The genetic diversity of natural populations of *Myzus persicae*, a polyphagous aphid species

Roland Henricus Maria Wouters

John Innes Centre

University of East Anglia
School of Biological Sciences

A thesis submitted for the degree of Doctor of Philosophy
February 2021

This page was intentionally left blank for display purposes in the paper copy of this document.

“Have no fear of perfection; you’ll never reach it.”

~ Salvador Dali

“The fool doth think he is wise, but the wise man knows himself to be a fool.”

~ William Shakespeare

“If everyone is moving forward together then success takes care of itself.”

~ Henry Ford

Acknowledgements

This PhD CASE studentship was funded by the UK research and innovation (UKRI) Biotechnology and Biological Sciences Research Council (BBSRC) as a National Productivity Investment fund CASE Award, BB/M011216/1, in collaboration with SESVanderHave.

Firstly, I must thank my PhD supervisor Professor Saskia Hogenhout for the excellent guidance, suggestions and feedback during my PhD. She taught me many essential skills in academia, and without her, I would have never reached this stage in my career. She encouraged me and gave me the interest to go further in academical sciences, especially in plant-insect and plant-microbe interaction. Thanks to her, I continued further in education via an MSc in Plant Biotechnology following my bachelor internship hosted by Saskia. This subsequently followed by this PhD CASE project. She was able to show her passion and convinced me how stunning the complex interaction between insects and plants are. I look back with most appreciation to Saskia's guidance and feel privileged to have been able to be part of the Hogenhout lab during my PhD project.

Furthermore, I would like to thank all the people who have been part of my supervisory team during my journey as a PhD student; Saskia Hogenhout, Diane Saunders, Thomas Mathers, and Sam Mugford Ian Munnery, Richard Robinson and Yazhou Chen. Both Diane Saunders and Thomas Mathers gave me great support during my PhD. They guided me in the population genetics analyses and gave great support to finalise my thesis. Thomas Mathers helped me most distinct on the teaching of the bioinformatics skill set and gave me a greater understanding of population genetics. Sam Mugford and Yazhou Chen guided me to develop the lab skills I needed to finalise my PhD. Ian Munnery and Richard Robinson helped me broadly to connect with industry and the CASE partner's company SESVanderHave. They helped me give me a greater understanding of the sugar beet industry and immensely increased my professional network.

I would also like to acknowledge the generous support from my CASE partner, SESVanderHave, during my PhD. I would like to thank Ian Munnery, Richard Robinson, Erik de Bruyne, Fokke-Jan Bulthuis, Aude Darracq, Niels Wynant and Glenda Willems. I frequently visited the company to network and discuss potential support from the company during my PhD. Erik and Fokke-Jan supported me by setting-up the capture plots and arranged the collection of the samples from the used capture plots around Europe.

Big thanks to everyone in the Hogenhout group for the enormous support they gave. Thanks for being great friends and colleagues, and also for having lots of fruitful discussion during lab meetings. I genuinely enjoyed working and having fun with all of you. Without the members of the Hogenhout group, my experience would less enjoyable and less rewarding.

I want to thank the JIC insectary staff. Without your help, my time in the insectary would have been a lot more difficult and less productive. In particular, I want to especially thank Susannah Gill for the insect assays she performed in the last legs of my studentship. Many thanks to Yogesh Gupta (The Sainsbury lab) and Darren Heavens (Earlham Institute) for nice discussions about the optimisation of ONT Nanopore sequencing of aphids. Of course, there are many more people within JIC I would like to thank that kept me in a good spirit and gave me the feeling that I would be able to succeed in my PhD project. In particular, I would like to acknowledge Anne Edwards and Gary Creissen. They were of great help with professional and leisure activities. They both have been my help and guidance around JIC, Norwich and Norfolk.

I would like to thank all the members of the local Morris dance side Kemp's Men Of Norwich for the great fun and to supply me with great memories, activities, weekends away to clear my mind and to were part of my needed breaks during my PhD. Therefore, I would like to express gratitude to the members of Kemp's Men of Norwich. I was able to explore the beautiful spots in the country and learn more about the English culture. Without the weekly practice and weekends of fun with Kemp's Men of Norwich, I would have had a lot more stress and never met this great and wonderful folk community.

Finally, I would like to dedicate this thesis to my parents Peggy and Wil Wouters. Both my parents had always supported me in all my choices. The gave an immense help during my education and the encouragement of pursuing my greatest interests. They both taught me essential values to provide me with the strength and willingness to keep pursuing my goals and interests. Without their love and guidance throughout my life in every way possible, I would have never gained this goal, and I would be forever grateful for all their support and help to accomplish my dreams. The most intriguing Dutch saying taught by my dad as a way of life was "Wie goed doet, goed ontmoet".

Abstract

The green peach aphid *Myzus persicae* has a broad plant host range of over 400 plant species from over 40 families and is a significant insect pest of many crops. Furthermore, *M. persicae* transmits approximately 100 plant viruses and has developed resistance to more than 70 insecticides. So far, knowledge of crop resistance mechanisms to aphids is limited. Sugar beet productions are particularly challenged by insect-vectored viruses and specifically those transmitted by *M. persicae*. The objectives of the research described in this thesis were (i) to identify resistance to *M. persicae* within existing sugar beet germplasm and (ii) to investigate the genetic diversity of *M. persicae* found on sugar beet in field plots around Europe. It was found there is variation in resistance to *M. persicae* among sugar beet lines and varieties. Moreover, *M. persicae* clones vary in their ability to colonise sugar beet in laboratory conditions. *M. persicae* genotype 4106a does have a lower fecundity and survival rate on sugar beet in comparison to the three other tested *M. persicae* genotypes (US1L, UK_SB and O). Existing microsatellite markers were not specific enough for aphid genotyping. I added to the current population genetics studies of *M. persicae* by a whole-genome sequencing approach of 99 *M. persicae* samples collected world-wide. I contributed to obtaining a chromosome-level *M. persicae* genome assembly of *M. persicae* genotype O, which was used as the reference genome in a genomics population approach to study the genetic diversity of 99 *M. persicae* genotypes. Between 8 to 10 populations were predicted within the sampleset of which 5 of the populations were further analysed. There was a considerable degree of differentiation among populations found with FST values ranging from 0.218 and 0.436. Population genetics analyses methods identified 11 selective sweeps, among the 5 populations. These 11 selective sweep areas were further investigated for the underlying genes involved. Multiple genes were found within the selective sweep areas. One of the sweeps associated with *M. persicae* plant host preference to *nicotianae* included candidate effector genes. In conclusion, research conducted in this PhD project showed that there is a level of resistance in sugar beet germplasm to some *M. persicae* clones and that field populations of *M. persicae* that colonise sugar beet in fields across Europe are highly diverse. These findings represent a significant step towards the identification of new approaches to obtain stable sugar beet resistance.

Table of Contents

List of tables

List of figures

List of abbreviations

ANOVA	Analysis of variance
Apaf-1	Apoptotic protease-activating factor-1
BAM	Binary sequence Alignment Map
BChV	*Beet chlorosis virus*
Blast	Basic local alignment search tool
BMYV	*Beet mild yellows virus*
BWYV	*Beet western yellows virus*
BYSV	*Beet yellow stunt virus*
BYV	*Beet yellows virus*
CC	Coiled coil
CED-4	Caenorhabditis Elegans death-4
CLR	Composite likelihood ratios
COI	Cytochrome oxidase subunit I
CTAB	Cetyl trimethylammonium bromide
DAMP	Damage-associated molecular pattern
DNA	Deoxyribonucleic acid
DPI	Days post-inoculation
dsRNA	Double-stranded RNA
ELISA	Enzyme-Linked Immuno Sorbent Assay
ETI	Effector triggered immunity
FRC	*Myzus persicae* French resistant clone
HAMP	Herbivore-associated molecular pattern
HMW	High molecular weight
HR	Hypersensitive response
LD	Linkage disequilibrium
Lec-RLK	Lectin receptor-like kinases
LRR-RLK	Leucine-rich repeat receptor-like kinases
MAMP	Microbe-associated molecular pattern
MR	Mediocre resistant
MRDH	Mediocre resistant double haplotype
mRNA	Messenger RNA
NB-ARC	Nucleotide binding with adaptor shared by APAF-1, certain R gene products and CED-4**.**
NB-LRR	Nucleotide binding-leucine rich repeat
NILR1	Nematode-induced LRR-RLK-1
NLR	Nucleotide-binding Leucine-rich-repeat receptors
Nr	*Nasonovia ribisnigri* resistance gene

NRC	NB-LRR required for hypersensitive response-associated cell death
ONT	Oxford Nanopore technology
ORF	Open reading frame
PAMP	Pathogen-associated molecular pattern
PCR	Poly chain reaction
PE	Paired end
PEPR1	Perception of the damage-associated molecular pattern peptides-1
PTI	Pattern triggered immunity
QTL	Quantitative trait loci
R	Resistant
R proteins	Resistance proteins
RAPD	Random Amplification of Polymorphic DNA
RFU	Relative fluorescent unit
RLK	Receptor like kinases
RNA	Ribonucleic acid
RNAi	RNA interference
SAM	Sequence Alignment Map
SDS	Sodium Dodecyl Sulfate
siRNA	Small-interfering RNA
SLI1	Small heat shock-like sieve element-lining chaperone-1
SNP	Single nucleotide polymorphism
SR	Semi-resistant
SRDH	Semi-resistant double haplotype
SUS	Susceptible
TUYV	Turnip yellows virus
VCF	Variant call format
WGS	Whole genome sequences
ZMW	Zero-mode waveguide

Chapter 1. General introduction

1.1 Overview

Plant pathogens and pests represent a significant threat to crops, causing dramatic yield losses annually worldwide, thereby challenging food security (Oerke, 2005, Popp and Hantos, 2013). Insects are the most abundant and diverse among eukaryotic organisms. Insects are able to survive in all habitats (Imms, 1964), and ± 500,000 insect species (half of all described species) known to date, feed on plants (Wu and Baldwin, 2010). Currently, insect pests are major contributors to crop yield losses (Oerke, 2005, Popp and Hantos, 2013) and pesticide costs. Annually, around 5.2 billion pounds are spent on pesticides worldwide (Alavanja, 2009, Grube et al., 2011). Aphids are plant-sucking insects from the order Hemiptera, which also includes, among others, whiteflies, psyllids, scales, leafhoppers, planthoppers and froghoppers/spittle bugs, and include major agricultural pests. Currently, aphid outbreaks are suppressed via insecticide applications. However, there are other ways to control aphid pests, including increasing plant resistance to these insects. Aphid resistance of crop varieties has been identified (e.g. *Nasonovia ribisnigri* resistance, Nr, in lettuce described by Helden et al. (1993). Nonetheless, obtaining durable plant resistance to aphids has been challenging. A factor contributing to this challenge is that aphids and other insects are known to have different genotypes/biotypes that overcome resistance (Arend, 2003). Examples are greenbug (*Schizaphis graminum*) biotypes that have overcome resistances of wheat & sorghum (Curvetto and Webster, 1989, Kindler et al., 2001); Russian wheat aphid (*Diuraphis noxia*) of wheat (Basky et al., 2001); rosy apple aphid (*Dysaphis plantaginea*) and woolly apple aphids (*Eriosoma lanigerum*) of apple (Rath-Morris et al., 1998, Young et al., 1982). Hence, more investigation on the nature of genetic diversity within an aphid species is needed.

1.2 Aphids

Aphids are members of the superfamily Aphidoidae in the suborder Sternorrhyncha. These insects are obligate herbivores that predominantly reproduce clonally via parthenogenesis (Figure 1), and can reach dense populations, often within only a few weeks. In total, ±4,000 aphid species across ten families are known. Of these, 250 species are described as significant pests in agriculture, horticulture and forestry (Dedryver et al., 2010, Dixon, 1998, Blackman and Eastop, 2000a). Aphid damage and yield loss are categorised as direct and indirect damage. Direct damage is caused by aphid

feeding from plant sieve elements and removal of photoassimilates, thereby weakening the plant and causing metabolic imbalance, that could trigger, in dense populations, curling of leaves and eventually plant death (Blackman and Eastop, 2000a). Additionally, some aphid species produce toxic saliva to inhibit or trigger plant defence responses and alter plant development (Belefant-Miller et al., 1994). Indirect damage is caused by pathogens transmitted by aphids. Over 2000 plant viruses are described to date, and many are known to cause yield losses in agriculture (Hull, 2014b). Over 30% of these viruses depend on aphids for transmission (Ng and Falk, 2006, Hohn, 2007, Hogenhout et al., 2008). Another indirect damage is due to the excretion of the excess sugars in the form of honeydew, which is often a preferred environment for opportunistic fungal pathogen colonisation leading to blockage of light and reduced photosynthesis. Taken together, aphids are among the most devastating pests in agriculture and horticulture (Dedryver et al., 2010).

