect gut. *Scientific reports*, *5*, 16823.

Xie, Y. (2020). *knitr: A general-purpose package for dynamic report generation in R*. In https://cran.r-project.org/web/packages/knitr/index.html

Xin, F., Zhu, T., Wei, S., Han, Y., Zhao, Y., Zhang, D., Ma, L., & Ding, Q. (2020). QTL mapping of kernel traits and validation of a major QTL for kernel length-width ratio using SNP and bulked segregant analysis in wheat. *Scientific reports*, *10*(1), 1-12.

Yan, G., Ferguson, A. R., McNeilage, M. A., & Murray, B. G. (1997). Numerically unreduced (2n) gametes and sexual polyploidization in *Actinidia*. *Euphytica*, *96*(2), 267-272.

Yao, Z., You, F. M., N'Diaye, A., Knox, R. E., McCartney, C., Hiebert, C. W., Pozniak, C., & Xu, W. (2020). Evaluation of variant calling tools for large plant genome re-sequencing. *BMC bioinformatics*, *21*(1), 1-16.

Zelkowski, M., Olson, M. A., Wang, M., & Pawlowski, W. (2019). Diversity and determinants of meiotic recombination landscapes. *Trends in Genetics*, *35*(5), 359-370.

Zeng, H., Li, D., & Huang, H. (2009). Distribution pattern of ploidy variation of *Actinidia chinensis* and *A. deliciosa*. *Journal of Wuhan Botanical Research*, *27*(3), 312-317.

Zhan, J., Thrall, P. H., & Burdon, J. J. (2014). Achieving sustainable plant disease management through evolutionary principles. *Trends in plant science*, *19*(9), 570-575.

Zhang, H., Hu, Z., Lei, C., Zheng, C., Wang, J., Shao, S., Li, X., Xia, X., Cai, X., & Zhou, J. (2018). A plant phytosulfokine peptide initiates auxin-dependent immunity through cytosolic Ca_2^+ signaling in tomato. *The Plant Cell*, *30*(3), 652-667.

Zhang, X.-C., Millet, Y. A., Cheng, Z., Bush, J., & Ausubel, F. M. (2015). Jasmonate signalling in Arabidopsis involves *SGT1b–HSP70–HSP90* chaperone complexes. *Nature Plants*, *1*(5), 1-8.

Zhang, Y.-L., Zhang, C.-L., Wang, G.-L., Wang, Y.-X., Qi, C.-H., Zhao, Q., You, C.-X., Li, Y.-Y., & Hao, Y.-J. (2019). The R2R3 MYB transcription factor MdMYB30 modulates plant resistance against pathogens by regulating cuticular wax biosynthesis. *BMC plant biology*, *19*(1), 1-14.

Zhang, Y., Fan, J., Francis, F. d. r., & Chen, J. (2017). Watery saliva secreted by the grain aphid (*Sitobion avenae*) stimulates aphid resistance in wheat. *Journal of agricultural and food chemistry*, *65*(40), 8798-8805.

Zhang, Y., Zhang, F., Li, X., Baller, J. A., Qi, Y., Starker, C. G., Bogdanove, A. J., & Voytas, D. F. (2013). Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiology*, *161*(1), 20-27.

Zheng, Z., Qualley, A., Fan, B., Dudareva, N., & Chen, Z. (2009). An important role of a *BAHD* acyl transferase-like protein in plant innate immunity. *The Plant Journal*, *57*(6), 1040-1053.

Zhou, J., Liu, Y., & Huang, H. (2011). Characterization of 15 novel single nucleotide polymorphisms (SNPs) in the *Actinidia chinensis* species complex (Actinidiaceae). *American journal of botany*, *98*(5), e100-e102.

Zhu, Y., Yu, J., Brecht, J. K., Jiang, T., & Zheng, X. (2016). Pre-harvest application of oxalic acid increases quality and resistance to *Penicillium expansum* in kiwifruit during postharvest storage. *Food chemistry*, *190*, 537-543.

Zhuang, Q., Hou, T., & Hill, M. (2016). Predicting white peach scale phenology on kiwifruit in Sichuan. *New Zealand Journal of Crop and Horticultural Science*, *44*(3), 218-226.