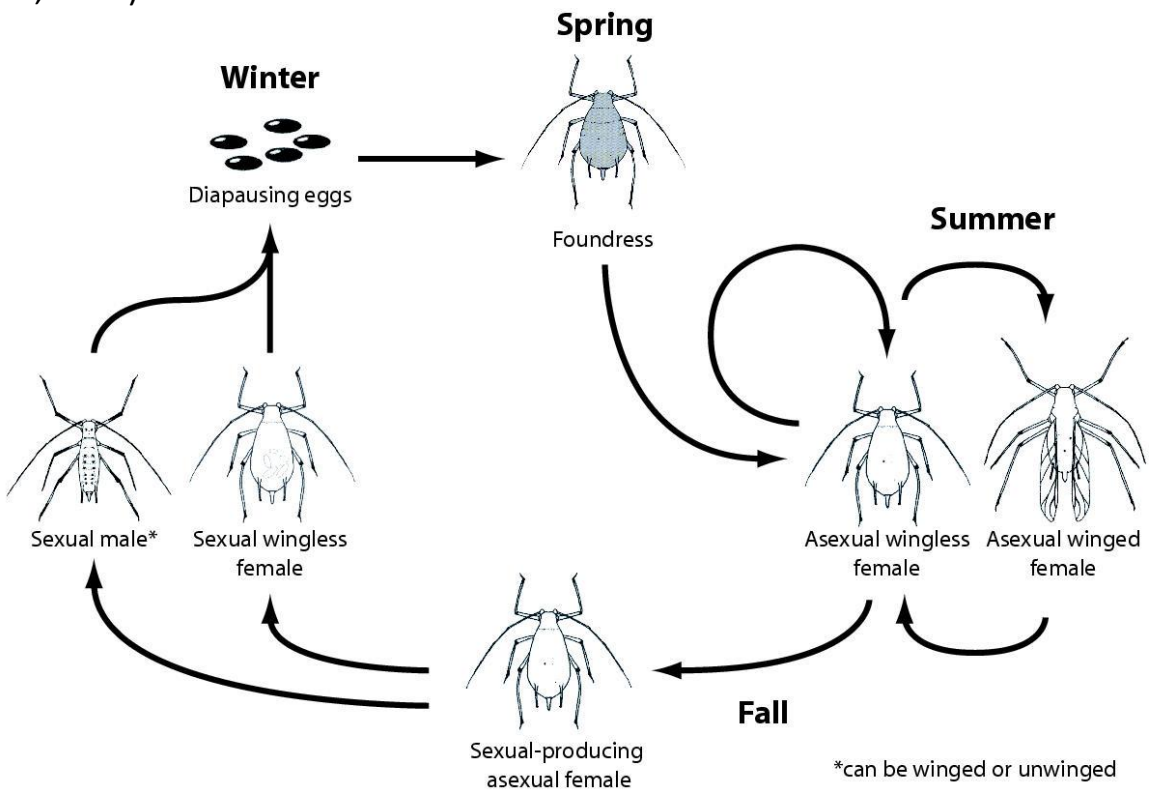


Figure 1.2.1 Lifecycle aphids. Most aphids reproduce clonally via the summer cycle. During the summer cycle, only female aphids are clonally reproduced that are either aptera (non-winged) or alate (winged) depending on the population density of the plant-host. In colder climates when there is suitable winter host available, some aphid species have different summer and winter hosts. During, shorter days in the late autumn, aphids will start to produce sexual-male and sexual-female aphids to reproduce sexually and overwinter via eggs on the winter-host. Reprinted by permission from Springer Nature (Shingleton et al., 2003). Copyright, 2003.

1.3 *Myzus persicae*

Myzus persicae (Sulzer), also known as the green peach aphid or potato-peach aphid, has a heteroecious life cycle in which the sexual reproduction of this species is restricted to peach, whereas asexual reproduction can occur on multiple plant species. The sexual females and males fly to peach trees in the autumn, and females lay eggs (oviparous reproduction) (Figure 1). When temperatures rise in spring, foundresses migrate to another plant species and produce asexual females, which reproduce clonally by giving birth to live nymphs (viviparous reproduction) (Figure 1). In the United Kingdom, *M. persicae* are thought to reproduce asexually predominantly and may 'skip' the sexual reproduction phase in winter; they may overwinter on any plant species. Hence, *M. persicae* populations in the United Kingdom are believed to predominantly consists of clonal lineages that may or may not have different genotypes.

M. persicae is known as one of the most notorious pests worldwide (Wu and Baldwin, 2010, Blackman and Eastop, 2000a). It is unusual among aphid species, as it can colonise over 400 different plant species from 50 different plant families. *M. persicae* is considered the most polyphagous among aphid species. In contrast, approximately 90% of the aphid species are specialised and do only colonise a single or up to three related plant host species (Schoonhoven et al., 2005). *M. persicae* also transmits over 100 different plant viruses, which may be acquired by this aphid when feeding on weeds adjacent to the crops.

M. persicae is used as a model organism for studying plant-insect interactions because it can colonise the model plants *Arabidopsis thaliana* and *Nicotiana benthamiana*. Additionally, the functions of *M. persicae* genes can be studied via knockdown by plant-mediated RNA interference (RNAi) (Pitino et al., 2011, Bos et al., 2010). For example, *M. persicae* genes encoding virulence factors (effectors) were knocked down via RNAi. The genomes of *M. persicae* clones G006 and O were sequenced (Mathers et al., 2017). *M. persicae* clone G006 was collected in the United States of America (USA), and clone O in the United Kingdom. The sizes of both *M. persicae* genomes are about 350Mb encoding ± 18,000 annotated genes (Mathers et al., 2017). The genome sequences of several other aphid species, including for example *Acyrthosiphon pisum* (pea aphid) and *Aphis glycines* (soybean aphid), are also available (International Aphid Genomics, 2010, Wenger et al., 2017), and can

be used for comparative genome analyses among aphids and other hemipterans.

Within the species *M. persicae*, there may be subspecies that appear to have different preferences for plant species. For example, *M. persicae nicotianae* are better at colonising *Nicotiana* species (Blackman, 1987, Field et al., 1994). However, the ability of *M. persicae* clones to colonise *Nicotiana* species appears more common, as clone O easily switches between *N. benthamiana*, *Arabidopsis thaliana* and *Brassica rapa* (Mathers et al., 2017) and other plant species (Chen et al., 2020). Other *M. persicae* clones can also colonise *N. benthamiana*, including clone FRC (pers. communication, Yazhou Chen in Hogenhout lab 2020).

M. persicae predominantly reproduce clonally during spring and summertime. When reared in greenhouses or controlled environments rooms, generally no male adults are produced, and all the reproduction occurs via female clonal production. Throughout clonal reproduction, nymphs are born from single female adults. Moreover, the nymphs have already developed nymphs within their bodies before they are born. This telescoping type of reproduction may be compared to a Russian nested doll set. Furthermore, the nymphs originated from a single adult aphid are genetically near-identical to their mothers. Thus, a population originated from a single adult aphid is seen as a clonal population. Multiple clonal populations can belong to the same genotype, but clonal populations are unlikely to have different genotypes. Previously, aphid colonies maintained at the John Innes Centre (JIC) Entomology Facility were not genotyped, except for clone O (Bos et al., 2010) and therefore they were referred to as clones. Upon genotyping them in Chapter 3 of this thesis, I referred to these colonies as genotypes.

1.4 Plant viruses

Countless plant species cause disease. Currently, the International Committee on Taxonomy of Viruses described over 6500 viruses that infect eukaryotic hosts, and of these, 2000 infect plants (Gergerich and Dolja, 2006, Hull, 2014a, ICTV, 2019). However, more plant viruses have been described in recent years. Virus particles do depend on the machinery of host cells for replication; plant viruses need other organisms or environmental support for dissemination. Relatively to animal viruses, plant viruses are more likely to be transmitted via a vector (Lefeuvre et al., 2019, Liu et al., 2016). Because plants are mostly sessile organisms that are restricted to one geographic area, plant

viruses have evolved a plethora of strategies to move between plants. Plant viruses are able to transfer between hosts via seed, pollen, touch and vectors such as nematodes, fungi, and insects, and parasitic plants (Jones and Naidu, 2019). Insect vectors transmit over 75% of the plant viruses. Insects of the Order Hemiptera, such as whiteflies, leafhoppers and aphids, transmit the majority of the insect-transmitted plant viruses (Andret-Link and Fuchs, 2005, Hogenhout et al., 2008, Gautam et al., 2020) and may be called the 'mosquitoes of the plant world' (quoted from Prof. Saskia Hogenhout).

Plant viruses constitute a significant threat to agriculture, with major economic losses estimated over 30 billion dollars a decade ago (Sastry and Zitter, 2013). Losses attributable to plant viruses are caused because of reduced plant growth and vigour and subsequent diminished gross yield. In addition, the symptoms induced by virus infections, such as leaf yellowing, leaf curling, or other growth distortions, can reduce the product quality and make it unsellable, thereby diminishing marketable yields (Stevens et al., 2004, Stevanato et al., 2019). Aphids transmit approximately 30% of the described plant viruses.

Aphids transmit viruses in a non-persistent, semi-persistent and persistent non-replicative manner or in a persistent replicative manner. Non-persistent viruses are acquired rapidly within seconds to a few minutes, but do only retain in their vector for a short time, and the virus will be lost when the vector moults. Semi-persistent viruses are acquired within minutes and are retained for up to a few days and persistent viruses that take hours to acquire, but survive in the vector for a day up to the vector's lifetime (Ng and Falk, 2006). Moreover, due to a longer retainment period to acquire the virus, semi-persistent viruses do require a suitable plant host for the vector in order to acquire the virus, but not crucial in transmission of the virus into the plant.

M. persicae is one of the most significant pests of sugar beet (*B. vulgaris*) (Hauer et al., 2017, Lange, 1987), mostly due to this aphid species being an efficient vector of multiple sugarbeet viruses. Additionally, *M. persicae* is a polyphagous insect species. At least five viruses contribute to yield losses in *B. vulgaris*; including *Beet mild yellows virus* (BMYV; Genus *Polerovirus*), *Beet yellows virus* (BYV; Genus *Closterovirus*), *Beet chlorosis virus* (BChV; *Polerovirus*), *Beet yellow stunt virus* (BYSV; *Closterovirus*) and *Beet western yellows virus* (BWYV; *Luteovirus*; table 3.1). In central and western Europe, BMYV and BYV are responsible for the most significant yield losses in sugar

beet, particularly if aphids are not being controlled with pesticide treatments (Qi et al., 2004, Noleppa, 2017, Hauer et al., 2017).

Table 1.4.1 Viruses that commonly infecting sugar beet (*B. vulgaris*) and that are predominantly transmitted by *M. persicae*. *Number of open reading frames (ORFs). ** S-Pers. = semi-persistent, Pers. = Persistent. ***predicted percentage of maximum yield loss. +ssRNA = positive-sense single-stranded RNA genomes.

Virus	Genus	Structure	Size (bases)	ORFs*	Type **	Yield loss***	References
***Beet yellows virus* (BYV)**	*Closterovirus*	+ssRNA	15,480	9	S-Pers.	50%	Agranovsky et al. (1994)
***Beet yellow stunt virus* (BYSV)**	*Closterovirus*	+ssRNA	10,545	10	S-Pers.	40%	Karasev et al. (1996)
***Beet mild yellowing virus* (BMYV)**	*Polerovirus*	+ssRNA	5,722	6	Pers.	40%	Guilley et al. (1995)
***Beet western yellows virus* (BWYV)**	*Luteovirus*	+ssRNA	5,641	6	Pers.	40%	Veidt et al. (1988)
***Beet chlorosis virus* (BChV)**	*Polerovirus*	+ssRNA	5,776	6	Pers.	30%	Hauser et al. (2002) & Stevens et al. (2004)

BYV belongs to the genus *Closterovirus* and is one of the earliest described, and most studied plant viruses. Aphids transmit closteroviruses in a semi-persistent manner. Semi-persistent viruses are acquired and transmitted within brief feeding episodes by aphids, and the virus particles adhere to the upper inner parts of aphid stylets and are lost upon aphid moulting. The aphid usually acquires the virus within about 5-10 minutes and is able to transmit the virus within a period of up to 2 days (Cockbain et al., 1963). The virus does not replicate in the aphids and the virus is therefore considered non-propagative. Semi-persistent viruses do not colonise gut, hemolymph and salivary glands of aphids and are therefore considered non-circulative. In plants, BYV is mainly restricted to the phloem, and is hence considered phloem-limited. Viruliferous aphids introduce BYV into the phloem sieve elements along with their saliva while salivating and acquire the virus when feeding. Various aphid species are known to transmit BYV to sugar beet, including *M. persicae*, *Rhopalosiphum padi (L.)*, *Acyrthosiphon kondoi* (Shinji), *Diuraphis noxia* (Kurdjumov) and *Aphis fabae* (Scopoli) (Summers et al., 1990). With a BYV transmission rate of 60%, apterous *M. persicae* clones are approximately twice as efficient in transmission of BYV compared to *A. fabae* (Limburg et al., 1997). Aphid species that colonise sugar beet throughout the

growing season are *M. persicae*, *A. fabae* and *Macrosiphon euphorbiae*, but most damage of BYV is caused by *M. persicae* that transmits this virus to sugar beet primarily at the start of the sugar beet growing season, because *M. persicae* appears earlier in the season (Qi et al., 2004). Sugar beet becomes more resilient to virus infection in later stages of development. Therefore, BYV transmission by aphid species that appear later in the year compared to *M. persicae* are considered to have minor contributions to sugar beet losses caused by BYV.

BYV is one of the most prominent examples of a destructive pathogen. This virus causes virus yellows disease that is responsible for up to 50% yield loss in root crops (Agranovsky et al., 1994). BYV is a fast-growing threat. In 2019, the proportion of infected sugar beet fields in the United Kingdom increased substantially to 55% from 18% in 2018 (Gillbard, 2020). BYV infection is detected all over Europe. In 2020, predicted yield losses by BYV of sugar beet amounted to 50% (Belgium), 30% (Germany), 25% (France) (BBRO, 2020). Due to increasing populations of (viruliferous) aphids and limited means for curbing the spread of these insects, the situation is highly likely to get worse in the future. Hence, the BYV threat is of extremely high economic importance for the European sugar beet production. As the 3rd largest sugar producer in the world, the EU sugar beet industry contributes €3.6 billion direct (and €15.6 billion indirect) to the GDPs of the European Union.

BMYV is a member of the genus *Polerovirus*, which belongs to the Family *Luteoviridae*. Poleroviruses are predominantly aphid-transmitted viruses. Aphids transmit poleroviruses in a persistent manner, meaning that these viruses do circulate throughout the aphid, including the gut, hemolymph and salivary glands from where the viruses are introduced into plants when aphids feed. However, the virus does not replicate in aphids. Upon acquisition of the virus, it takes days for aphids to become vectors, because the virus needs to circulate within aphid bodies to reach aphid salivary glands. However, aphids carry the virus throughout their lifespan (Sylvester, 1980, Schliephake et al., 2000). The aphid needs sustained feeding from the plant phloem to acquire and transmit BMYV. With 90% transmission efficiency, *M. persicae* is thought to be the predominant vector of BMYV in sugar beet (Schliephake et al., 2000). Like BYV, BMYV is mainly restricted to the phloem tissues of plants. BMYV is considered to be one of the biggest threats to *B. vulgaris* production in the growing season (Hossain et al., 2020). Especially with the current restrictions of the use of neonicotinoids in Europe.

1.5 Sugar beet

Sugar beet (*Beta vulgaris* L.) is grown in temperate and continental climates all over the world. More recently, sugar beet is also grown in more tropical areas (Zhang et al., 2008, Cooke and Scott, 2012). Sugar beet is an important high-value field crop conventionally used for sugar production. Presently, a growing trend is observed for biofuel and biogas production (Zhang et al., 2008). *B. vulgaris* has several subspecies. The subspecies *vulgaris* is known as the cultivated sugar beet, and wild (non-cultivated) subspecies are known as subspecies *maritima* (L.), *adanensis* (Pamuk.), *macrocarpa* (Guss.), *patula* (Ait) and *trigyna* (Waldst & Kit) (Lange, 1999, Castro et al., 2013). The subspecies *vulgaris* is clustered in 5 groups; Sugar beet (Altissima), fodder beet (Crassa), garden beet (Conditiva), spinach beet (Cicla) and swiss chard (Flavescens) (Porcher, 2005).

In current regulations, prices of sugar beet are contracted before sowing (Qi et al., 2004). Therefore, the profit of the crop can be directly translated from the yield. For the biggest yield, an optimal growing environment needs to be sustained during the growing season, and various strategic decisions are continuously being made and improved in the sugar beet industry (McLean et al., 1986, Qi et al., 2004, Hauer et al., 2017). Pests and pathogens during the sugar beet growing season are well monitored. Many companies have developed guidelines for usage of treatments against the pathogens and pests. These guidelines are used to assess the spraying regimes and the usage of neonicotinoids. Currently, there are no *M. persicae* resistant sugar beet cultivars on the market. Thus, the farmer depends on the use of these pest management strategies to control this pest.

1.6 Pest management

Application of chemical insecticides is the predominant method for aphid pest management. Hence, most research of industry and many academic labs have focused on developing new insecticides to control *M. persicae* and other aphid/hemipteran pests. Pesticides of the neonicotinoid class were discovered in the 1980s. However, worldwide, neonicotinoids have been commonly used for insect control only since 1991. Neonicotinoids are nicotine-like chemicals classified as neuro-active insecticides (Jeschke et al., 2011). These insecticides move systemically within the plant and hence can be applied as a seed coating, which protects, for example, sugar beet and oilseed

rape, against insect pests, such as *M. persicae,* for at least the first six weeks of plant growth (Nauen et al., 2008, Jeschke et al., 2011, Simon-Delso et al., 2015). In contrast, other pesticides, such as organophosphates, carbamates and pyrethroids, require application via foliar spraying. Neonicotinoids are also effective against a broader range of other pests, including most hemipterans (aphids, whiteflies and leafhoppers) and coleopterans (beetles), but also Gastropoda (slugs and snails) (Nauen et al., 2008, Thompson, 2012, Noleppa, 2017), contributing to an increase of neonicotinoid usage globally. By 2008, 80% of all crops were protected by neonicotinoid, mostly applied as a seed coating. Neonicotinoids are currently registered for use in crop protection in more than 120 countries. In 2006, neonicotinoids had a market value of ±1.7 billion US-dollars (Nauen et al., 2008). Pesticides in the neonicotinoid family include imidacloprid, acetamiprid, nitenpyram, clothianidin, thiacloprid, thiamethoxam and nithiazine; until 1999 imidacloprid was the most used insecticide worldwide, followed by thiamethoxam (Maienfisch et al., 2001, Simon-Delso et al., 2015, Hauer et al., 2017). Thiamethoxam and imidacloprid are effective against sucking insects, some chewing insects, soil-inhabiting insects and animal fleas.

Without pesticide control of aphids, aphid-vectored viruses are known to cause yield losses of up to 50% in sugar beet alone, rendering this crop unprofitable in the United Kingdom (Noleppa, 2017, Hauer et al., 2017). Application of chemical insecticides is the predominant method for aphid pest management. Hence, most research has focused on developing new insecticides to control *M. persicae*.

M. persicae has evolved resistance to thiamethoxam, and possibly other neonicotinoids (Bass et al., 2014) and to around 70 non-neonicotinoid compounds, such as organophosphates, carbamates and pyrethroids (Silva et al., 2012, Bass et al., 2014). At least two neonicotinoid resistant genotypes of *M. persicae* are found in Europe. One is found in South France (French Resistance clone; FRC) and the other in Greece (5191A). FRC and 5191A have evolved two distinct mechanisms for neonicotinoid resistance: 1. Over-expression of detoxifying cytochrome P450s monooxygenases, which induce production of detoxifying enzymes (Puinean et al., 2010); 2. Mutations in nicotinic acetylcholine receptor β subunit resulting in reduced sensitivity to imidacloprid (Bass et al., 2011).

Neonicotinoids have damaging effects on bee and bird populations (Heimbach et al., 2016, Peters et al., 2016, Rolke et al., 2016a, Rolke et al., 2016b, Sterk et al., 2016, Schmuck and Lewis, 2016, klages, 2016, Thompson, 2012, Wu-Smart and Spivak, 2016). All neonicotinoids are soluble in water causing an unwanted dispersal of neonicotinoids in nature (Adak et al., 2012, Goulson and Kleijn, 2013, Bonmatin et al., 2015). Specific neonicotinoids on all flowering field crops and vegetables (e.g. oilseed rape) have been banned since 2013 in Europe (Hauer et al., 2017). Neonicotinoids have been entirely banned in Europe since 2019. Yield loss has already occurred for oilseed rape productions that were challenged by outbreaks of stem flea beetles, slugs and cabbage aphids (Noleppa, 2017). It is estimated that a lack of effective pest management strategies will cause yield losses of up to 80% due to damage by insect pests (Qi et al., 2004) and will cost the economy of Europe approximately 3.4 billion euro per year (Hauer et al., 2017). Thus, with the recent ban of neonicotinoids in Europe, there is a greater need for alternative pest control methods, such as biocontrol and increasing plant resistance via breeding efforts.

Upon banning of the core neonicotinoid active compounds clothianidin, imidacloprid and thiamethoxam were banned (Noleppa, 2017, Jactel et al., 2019), SESVanderHave and others noticed significant losses because of higher colonisation of *M. persicae* in *B. vulgaris* fields and virus occurrence, causing sugar beet yield losses at a similar level as was observed before neonicotinoids were introduced (Noleppa, 2017, Hauer et al., 2017). It is estimated that a lack of effective pest management strategies will cause yield losses of up to 80% in sugar beet (Qi et al., 2004), and will cost the economy of Europe approximately 3.4 billion euro per year (Hauer et al., 2017). There is clearly a greater need for alternative pest control methods, such as biocontrol and possibly increasing plant resistance via breeding efforts.

Because of the devastating yield losses in several crops, and from BYV in sugar beet, there is a temporary removal of the neonicotinoid ban in some countries, such as France. The United Kingdom has subsidised farmers who suffered yield losses because of BYV infection and has allowed small quantities of neonicotinoid-based controls in order to mitigate the possibility of higher aphid occurrence due to lower temperatures of previous winters. It is hoped that with these measures, UK farmers will still grow sugar beet. It of utmost urgency to develop new technologies to help stop BYV spread. Specific, currently available plant protection methods against viruses are based on breeding disease-resistant plant varieties and crop management. However,

the above solutions for BMYV and BYV control have been slow and/or ineffective. There are presently no BYV-resistant beet varieties. Breeding new beet varieties are costly (~€60m EUR per annum (Bruins, 2020) and slow (1 year per generation) (Panella and Lewellen, 2006). A partial crop management solution is to remove beet debris after harvest as this was shown to decrease the chance BYV occurrence in the next growing season. However, removing beet debris is not practical for medium-and large-size farms, nor does it curb the transmission when the disease has already spread.

1.7 Plant resistance

1.7.1 Defence mechanisms

Plants have evolved numerous defences and barriers to defend against pathogens and pests. Plant defence first starts with barriers such as trichomes, lignin-rich areas such as bark, waxy epidermal cuticles, thorns, cell walls and secondary metabolites. These barriers are constitutive (continuous) and are not induced by pest or pathogens (Freeman and Beattie, 2008). Moreover, plants have developed defence responses that are induced upon recognition of pathogens and pests. Plants rely on an innate immune response because their defence lacks an adaptive immune system (Boller and Felix, 2009, Jones and Dangl, 2006). To recognise non-self signals and respond to these, plants have diverse membrane-spanning receptors that have extracellular domains, which sense pest-derived signals (often called Pathogen-/Microbe-/Damage-/Herbivore association molecular patterns; PAMPs/MAMPs/DAMPs/HAMPs) in the apoplast and that are connected to kinase-containing domains or co-receptors inside the cell to induce a specific signalling pathway, known as pattern triggered immunity (PTI) (figure 1.7.1a), which propagates intracellularly (Zipfel, 2008) and extracellularly (Wu et al., 2014). Together, the physical barriers and PTI give plants resistance to the majority of pathogens and pests and are likely to be responsible for the phenomenon known as 'non-host resistance'.

Most successful pathogens and pests evolved to cope with the plant physical barriers and PTI. These pathogens and pests secrete virulence proteins, also known as effectors, to suppress PTI enabling colonisation of the plant host (figure 1.7.1b) (Hogenhout et al., 2009, Wu and Baldwin, 2010, Dodds and Rathjen, 2010). Some effectors modulate or target specific plant

targets, which are known as susceptibility genes (S genes), to increase plant susceptibility (Eckardt, 2002). Nevertheless, plants developed mechanisms to sense effectors or their actions resulting in a proper defence against specific pathogens and pests. The effector (actions) are often detected by cytoplasmic receptors, known as intracellular resistance (R) proteins or NLRs (nucleotide-binding Leucine-rich-repeat receptors) (Belkhadir et al., 2004), resulting in effector-triggered immunity (ETI) (figure 1.7.1c) (Takken and Goverse, 2012, Dodds and Rathjen, 2010). ETI is often associated with a hypersensitive response (HR), which results in local cell death (and limiting spread of a pathogen or feeding of a pest). Pathogen/pest effectors and NLRs are involved in an arms race; both are under selection for mutations or polymorphisms to increase or prevent infection, respectively.

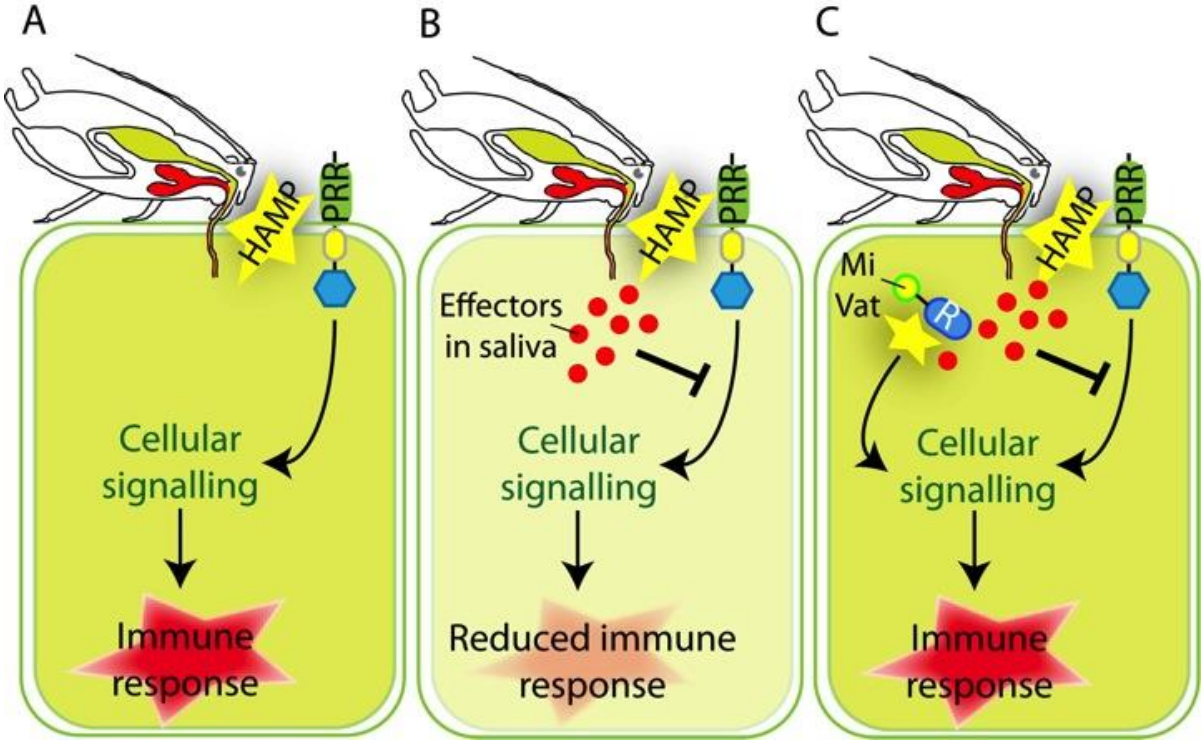


Figure 1.7.1 The 'ZigZag model' for plant-aphid interaction. Model of the co-evolution of plant receptors and insect effectors. **A.** HAMPs (Herbivore Associated Molecular Patterns) are recognised resulting into a PTI-like immune response. **B.** Aphid effectors suppress the recognition or downstream response of PTI induction. **C.** Plant NLRs (Nucleotide-binding Leucine-rich-repeat receptors) recognise an effector and trigger the hypersensitive response, which leads to an immune response. Reprinted from (Hogenhout and Bos, 2011). Copyright (2011), with permission from Elsevier.

The different levels of plant-biotic interactions is often referred to as the "ZigZag model". The "ZigZag model" describes the molecular dialogue between plants and pathogens (Jones and Dangl, 2006) and has gained support in the well-studied model plants *Nicotiana benthamiana* and *Arabidopsis thaliana* (Segonzac and Zipfel, 2011, Dodds and Rathjen, 2010,

Cook et al., 2015). In the apoplast, a conserved molecule, called pathogen/microbial-associated molecular pattern (P/M/AMP), interacts with the extracellular domain of a transmembrane pattern recognition receptor (PRR) triggering PTI. In compatible interactions, effectors produced by pathogens/pests are able to suppress this defence response, leading to effector-triggered susceptibility (ETS) and colonisation of the plant host. However, plants have specific R genes that can recognise specific effectors triggering ETI, which is also known as 'gene-for-gene interaction'. PTI, ETS and ETI of the "ZigZag model" are part of an ongoing arms race between plants and pathogens (Jones and Dangl, 2006). Nevertheless, recently Cook et al. (2015) discussed that the model is too specific and proposed a uniform name in plant response that combines PTI and ETI, named invasion pattern triggered response (IPTR).

The "ZigZag model" was further expanded to also include insect herbivores, such as aphids, in which PTI may be triggered by herbivore-associated molecular patterns (HAMPs) or egg-associated molecular patterns (EAMS) (Figure 1) (Hogenhout et al., 2009, Hogenhout and Bos, 2011, Reymond, 2013, Reymond and Calandra, 2015). During feeding, aphids secrete saliva, including effectors, into plant cells (Bos et al., 2010, Hogenhout and Bos, 2011, Rodriguez et al., 2014b, Jaouannet et al., 2014, Reymond and Calandra, 2015, Coleman et al., 2016) and these effectors are thought to modulate host processes to enable aphid colonisation (Will et al., 2007, Mutti et al., 2008, Mugford et al., 2016). The aphid salivary effectors are likely to target similar defence responses as described for bacterial and fungal effectors. Hogenhout and Bos (2011) gave an overview of a model of herbivore-plant interaction (figure 1).

1.7.2 Nucleotide-binding domain, leucine-rich repeat-containing receptor proteins (NLRs) and Receptor-like kinases (RLKs)

Nucleotide-binding domain, leucine-rich repeat-containing receptor proteins (NLRs) and receptor-like kinases (RLKs) are the eyes and whistle-blowers of the plant. RLKs are surface localised receptors, and NLRs are cytoplasmic receptors. These two groups of receptors are the main components of the plant immune system. Yet, RLKs are also receptors for many essential plant physiological and cellular processes (Tor et al., 2009). For instance, RLKs have crucial roles in stem cell maintenance (De Smet et al., 2009), lateral root meristem initiation (De Smet et al., 2008), and stomata development (Geisler et al., 2000).

Plants rely on receptor proteins to activate immune responses upon recognition of non-self particles. Both NLRs and RLKs play a crucial role in the recognition and activation of defence responses (Dodds and Rathjen, 2010). NLRs are part of an arms race of recognition and evasion of effector particles. This results in that NLRs are rapid-evolving and found to be one of the highly diverse protein families in plants (Jacob et al., 2013). NLRs can either directly or indirectly bind the non-self particle. The NLR that induce an immune response is seen as a sensor that needs, in some cases, a helper NLR to trigger the defence responses such as a hypersensitive cell-death response (Wu et al., 2016). Genes of sensor NLR proteins are highly divergent among plant species and hence appear to evolve rapidly, while helper NLR proteins, including for example the nucleotide binding-Leucine rich repeat (NB-LRR) called REQUIRED FOR HYPERSENSITIVE RESPONSE-ASSOCIATED CELL DEATH 1 (*NRC1*) (Gabriels et al., 2007), are more conserved among plant species (Baggs et al., 2017). This makes sense, because NLR proteins have to recognise effectors of highly divergent pathogens and pests, whereas helper NLR proteins assist with signalling.

NLR proteins consist of a N-terminal nucleotide-binding domain followed by a non-conserved domain and a leucine-rich repeat at C-terminal. The nucleotide-binding domain is known as NB-ARC, named after the nucleotide-binding resistance protein discovered in the NB-ARC members *Apaf-1* (apoptotic protease-activating factor-1) and *CED-4* (Caenorhabditis elegans death-4 protein; (van der Biezen and Jones, 1998)). Many genetic redundancies are built in the NLR network to make it a robust network as protection against non-self molecules from pathogens, pests, or any other threats. Interestingly, *Beta vulgaris* has only one NLR-helper, while most other plant species have multiples of NLR-helper of which some over 10 (e.g. *Solanum tuberosum* and *Coffea arabica*) (Wu et al., 2017).

RLKs have an extracellular domain that is linked to a cytoplasmic kinase domain via a transmembrane region. RLKs comprise the largest gene family of receptors in plants, with over, for instance, 600 found in Arabidopsis and 1100 in rice (Shiu et al., 2004). Moreover, multiple types of RLKs are defined, of which one type is defined as lectin receptor-like kinases (Lec-RLKs) that are effective against insects. These Lec-RLKs are membrane-spanning with the Kinase intracellular and the lectin domain extracellular (Bouwmeester and Govers, 2009). The Lec-RLK family members are functional in diverse plant

processes from plant growth, abiotic stress recognition and recognition of plant pathogens and pests (Humphrey et al., 2007). The membrane-spanning receptors are categorised in three subgroups (L, C and G type) due to the different lectin domain structures (Vaid et al., 2013). The C-type LEcRLKs is named for their calcium-dependent lectin structure (Shiu and Bleecker, 2001). The G-type for their GNA-related (*Galanthus nivalis* agglutinin-related) structure at the extracellular lectin domain (Van Damme et al., 2007), and the L-type for their legume-lectin domain (Herve et al., 1999). Distinct functions in the plant are categorised of each type of Lec-RLKs, but all three types are found to be complemented in defence responses and immunity. L-type Lec-RLKs are well described as receptors that confer resistance to a variety of plant-pathogens (Chen et al., 2006, Gilardoni et al., 2011, Singh et al., 2012) and insect feeding (Gilardoni et al., 2011, Bonaventure, 2011). G-type Lec-RLKs are known to function in both plant defence as self-incompatibility receptors (Kusaba et al., 2001, Cheng et al., 2013). C-type Lec-RLKs are described as factors in immunity in mammals (Geijtenbeek and Gringhuis, 2009), but are rare in plants.

1.7.3 Secondary metabolites

Other more passive resistance mechanisms are chemical compounds made by the plant and are often called secondary metabolites. Secondary metabolites are organic compounds that are not primarily involved in growth and development, but they are used for interaction with its environment such as attractants for pollinators and repellents and toxins for pathogens and pests (D'Alessandro and Turlings, 2006, Snoeren et al., 2010, Piasecka et al., 2015) . Moreover, secondary metabolites that function as toxins against pathogens and pests are commonly used as commercial fungicides and pesticides (Rattan, 2010).

Some examples of secondary metabolites that are involved in plant defences against pest and pathogens are camalexin and pyrethroid. Camalexin is an alkaloid compound that is found in crucifer species such as *A. thaliana* and works as a phytoalexin to deter insect pests and fungal and bacterial pathogens (Glawischnig, 2007, Ausubel et al., 1995, Kettles et al., 2013). Pyrethroid is a secondary metabolite group of organic compounds such as pyrethrins, produced in flowers of pyrethrums. Pyrethroids are commercial used products as pesticides against most insect pests (Soderlund, 2012, Ranson et al., 2011). These pyrethroid chemical compounds are incorporated in many pesticide topical sprays to kill the insect pest but do not work

systemically within the plant. These secondary metabolites used commercial as sprays, are not specific and could harm beneficial insects. Therefore, while these products were used predominantly in the past years (Schuman and Baldwin, 2016) , more recently, more substantial development occurred for resistance breeding against specific insect pests.

1.7.4 Resistance breeding

1.7.4.1 Resistance mechanisms against insects

Resistance against aphids is defined by antixenosis and antibiosis. Antixenosis disrupts insect behaviour (Kennedy and Kishaba, 1977, Klingler et al., 2005), and antibiosis disrupts insect survival, growth and development (Kaloshian et al., 2000). Aphid behaviour may be monitored via the Electrical Penetration Graph (EPG) technique (Tjallingii, 1978), which measures small changes in electrical currents between the aphid and soil to assess the location of the stylets within plant tissues and cells and whether the aphid salivates or acquires cell contents (Tjallingii, 1978, Tjallingii, 1985). Aphid antibiosis is often investigated by counting reproduction and survival rates of this insect.

RLK/Receptor like proteins (RLP) cell surface receptors and NBS-LRR-type intracellular receptors both play a role in plant defence responses to insect herbivores (Table 1.7.1). So far, there are two examples of RLK/RLP-mediated resistance to insect herbivores. One RLK found in rice is mediated by the *Bph17* locus, which encodes genes *OsLecRK1-OsLecRK*, and give resistance to biotypes of the brown planthopper *Nilaparvata lugens*. Given that these proteins recognise lectins, the resistance response is mostly triggered by a lectin-like HAMP of the planthopper. However, no HAMP has been identified. RLK/RLP-mediated defence pathways also play a role in plant responses to aphids. Firstly, BAK1 that is an RLK co-receptor is involved in triggering defence responses to aphid elicitors (Prince et al., 2014, Kaloshian and Walling, 2016). Moreover, a locus enriched RLKs has been associated with resistance to *M. persicae* in pepper (Sun et al., 2020b, Sun et al., 2020a).

There are more NBS-LRR-type genes that give resistance to insects cloned (Table 1.7.1). The first one is *Mi1-2* cloned from tomato (*Solanum peruvianum*). This gene was initially characterised to give resistance to the nematode *Meloidogyne incognita*. However, it also gives resistance to the potato aphid *M. euphorbiae* and to whiteflies. The resistance to *M. euphorbiae* is biotype specific. The resistance occurs at the time the aphid stylets approach the phloem (Kaloshian et al., 2000) and it has been noted

that these aphids probe more on *Mi1-2* plants (Kaloshian et al., 2000) thereby possibly transmitting more potyviruses (Boissot et al., 2016a). Similarly, *Vat* is an NBS-LRR gene that gives resistance to the cotton-melon aphid *Aphis gossypii*, particularly a biotype of this aphid that is prevalent in France. *Vat* also gives resistance to several potyviruses transmitted by *A. gossypii*. Like *Mi1-2*, the resistance of *Vat* is phloem-mediated (Walling, 2008). By far the most NBS-LRR-type genes are cloned from rice to the brown planthopper (Table 1.7.1). There are many rice brown planthopper biotypes known (Khush et al., 1985, Jena and Kim, 2010). These NBS-LRR-type genes give resistance to one or more of these biotypes. Several resistance genes were mapped but not yet cloned and many mapped loci appear to be enriched for NBS-LRR genes. For instance *Gm* genes against the gall midge *Orseolia oryzae* in rice (Li et al., 2020b, Suvendhu et al., 2014), *H* genes in the hessian fly (*Mayetiola destructor)* in wheat (Yu et al., 2009), *RAG* genes against the aphid *Aphis glycines* in soybean (Zhang et al., 2010), *Nr* genes against the aphid *Nasonovia ribisnigri* in lettuce (Helden et al., 1993) and *Dn* genes against the aphid *Diuraphis noxia* in wheat (Liu et al., 2005).

Partial resistance against *M. persicae* are described for some crops and trees. Yet, all these found resistance genes are mapped and not cloned. For example, firstly in pepper, a reduction of phloem intake is found by *M. persicae* on a wild accession *Capsicum baccatum*. This reduction of phloem intake by *M. persicae* was induced by callose deposition in the sieve elements (Sun et al., 2018). More recently, Sun et al. (Sun et al., 2020a) described a quantitative trait loci (QTL) finding in pepper for aphid resistance. This resistance resulted in reduced aphid survival rates and fecundity. Interestingly this QTL was fine mapped to a region with leucine-rich repeat receptor-like kinases (LRR-RLK) analogues. Moreover, in Arabidopsis, (Kloth et al., 2017) a the small HEAT SHOCK-LIKE SIEVE ELEMENT-LINING CHAPERONE1 (*SLI1*) was found to reduce phloem intake of *M. persicae* and to increase salivation during the feeding. Thirdly, in peach, eight QTLs were found to have increased resistance via an underlying change in aphid feeding. *M. persicae* feeding on the peach plants with the QTLs were salivating more in the sieve element than on other peach trees (Sauge et al., 2012, Lambert and Pascal, 2011). Thus, even though partial resistance against *M. persicae* does exist in nature and is described, no genes have been cloned so far.

The NBS-LRR-type genes identified to give resistance to insect are all CC-NBS-LRRs, in which the CC stand for coiled-coil that is found to be involved with communicating with the helper NLRs, such NRC1, 2 and 3 (Wu et al.,

2016). One caveat of CC-NBS-LRRs is that some appear to lose their activity in warm temperatures, for example Mi1-2 does not work at temperatures above 32°C (Cooper et al., 2005) This makes this gene not useable for warmer countries. So far, there are no TIR-NBS LRRs that give resistance to insects cloned. Unlike CC-NBS-LRRs, TIR-NBS-LRRs trigger defence responses via *EDS1*/*PAD4* (Nandety et al., 2013). Nonetheless, PAD4 does mediate resistance to *M. persicae*, whereas EDS1 does not (Louis and Shah, 2015, Louis et al., 2012).

So far, only few plant genes that give resistance to insects have been cloned; these include seventeen genes that give resistance to 4 insect species of the order Hemiptera and one to an insect species of the order Lepidoptera. Most resistance genes are cloned from rice against the brown plant hopper (*Nilaparvata lugens)* (Yang et al., 2019, Du et al., 2020). Only three resistance genes against aphids are cloned and one against a lepidopteran (table 1.7.1). Interestingly, most resistance characterised in rice were found to have a CC-NBS-NBS-LRR structure (two NBS domains instead of one). NBS or NB-ARC domains are described as regulators ("molecular switch") for activation of R-genes (Takken and Goverse, 2012, Takken et al., 2006, Tameling et al., 2006).

Table 1.7.1 Cloned plant genes that give resistance to insects.

Insect species	Gene name	type	host species	Reference
Aphids				
Macroshiphum euphorbiae	*MI-1.2*	CC-NBS-LRR	Tomato	Nombela et al. (2003)
Aphis gossypii	*VAT*	CC-NBS-LRR	Melon	Boissot et al. (2016a)
Acyrthosiphon kondoi	*AIN*	CC-NBS-LRR	Medicago	Klingler et al. (2009)
Planthoppers				
Nilaparvata lugens	*Bph1*	CC-NBS-NBS-LRR	Rice	Zhao et al. (2016)
Nilaparvata lugens	*Bph2*	CC-NBS-NBS-LRR	Rice	Tamura et al. (2014)
Nilaparvata lugens	*Bph7*	CC-NBS-NBS-LRR	Rice	Zhao et al. (2016)
Nilaparvata lugens	*Bph9*	CC-NBS-NBS-LRR	Rice	Zhao et al. (2016)
Nilaparvata lugens	*Bph10*	CC-NBS-NBS-LRR	Rice	Zhao et al. (2016)
Nilaparvata lugens	*Bph14*	CC-NBS-LRR	Rice	Du et al. (2009)
Nilaparvata lugens	*Bph18*	CC-NBS-NBS-LRR	Rice	Ji et al. (2016)
Nilaparvata lugens	*Bph21*	CC-NBS-NBS-LRR	Rice	Zhao et al. (2016)
Nilaparvata lugens	*Bp26*	CC-NBS-LRR	Rice	Tamura et al. (2014)
Nilaparvata lugens	*Bph15*	LEC-RLKs	Rice	Cheng et al. (2013)
Nilaparvata lugens	*Bph3/ Bph17*	LEC-RLKs 1-3	Rice	Liu et al. (2015)

Nilaparvata lugens	*Bph6*	Exocyst localised protein / Atypical LRR	Rice	Guo et al. (2018)
Nilaparvata lugens	*Bph29*	B3 DNA binding protein	Rice	Wang et al. (2015b)
Nilaparvata lugens	*Bph32*	SCR domain containing protein	Rice	Ren et al. (2016)
Lepidopteran				
Spodoptera exigua	*INR*	LRR-RLP	Cowpea	Steinbrenner et al. (2020)

1.7.4.2 Characterised effectors of hemipterans

Defence pathways mediated by NBS-LRRs are often triggered by effectors or their actions (Takken et al., 2006, Li et al., 2019). For plant pathogens, it has been shown that effectors (in)directly recognised by NBS-LRRs lead to severe cell defence responses, such as cell death, often referred to as hypersensitive response (HR). In this scenario, the effectors are referred to as avirulence factors. In the absence of NBS-LRRs, the 'avirulence factors' often are virulence factor/effectors that suppress cell defence responses. The interplay of (a)virulence and (NBS-LRR-mediated) plant defence is subject to evolutionary arms races leading to variations of effector genes of pathogens/pests and NBS-LRRs of plants, particularly for specialist systems in which the pathogen/pest host range is predominantly restricted to one or a few related plant species. How these arms races work for generalist pathogens that have more plant species as hosts is less well known. Nonetheless, understanding genetic variations at the whole-genome scale of pathogen and pest populations is likely to reveal the diversity of populations and genes and effectors under selection. Moreover, possibly some conserved effectors could be found that are necessary for insect colonisation on certain plant hosts.

Table 1.7.2 Table of putative effectors in Hemipteran

Insect species	Effector	Phenotype	Reference
Myzus persicae	MpC002	Increased fecundity	Pitino and Hogenhout (2013)
Myzus persicae	Mp2	Increased fecundity	Pitino and Hogenhout (2013)
Myzus persicae	Mp10	Suppress reactive oxygen species burst	Rodriguez et al. (2014b)
Myzus persicae	Mp42	Suppress plant immune signalling	Rodriguez et al. (2014b)
Myzus persicae	Mp1	Increased fecundity	Rodriguez et al. (2017)

Myzus persicae	Mp56	Reduced fecundity	Elzinga et al. (2014)
Myzus persicae	Mp57	Reduced fecundity	Elzinga et al. (2014)
Myzus persicae	Mp58	Reduced fecundity	Elzinga et al. (2014)
Myzus persicae	Armet	Activation salicylic acid signalling in plants	Cui et al. (2019)
Acyrthosihon pisum	Armet	Increased fecundity	Wang et al. (2015a)
Acyrthosihon pisum	ApC002	Increased fecundity	Pitino and Hogenhout (2013), Mutti et al. (2005)
Acyrthosihon pisum	Ap25	Increased fecundity	Guy et al. (2016)
Macrosiphum euphorbiae	Me10	Increased fecundity	Atamian et al. (2013)
Macrosiphum euphorbiae	Me23	Increased fecundity	Atamian et al. (2013)
Macrosiphum euphorbiae	Me47	Increased fecundity	Kettles and Kaloshian (2016)
Rhopalosihum padi	RpC002	Suppress plant immune signalling	Escudero-Martinez et al. (2020)
Rhopalosihum padi	Rp1	Suppress plant immune signalling	Escudero-Martinez et al. (2020)
Bemisia tabaci	laccase 1	Overcome chemical defences of host plants	Yang et al. (2017)
Bemisia tabaci	Bt56	Activation salicylic acid signalling in plants	Xu et al. (2019)
Bemisia tabaci	2G4	Reduce disease development of pathogens	Lee et al. (2018)
Bemisia tabaci	2G5	Reduce disease development of pathogens	Lee et al. (2018)
Bemisia tabaci	6A10	Reduce disease development of pathogens	Lee et al. (2018)
Bemisia tabaci	BtFer1	Suppressing H2O2 oxidative signalling in tomato	Su et al. (2019)
Bemisia tabaci	Bsp9	Suppress plant immune signalling	Wang et al. (2019)
Nilaparvata lugens	Nisef	Suppress plant immune signalling	Ye et al. (2017)

Nilaparvata lugens	NiEG1	Cellulose degradation	Ji et al. (2013)
Nilaparvata lugens	Ni12	Activate defence responses	Rao et al. (2019)
Nilaparvata lugens	Ni16	Activate defence responses	Rao et al. (2019)
Nilaparvata lugens	Ni28	Activate defence responses	Rao et al. (2019)
Nilaparvata lugens	Ni43	Activate defence responses	Rao et al. (2019)
Nilaparvata lugens	Ni40	Activate defence responses	Rao et al. (2019)
Nilaparvata lugens	Ni32	Activate defence responses	Rao et al. (2019)
Nilaparvata lugens	NiMLP	Activate defence responses	Shangguan et al. (2018)
Nephotettix cincticeps	NcSP75	support phloem ingestion	Matsumoto and Hattori (2018)

M. persicae has a plethora of effector genes that they deploy for plant colonisation (Bos et al., 2010). The diversity of many putative effector genes within the species of *M. persicae* is unknown. Some more conserved effector genes such as Mp1, Mp10 and MpC002 are shown to be exploited for the colonisation of plants. Mp1 is interacting with the plant hosts' Vacuolar Protein Sorting Associated Protein 52 (*VPS52*). By reducing the VPS52 levels, fecundity was improved (Rodriguez et al., 2017). Mp10 is found to suppress the reactive oxygen species (ROS) burst that results in chlorosis and local cell death in response to flg22 when overexpressed in *N. benthamiana* (Bos et al., 2010, Rodriguez et al., 2014a). The effector MpC002 was found to improve *M. persicae* colonisation (Pitino and Hogenhout, 2013, Bos et al., 2010). Moreover, *M. persicae* had reduced fecundity when MpC002 was silenced using RNAi (Pitino and Hogenhout, 2013). The homolog in the *A. pisum* does also interfere with the ability of this aphid to colonise plant hosts (Mutti et al., 2008). However, so far, only a few effectors have been investigated.

Stable resistance of crops is achieved when the resistance gene(s) hold up against all genotypes (Dogimont et al., 2010). Multiple approaches are suggested in the literature to achieve stable long-term resistance against pathogens or pests, as follows; **1.** It is stacking genes that give resistance to multiple pest biotypes/clones (Halpin, 2005). **2.** Prevent the ability of an insect to colonise plants by mutating genes (susceptibility-genes (S-genes)) for plant components required for insect attraction, feeding and colonisation (Pavan et

al., 2010). **3.** Targeting insect pest 'Achilles' heels', such as conserved effectors, by plant proteases or other activities (Ahman et al., 2019). To fully exploit these approaches, it has to be assessed whether there is variation in (i) resistance to *M. persicae* in sugar beet germplasm and (ii) among *M. persicae* biotypes/clones/genotypes in the ability to colonise sugar beet varieties/lines.

1.8 Focus and aims described in this thesis

Aphids caused a 20-30% yield loss of sugar beet in Europe and the United Kingdom many years ago before neonicotinoids were introduced. Recently legislation of the European Union has banned neonicotinoids. Therefore, sugar beet growers are in great danger again of suffering significant yield losses due to aphid infestations in Europe. The sugar beet breeding company SESVanderHave is interested in breeding for stable resistance in sugar beet to aphids such as *M. persicae*. So far, little research has been done on aphid resistance breeding in sugar beet, and general plant resistance mechanisms to aphids are poorly understood. Also, there more knowledge needed about the diversity within aphid populations, even of single species. For example, some *M. persicae* clones can colonise *Nicotiana* species, and others do not.

SESVanderHave has access to a private sugar beet germplasm seed bank and has the sequences available of many of these lines. Before the start of this PhD project, SESVanderHave did an aphid resistance screen of some of its sugar beet germplasm with one aphid colony (likely derived from one clone) and identified some variation in resistance to *M. persicae* among lines and varieties.

The sugar beet germplasm's resistance screen outcomes to *M. persicae* are used as preliminary data in **chapter 3**. In this chapter, I tested sugar beet varieties/lines again, but now with more *M. persicae* genotypes to determine whether (**i**) previous findings of the resistance variation could be confirmed and (**ii**) if there is variation among *M. persicae* genotypes for colonisation of the sugar beet lines/varieties. Furthermore, I tested additional SESVanderHave germplasm for resistance to aphids and BMYV. In chapter 3 it became apparent that there are differences among *M. persicae* clones reared in the JIC insectary in their ability to colonise sugar beet. Based on these results, it was decided to genotype these clones and to investigate genetic

variation of *M. persicae* on sugar beet fields and neighbouring plants in UK and Europe.

Microsatellite markers to genotype *M. persicae* clones in Europe were described (refs). In **chapter 4**, I used the published markers to genotype the JIC *M. persicae* clones. However, this was not as straightforward as anticipated because the primers for microsatellite typing amplified multiple aspecific bands, which made the distinction of the real markers very difficult. Attempts to further optimise the experimental conditions did not improve the situation. Furthermore, I did not find evidence of specific amplification of the microsatellites in the existing literature. In conclusion, these results revealed that more work has to be done to properly genotype *M. persicae* clones.

Given the advances made in new generation sequence technologies, it was decided to develop markers based on whole-genome sequence information of *M. persicae*. A draft genome of *M. persicae* clones O and G006 was already available (Mathers et al., 2017) at the time this work started. However, the genome assemblies were fragmented and required further optimization. Therefore, in **chapter 5**, I contributed to this effort by optimizing high molecular weight DNA extractions because long DNA fragments were required for long-read PacBio and Oxford Nanopore technology Minion sequencing. I was successful in obtaining long reads via Minion sequencing, and these were used to improve the assembly of the M. persicae genome dramatically. A high-quality chromosome-level assembly of *M. persicae* clone O was obtained. This assembly was the perfect basis for conducting population genomics analyses.

The next step was to sample *M. persicae* populations from the field. Therefore, in **chapter 6**, I worked with colleagues from SESVanderHave to setup 20 m^2 capture plots in many countries of Europe, including the UK. The capture crops were similarly organized, containing 5 m^2 areas with only sugar beet, only canola, potato and Chinese cabbage. These were sampled for three consecutive years, and aphid samples on the plants in the plots were sent off to me. In addition to aphids from the field plots, I also received *M. persicae* samples from other fields and lab-grown colonies from other countries. In order to optimize aphid shipment, I developed a silica-tube storage system that enabled a cheap way of shipment of single aphid samples while maximizing DNA yield. I also optimized DNA extraction procedures from single (dried) aphids. Many aphid samples were received, but it became apparent that not all aphids were *M. persicae*. Therefore, I genotyped a subset of the

received samples via mitochondrial cytochrome oxidase I (COI) sequencing. Approximately 16 aphid species were identified. Of these, 129 samples genotyped as *M. persicae* were selected for whole-genome resequencing at > 10-20 fold genome coverage.

In **chapter 7**, I further analysed the genome sequences derived from the 129 samples of chapter 6. I found that the reads of 99 samples were appropriately aligned to the *M. persicae* clone O genome sequence (chapter 5). With the alignments of the reads of 99 genomes to the chromosome-level genome of clone O, I identified single nucleotide polymorphisms (SNPs) throughout the genome. The SNPs were, in turn, used to conduct an assessment of field *M. persicae* population structures. The resulting data were used to select clusters of aphids that shared SNPs and analysed if the SNPs may have been part of selective sweeps. Various population genetics analyses methods identified 11 sweeps that were further investigated for the underlying genes involved. One of the sweeps associated with *M. persicae* plant host preference included candidate effector genes.

In conclusion, research conducted in this PhD project showed that there is a level of resistance in sugar beet germplasm to some *M. persicae* clones. In addition, I found that field populations of *M. persicae* that colonize sugar beet are highly diverse. In **chapter 8**, I discussed how these findings might be used towards obtaining stable sugar beet resistance.

Chapter 2. Material and methods

2.1 *Myzus persicae* rearing and maintaining of *Beet mild yellowing virus*

2.1.1 Insect rearing and plant growth/maintenance conditions

M. persicae colonies of genotypes; O, 4106a, US1L, UK_SB and FRC (table 2.1.1), were maintained and reared in a 25x50x50cm custom-build acrylic cage with ultra-fine insect mesh (thrips mesh) on the two narrowest sides for ventilation. The aphid colonies were reared on Chinese cabbage (*Brassica rapa*) in a long day photoperiod (16/8-h light/dark) at 18°C. Chinese cabbages used for insect rearing were sown on insecticide-free F2 compost soil (Levington, UK) and grown at 20°C, long-day photoperiod (14-h day at 18°C and 10-h night at 15°C photoperiods with 60% relative humidity) for at least four weeks before use.

Table 2.1.1 *M. persicae* genotypes reared in the Insectary at JIC.

Genotype	Location isolated	Original host	Year population started at JIC	Originally reared on
O	Scotland	Potato	2008	Chinese cabbage
US1L	South-East England	Sugar beet	2009	Chinese cabbage
4106a	Scotland	Potato	2016	Chinese cabbage
UK_SB	England	Sugar beet	2016	Sugar beet
FRC	Southern France	Peach	2015	Chinese cabbage

M. persicae is a generalist and can colonise divergent plant species (Mathers et al., 2017, Chen et al., 2020), although it is not known if all *M. persicae* clones can survive on all hosts. At the time this study started, the JIC entomology facility reared four *M. persicae* clones that were maintained on *Brassica rapa* and sugar beet (*Beta vulgaris*) for some time. This included *M. persicae* clones O, US1L, UK_SB and 4106A. Clone O was genotyped as clone O using microsatellite markers (Fenton et al., 2010), and the genome of this aphid clone was sequenced to completion by the Hogenhout group in John Innes Centre (Mathers et al., 2017). Clone O has been reared on *B. rapa* since 2007. Clone US1L was collected from sugar beet in South-East England and reared on *B. rapa* since 2007. Clone UK_SB was originally collected from sugar beet and reared on sugar beet since 2016. Clone 4106a is also known as Braveheart and was collected from potato and is cited in several studies (Bass et al., 2014, Voudouris et al., 2017, Mingeot et al., 2021) as an insecticide susceptible clone.

2.1.2 Maintaining stock cages of *Beet mild yellowing virus*-infected sugar beet plants

Sugar beet (*Beta vulgaris*) plants infected with BMYV were obtained from the British Beet Research Organisation (BBRO), Norwich, UK and verified using a multiplex RT-PCR. The multiplex analysis was needed to ensure that we only had one type of virus within the sugar beet plants. The RT-PCR was conducted for the detection of open reading frame 3 (ORF3) for several Polerovirus (such as Beet mild yellowing polerovirus (BMYV), Beet chlorosis polerovirus (BChV), Turnip yellows polerovirus (TuYV)), the ORF0 of BMYV, the ORF0 of BChV, ORF0 of TuYV and targeting the coat protein of Beet yellows virus closterovirus (BYV).

M. persicae genotype UK_SB was reared on the BMYV infected sugar beet plants and new 4-week old non-infected sugar beet plants were added into the stock cage on a 2/4 weekly basis, depending on the needed colony size. Plants and insects have been maintained in 25x50x50cm custom-build acrylic cage with thrips mesh on the two narrowest sides for ventilation. The colonies were located in a controlled environment that was conditioned at 14-h day at 18°C and 10-h night at 15°C photoperiods with 60% relative humidity.

2.2 Myzus persicae survival, fecundity and BMYV transmission assays

2.2.1 Preparation of host plants and aged *Myzus persicae* for whole-plant bioassays

All sugar beet lines used in the fecundity assay were received from SESVanderHave. The sugar beets lines are sown per 30 in a 10cm square black pots and grown in a 25x50x50cm custom-build acrylic cage in a climate chamber under 14-h day (18°C) and 10-h night (15°C) photoperiod with 60% relative humidity. Chinese cabbage was grown in 12cm Ø black pots in a greenhouse under 14-h day and 10-h night at 18°C. Single 14-day old plants were separated and transplanted to a 12cm Ø black pot, and either caged with an 11 Ø custom made bottle cage (experiment 1, 2, 3)or put in a bigger cage together (experiment 4). Five different *M. persicae* fecundity and virus transmission assays were performed and described in this thesis (Table 2.2.1).

In three fecundity experiments, we used age synchronised aphids (experiment 1, 2 and 3), and in two experiments (4 and 5), we start with adult aphids. Seventeen-day-old plants were inoculated with aphids for age synchronisation. Aphids were age synchronised inoculating three adult aphids to the test plants. Two days post-inoculation (DPI) all adults were removed to keep 0-2 days old nymphs on the plant. Five DPI the nymphs were counted, and most of the nymphs were removed to leave five nymphs on the plant that were used as the experimental insects.

Table 2.2.1 Beta vulgaris lines used in fecundity and virus transmission assays. Assay type shows either fecundity assays or virus transmission. Multiple different *M. persicae* genotypes were used over the five assays O, 4106a, UK_SB and US1L.

Line	Experiment	Assay-type	*M. persicae* genotypes tested
SUS	2, 3	Fecundity	O, 4106a, UK_SB, US1L
SRDH	2, 3	Fecundity	O, 4106a, UK_SB, US1L
SR	2, 3	Fecundity	O, 4106a, UK_SB, US1L
MR	2, 3	Fecundity	O, 4106a, UK_SB, US1L
MRDH	1, 2, 3, 4, 5	Fecundity/virus transm.	O, 4106a, UK_SB, US1L
HR	2, 3, 4, 5	Fecundity/virus transm.	O, 4106a, UK_SB, US1L
R1	4	Fecundity	4106a, UK_SB
R2	4	Fecundity	4106a, UK_SB
T1	4	Fecundity	4106a, UK_SB
T2	4	Fecundity	4106a, UK_SB
V1	4	Fecundity	4106a, UK_SB
V2	4	Fecundity	4106a, UK_SB
W1	4	Fecundity	4106a, UK_SB
W2	4	Fecundity	4106a, UK_SB
X1	4	Fecundity	4106a, UK_SB
X2	4	Fecundity	4106a, UK_SB
Y1	4	Fecundity	4106a, UK_SB
Y2	4	Fecundity	4106a, UK_SB
Z1	4	Fecundity	4106a, UK_SB
Z1	4	Fecundity	4106a, UK_SB
SV-JIC 1	5	virus transmission	UK_SB
SV-JIC 2	5	virus transmission	UK_SB
SV-JIC 3	5	virus transmission	UK_SB
SV-JIC 4	5	virus transmission	UK_SB
SV-JIC 5	5	virus transmission	UK_SB
SV-JIC 6	5	virus transmission	UK_SB
SV-JIC 7	5	virus transmission	UK_SB
SV-JIC 8	5	virus transmission	UK_SB

2.2.2 Fecundity assays (experiment 1, 2, 3 and 4)

Five different *M. persicae* fecundity and virus transmission assays were performed and described in thesis chapter 3. These five different assays had slight changes in the protocol, as indicated below:

(**1**) The first experiment was a fecundity assay to compare the fecundity of *M. persicae* genotypes on *B. vulgaris* and *B. rapa*. Five technical replicates (plants) were used (N=5). Fourteen-day-old plants were inoculated with five age-synchronised *M. persicae* nymphs. Four *M. persicae* genotypes were used; 4106a, UK_SB, US1L and O. The offspring were counted and removed on the 8^{th}, 11^{th} and 14^{th}-day post-inoculation (DPI). The average survival rate and production of nymphs were calculated for 14 DPI.

(**2**) The second experiment was a fecundity assay to compare the fecundity of a single *M. persicae* genotype on different *B. vulgaris* lines. A single *M. persicae* genotype was used; US1L. Fourteen-day-old plants were inoculated with ten age-synchronised *M. persicae* nymphs. The offspring were counted and removed on the 8^{th}, 11^{th} and 14^{th}-DPI. The average survival rate and production of nymphs were calculated for 14 DPI. Six technical replicates were used for three biological replicates (N=18) as independent time points.

(**3**) The third experiment was a fecundity assay to compare the fecundity of four *M. persicae* genotypes on different *B. vulgaris* lines. Four *M. persicae* genotypes were used; 4106a, UK_SB, US1L and O. Fourteen-day-old plants were inoculated with five age-synchronised *M. persicae* nymphs. The offspring were counted and removed on the 8^{th}, 11^{th} and 14^{th}-DPI. The average survival rate and production of nymphs were calculated for 14 DPI. Five technical replicates were used for three biological replicates (N=15) as independent time points.

(**4**) The fourth experiment was a fecundity assay to compare the fecundity of two *M. persicae* genotypes on different *B. vulgaris* lines. Two *M. persicae* genotypes were used; 4106a and UK_SB. Fourteen-day-old plants were inoculated with five non-synchronised adult *M. persicae* nymphs. Both seven and fourteen DPI surviving aphids found per plant were counted. Three technical replicates were used for three biological replicates (N=9) as independent time points.

2.2.3 Sugar beet BMYV transmission assay (experiment 5)

(**5**) The fifth experiment was a BMYV transmission assay to compare virus transmission of a single *M. persicae* genotype on ten different *B. vulgaris* lines. A single *M. persicae* genotype; UK_SB. Five BMYV infected adult aphids were transferred on 16-day old sugar beet plants. 7 DPI sugar beet aphids were removed, and plants were treated with 4ml/L Pyrethrum mix insecticide spray for two times before transported from a CER insectary to a greenhouse under 14-h day and 10-h night at 18°C conditions. 28 DPI four leaves were harvested (leaves 1, 2, 3 & 4, counting out-inwards)and snap-frozen for storage at -80°C. Two leaves of each 44-day old plant were shipped and analysed with an ELISA analysis at SESVanderHave Belgium. The ELISA was performed using Turnip Yellow virus (TuYV) antibodies. Samples were incubated for 50 minutes. Variation and assay were verified by checking that the difference between leaf 1&2 was minimal and much lower than the variation between the batches.

2.2.4 Statistical analyses and graphs

All calculations were performed in R version 3.5.1 for Windows and Microsoft 365 Excel. For replications, I use 'n=#' for the biological replicates (plants used per experiment) and call repeats within a separate experiment' time point', For insect bioassays, 'survival rate' refers to the percentage of adult aphids alive at each measurement point, and 'fecundity' refers to the total number of nymphs produced.

Insect fecundity data and virus transmission data were analysed using a two-tailed paired t-test for comparative analysis between time points and an ANOVA with post-hoc Tukey-HSD test for comparison between the different *M. persicae* genotypes. All insect fecundity assays in this thesis were illustrated with boxplots and overlaying scatterplots.

2.3 Sample collection

2.3.1 Sample collection tubes for mid-term storage of DNA at room temperature, as described in (Wouters et al., 2020b).

In order to keep DNA quality and quantity high enough for population genomics studies from field-collected samples, I developed a storage system that is optimised from (Maddison, 2013). This storage system was able to keep

dried aphid samples at ambient temperature for mid-term storage for up to 8 weeks. In chapter 6, I present a method for the construction of low-cost sample collection and storage tubes from commonly available laboratory consumables (Wouters et al., 2020b) (figure 2.1). These collection and storage tubes were used in all collection attempts of aphid samples shown in this thesis.

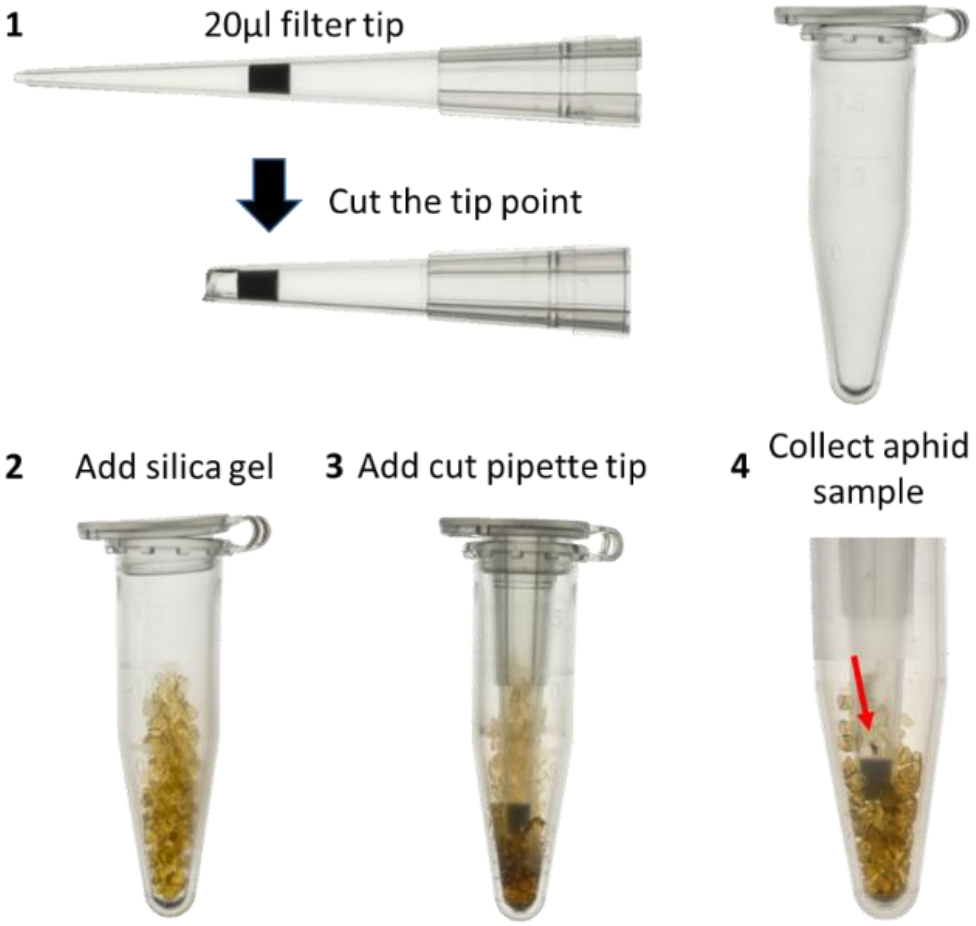


Figure 2.3.1 Development protocol for collection and storage tubes. (1). A 20µl filter pipette point was cut beneath the filter. **(2).** 100µl silica gel was added into a 1.5ml Eppendorf tube. **(3).** A cut pipette tip is pushed in the collection tube with orange silica gel on the bottom. **(4).** Collection of aphids within the pipette point to avoid crushing of the aphid during transport. The red arrow points to an aphid sample within the silica gel tube.

2.3.2 Sample collection tubes storage experiment

The effectivity of the storage tubes was tested via an 8-week storage experiment. Single adult aphids were stored for 1, 2, 3, 4, 6 and 8 weeks in, above-mentioned collection tubes, either with or without silica gel. DNA was

extracted as explained in section 2.4.1 (CTAB DNA extraction for single aphids) and in Mugford et al. (2020). DNA quality and quantity were measured with the Nanodrop (Thermofisher Scientific) and the Qubit DNA broad range kit (Invitrogen) using 2µl of eluted samples per quantification.

2.4 Sample preparation for sequencing

2.4.1 Aphid CTAB DNA extraction for single aphids

DNA from whole infected tissues was extracted using a CTAB DNA extraction protocol from Marzachì et al. (1999)and was further optimised by Sam Mugford for aphid and leafhopper samples (Mugford et al., 2020). Each extraction was carried out with a single adult *M. persicae* aphid in 1.5 mL Eppendorf tubes, each before freezing in liquid nitrogen. We lysed the tissue by grinding the aphid with a plastic pestle precooled in liquid nitrogen. Ground tissue kept fully frozen until suspended in 100 µL CTAB lysis buffer (2% w/v CTAB, 1.4 M NaCl, 20mM EDTA, 100 mM Tris-HCl pH 8.0, Milli-Q water and autoclaved to inactivated DNases. Two µL (0.2%) RNase A (100 mg/mL) (Qiagen) was added to the mixture before starting the extraction, followed by incubation on room temperature for 60 minutes. The supernatant is extracted from the aphid material by centrifugation at 13300g for five minutes at room temperature. After the supernatant was transferred to a new tube 1 volume, chloroform was added and mixed for 10 minutes. The phases were separated by centrifugation at 13300g for 5 minutes at room temperature. We transferred the upper DNA containing phase to a fresh tube, and 1 volume of ice-cold isopropanol was added. The solutions were mixed and stored for 30 minutes in a -20 freezer with occasional mixing, followed by centrifugation at 13,300g for 15 min at four °C. The isopropanol was discarded without disturbing the pellet and washed with 1ml 70% ethanol. All the ethanol was removed before adding 25 µL MiliQ water to resuspend the pellet. The solubilised DNA was stored at -20°C.

2.4.2 Aphid DNA extraction for long-read sequencing platforms

Around 20 aphids were collected alive in 1.5 ml low-bind Eppendorf tubes and snap-frozen in liquid nitrogen. I extracted high molecular weight DNA with the Illustra Nucleon PhytoPure kit (GE Healthcare, RPN8511) following the manufacturer's protocol. All centrifuge steps were performed on the lowest manufacturers recommended setting. 200 µL wide bore filtered pipette tips (Thermofisher) were used when transferring solutions to

circumvent shearing of DNA. DNA was precipitated in 50 µL MiliQ water to resuspend the pellet.

2.4.3 DNA quantitating and quality checks

2.4.3.1 Samples for microsatellites and short read whole genome sequencing

DNA quantification check was performed using 1 µl of the aliquoted DNA sample with both the Nanodrop and the Qubit (DNA broad range). No qualification assay with the 260/280 and the 260/230 ratio on the Nanodrop was performed because the yield was too low to see accurate absorbance ratios. Lower 260 values will skew the ratios of other contaminants in low abundance. The concentration (ng/µl) was calculated in both Nanodrop and Qubit. Samples with a total yield of more than 20 ng/µl on the Qubit were further used for either the microsatellites or Illumina whole-genome sequences.

2.4.3.2 Samples for whole genome long-read sequencing

DNA quantification and a quality check were performed as described in (Wouters et al., 2020a). The amount of DNA was quantified by using 2 µl of the aliquoted DNA sample with both the Nanodrop and the Qubit (DNA broad range). Qualification analysis was measured via the 260/280 and the 260/230 ratio on the Nanodrop. These ratios should be between a specific absorbance range. The absorbance ratio of 260/280 should be between 1.8 and 2 to indicate suitable quality DNA without protein or RNA contamination. The absorbance ratio of 260/230 should be between 2 and 2.2 to indicate limit contamination of organic compounds or chaotropic salts were present. Additionally, the DNA concentration (ng/µl) ratio of the Nanodrop/Qubit should be around 1.5 or below to indicate less contamination (Schalamun et al., 2018). Fragment sizes of the genomic DNA were checked via either the pulse-field gel electrophoresis or via the Femto fragment analyser-pulse (Agilent). When all quality checks passed, the samples were used for whole-genome sequences.

2.4.4 Confirmation of aphid species using the cytochrome oxidase subunit 1 (COI) gene

In order to select samples for whole-genome Illumina sequencing, I first needed to verify the aphid species and if parasitoid wasps predated the aphid. Dried aphid samples collected around the world were not identified by phenotype but only via sequencing of a conserved gene. DNA was extracted from single aphids following the DNA extraction protocol and quality check, as stated in 2.4.1 and 2.4.3.1.

2.4.4.1 Amplification and purification of the COI gene

A 710-bp conserved mitochondrial gene fragment, COI, was amplified using the following primers; LCO1490 (GGTCAACAAATCATAAAGATATTGG) and the complement sequence HCO2198 (TAAACTTCAGGCTGACCAAAAAATCA) (Folmer et al., 1994). Phusion DNA polymerase was used for amplification of the COI fragment using manufacturer protocol for 20µL amplification reactions. The PCR-programme was followed as advised in Lee et al. (2011); Commenced with 5 minutes at 94°C, followed by 34 cycles of three times 1-minute steps of 94°C, 45°C and 72°C. The programme was finished with a final extension step at 72°C for five minutes. These PCR products were verified and purified from an electrophoresis 0.8% agarose gel using a QIAquick® PCR purification kit (QIAGEN Inc).

2.4.4.2 Sanger sequencing and data evaluation

Purified PCR-products were sent for Sanger sequencing. Samples were sequenced in both directions using the primers defined in 2.4.4.1. Consensus sequences were trimmed and aligned using CLC workbench V8. Alignments were plotted on a neighbour-joining phylogenetic tree and then exported to the web-based programme Interactive Tree Of Life (https://itol.embl.de/). Furthermore, consensus sequences were blasted for sequence similarities on blastN optimised for highly similar sequences (https://blast.ncbi.nlm.nih.gov/). The alignments were verified checked on 100% identity (up to 1% error margin allowed) and around 99/100% query cover.

2.4.5 ONT MinION library preparation

High molecular-weight genomic DNA libraries were prepared using the Ligation Sequencing Kit SQK-LSK109 according to the manufacturing protocol. **ONT LSK109 Protocol:** The DNA was equipped with a DNA repair and end-prep step. 47µl with in total 2µg DNA(diluted with nuclease-free water) was transferred in a 0.2ml thin-walled PCR tube and mixed with 3.5µl NEBNext

FFPE DNA repair buffer, 2µl FFPE DNA repair mix (NEB), 3.5µl Ultra II End-prep reaction buffer (NEB) and 3µl Ultra II End-prep enzyme mix (NEB). The solution was gently mixed by inversion before incubated in a thermal cycler for 5 minutes at 20°C followed by 5 minutes at 65°C. The samples were cleaned with AMPure XP beads and 70% ethanol. The DNA was dissolved in 61 µl of nuclease-free water and transferred in a fresh 1.5ml Eppendorf LoBind tube with Wide-borer pipette tips. For adapter ligation, 60 µl of eluted DNA was added with 25 µl ligation buffer (ONT LSK109 manufacturer-provided), 10 µl NEBNext Quick T4 DNA ligase (NEB), 5 µl Adapter mix (manufacturer-provided) and gently mixed by inversion before incubation of 10 minutes on room temperature. The samples were cleaned with AMPure XP beads and 70% ethanol. The DNA was dissolved in 20 µl dilution buffer (manufacturer-provided) and transferred in a fresh 1.5ml Eppendorf LoBind tube with Wide-borer pipette tips. 1µl of the eluted DNA sample was used to measure the concentration on the Qubit fluorometer.

2.5 Sequence techniques for resequencing and long-read sequencing

2.5.1 Illumina sequencing for *Myzus persicae* resequencing

Genomic DNA was extracted from single aphids described in 2.4.1, and quality was assessed as described in 2.4.3.1. Samples verified as *M. persicae* without contamination with the COI sequence reads were sequenced with PCR-based 150bp low-input Paired-end libraries by Novogene with the NEBNext® DNA Library Prep Kit, according to manufacturer's protocol. PCR-based 150bp low-input Paired-end libraries were sequenced with Illumina on a depth of 20x coverage with 80% of the read of at least a quality Phred score of 30 (error base rate maximal 1/1000).

2.5.2 Nanopore long-read sequencing on the ONT MinION

A new 106 REV-D flow cell (9.4) was prepared and checked for active pores, and primed according to manufacturing protocol (SQK-LSK108 or SQK-LSK109). 20 µl DNA library was prepared with 37.5 µl sequencing buffer and 25.5 µl loading beads before added into the flow cell for sequencing on the ONT Minion. The sequencing programme was run as default for maximal 72 hours on default settings for the SQK-LSK108/SQK-LSK109 without base-

calling and was started before the library was loaded on the flow cell. The raw reads were processed with the base-calling programme Guppy V2.3.1.

2.5.3 Previous published whole-genome sequences (Singh et al., 2020a).

Twelve WGS samples were obtained from previously published works (Singh et al., 2020a). These twelve samples were initially collected from Germany (SUS_NS), Italy (NIC_8124, NIC_67, SUS_23, SUS_4255A,), Zimbabwe (NIC410R, NIC_5191A), Greece (NIC_926B, NIC_410G), the United Kingdom (SUS_US1L, SUS_4106a, SUS_1X). All *M. persicae* genotypes were collected from the field and reared in clonal populations on Chinese cabbage for at least a year before sequencing. Genomic sequences were obtained on the Illumina HiSeq 2500 with PCR-free libraries. SUS_NS and SUS_1X were sequenced using a 100-bp paired-end library, and the other ten samples were sequenced using a 250-bp paired-end library. The sequenced samples of depths between 38x and 200x were subsampled down to around 20x coverage.

2.6 Bioinformatics analyses

2.6.1 Subsampling, trimming of the reads and evaluation of sequencing quality

All the commands, workflows, algorithms and programmes used in this thesis have been used on default settings unless stated differently.

Short-read Illumina sequenced WGS samples were received from different sources and had different sequence depths from 15x up to 250x coverage. Re-sequenced samples from fields were sequenced using a PCR-based low-input library for around 20x coverage. While samples from Singh et al. (2020a) and the *M. persicae* samples reared at the John Innes Centre were sequenced PCR free on coverage between 15x and 250x. Therefore, all sequences were subsampled in the partition according to the initially estimated depth to have around 15x coverage using the command seqkit from the programme SeqKit version 0.10.0 (Shen et al., 2016) subsampling. Subsample selection seed number 1234 was used for all samples.

To remove the adapters and poor-quality sequences (below a quality score of 20), I needed to trim the reads. I trimmed the sequences, and

subsequently, they were checked for sequencing quality, GC content and yield using the FastQC command from the programme Trim galore version 0.4.2 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore).

2.6.2 *De-novo* assembly for the examination of genomic contamination

In order to analyse that we have sequenced *M. persicae* genomes alone without contamination, I used the De-novo Blobtools workflow (Laetsch et al., 2017) to visualise the quality and the taxonomic partitioning of the whole genome sequence. The programme illustrates the depth, GC content and is coloured on the genus or family level. Therefore, this workflow illustrates an overview if there are multiple aphids mixed sequence if there was fungal, bacterial, or parasitoid wasp contamination.

2.6.3 Mapping of sequencing to the reference genome

The genome reference genome of *M. persicae* genotype O was used (Mathers et al., 2020b). This optimised genome sequence with 97% mapped to the first six super-scaffolds indicates that most of the reads are assembled on six, the previously shown (Spence et al., 1998, Mandrioli and Manicardi, 2012), the number of chromosomes. The re-sequenced samples to the reference genome, I indexed the reference genome using Burrow Wheelers Aligner (BWA)-index command; BWA V version 0.7.15 (Li and Durbin, 2009). Following, by the mapping of the trimmed sequence reads to the reference genome with the BWA version 0.7.15 (Li and Durbin, 2009) BWA-mem. These mapped reads were sorted on scaffolds level with SAMtools Version 1.3 (Li et al., 2009). Additional sorting on the genetic coordinate level was done by the sortSAM command of Picardtools Version 2.1.1. (http://broadinstitute.github.io/picard/).

To ensure the sequence read duplicates were marked and removed before further analysis, I marked the duplicates using the command MarkDups from the programme Picardtools Version 2.1.1. I then indexed these files using the command Index with SAMtools version 1.3. After the duplicates are marked and the files are indexed, the sequences were realigned to the reference genome using the command RealignerTargetCreator, from the programme GATK version 3.7.0 (Poplin et al., 2018), to make an interval file and that was used for the realignment of the indels using the command IndelRealigner.

2.6.4 Development of variance file and filtering

The realigned files of the *M. persicae* WGS samples that were mapped to the reference genome were combined within a core variance file using the command Mpileup in the programme BCFtools version 1.8 (Li, 2011). Subsequently, in order to remove erroneous variants, I filtered the variance on quality scores (>Q30), biallelic, without indels, and total depth per sample per SNP (>4 depth) with the programme VCFtools Version 1.1.15. Furthermore, I removed samples within the VCF file with the rmv-indv command within the VCFtools programme.

The development of the variance file for the mitochondrial genome analyses used the same workflow as discussed above in the previous sections, but the reference where the samples were mapped to was different. The reference genome was the mitochondrial genome from *M. persicae*, obtained from GenBank (NC_029727.1). All variances were only filtered on quality score (>Q30) and total depth per sample per SNP (>4 depth).

2.6.5 Quality control of genomic samples

In order to verify the quality of the samples within the variance files, I analysed the percentage missing, the heterozygosity, the number of unique variances within each sample, checked variability within a conserved fragment of mitochondrial sequence and analysed the similarity of each sample.

The percentage missing and the heterozygosity per sample was identified by using the commands “missing-indv” and “het” within the software VCFtools version 1.1.15. The unique variation and unique homozygous sites were analysed using the programme VCFtools version 1.15 and a python script developed by SESVanderHave.

In order to verify the variances found within the *M. persicae* populations, I checked the variances within a 1.5 kb mitochondrial fragment and a similarity matrix of all variances. The analysis of the 1.5 kb mitochondrial fragment was performed by making a minimum spanning network using POPART version 1.7 for windows (Bandelt et al., 1999). The summary statistics was obtained using the same programme. Furthermore, to develop the similarity matrix I used an established Perl script written by C. Burbridge and published on github

(https://gist.github.com/cbe453/b3a28923db7b37435d4af682da260c7c). Subsequently, the matrix was plotted using the R packages reshape2 and ggplot2.

2.6.6 Variance analysis of re-sequenced *Myzus persicae* genomic samples

To analyse the principal component analysis of the 99 WGS *M. persicae* samples, I used the R packages gdsfmt, SNPRelate and ggrepel. For pruning, I used a linkage disequilibrium (LD) threshold of 0.2 in a sliding window of 20000 bases, and I did not remove the monoSNPs. Finally, only 148 monomorphic SNPs were discarded, and all other 2.8 million SNPs were used for the principal component analysis. Samples were colour labelled on either host or country via information given by collectors.

2.6.7 Network and LD analysis

In order to get the network analysis, the VCF file needed to be transformed to fasta using the programme PGDspider version 2.1.1.5 (Lischer and Excoffier, 2012), using Java JDK-1.7.0_25. A single haplotype was randomly selected per WGS sample with the programme seqkit version 0.9.1 (Shen et al., 2016). Subsequently, the network analysis was illustrated using the Neighbour-Net in splits tree for windows version 4.15.1 (Huson and Bryant, 2006).

The LD was calculated using the command geno-r2 in VCFTools version 0.1.15. I had calculated the LD over 10kb windows and limited the maximum distance from the SNP by 10 kb. The graphs were illustrated using the ggplot2 package from R.

2.6.8 Genetic structure analysis

To predict the number of potential clusters found within the subset, I used the analysis programme ADMIXTURE version 1.3.0 (Alexander and Lange, 2011).

In order to get the right format to read into the structure analysis, I sorted the VCF file on location and indexed the file using BCFtools version 1.8 (Li, 2011). The dataset was pruned on linkage disequilibrium within 50kb

windows using plink version 1.9 (Purcell et al., 2007). The pruned VCF file was used in ADMIXTURE for K=2 up to K= 40 and the number of clusters with the lowest cross-validation error was selected.

2.6.9 Selective sweep analysis

To obtain a selection of selective sweep areas with high probability, I calculated the composite likelihood ratio (CLR) within the 2.8 million variants per selected genomic group using the programme SweeD version 3.3.2 (Pavlidis et al., 2013). I quantified the maximum number of sites calculated per scaffold to 100000 via the grid function and folded the site frequency spectrum to not distinguish between the ancestral and the derived allele. The likelihood scores were reformatted in columns via a Python script before plotted in graphs via R using the package ggplot2.

The three other parameters, Pi, TajimaD and FST, were all three calculated using the programme VCFtools version 0.1.15. All three parameters, Pi, FST, Tajima D, used a 10 kb window and 5 kb slides. FST estimates were calculated using Weir and Cockerham's (Weir and Cockerham, 1984) formula against each of the selected groups, and the weighted FST was used. The parameters were calculated separate per group, and regions were selected subsequently using Linux command line. All four parameter graphs were plotted using the R package ggplot2.

2.6.10 Gene-specific analysis

In order to analyse the genes underlying the selective sweep areas, I obtained the gene IDs from the annotation of *M. persicae* clone O as described in (Chen et al., 2020). Functional annotation was performed to predict the function of each gene using the programme interproscan version 5.22.61. Effectors were predicted by Sam Mugford using the following valuations: **1.** Transcripts were showing high expression in salivary glands, **2a.** encoding proteins with predicted secretory signal peptides using the program SingalP version 4.0 (Petersen et al., 2011), or **2b.** that were detected in the saliva of aphids by mass spectrometry (Bos et al., 2010, Harmel et al., 2008) this led to an identification of a set of 496 candidate effectors

2.6.11 Statistical analyses and graphs

Most statistical analyses were performed using the R version 3.2.2. All graphs were produced in R (Team, 2013) unless otherwise stated. All bioinformatics analyses were performed using the high-performance computing (HPC) resources of the Norwich Bioscience Institute Partnership Computing infrastructure for the Science group.

Chapter 3. Analysis of resistance of *Beta vulgaris* varieties to *Myzus persicae* clones.

3.1 Introduction

In the last three decades, neonicotinoids have been effective at controlling *M. persicae* in the UK. However, *M. persicae* clones that are resistant to these pesticides were found in France and Greece; French Resistance clone (FRC) and 5191A (Puinean et al., 2010, Bass et al., 2011). These two clones (FRC and 5191A) use two distinct mechanisms for neonicotinoid resistance: 1. over-expression of detoxifying cytochrome P450 monooxygenases, which inhibits detoxifying enzymes (Puinean et al., 2010); 2. Mutations in nicotinic acetylcholine receptor β subunit, resulting in reduced sensitivity to imidacloprid (Bass et al., 2011). Moreover, *M. persicae* is known to have evolved resistance to ± 70 other pesticides, for example organophosphates, carbamates and pyrethroids (Silva et al., 2012, Bass et al., 2014). These insecticides include many that were commonly used in the UK before neonicotinoids were introduced, and aphids have over time developed effective resistances to these pesticides. Therefore, it follows that it is could be a matter of time before *M. persicae* also develop resistance to neonicotinoids and other insecticides in the UK. Given this situation, it is essential to identify different ways to control aphid outbreaks in sugar beet. One way is to identify sugar beet varieties that have resistance to aphids, and particularly *M. persicae*.

Partial resistance against *M. persicae* is described for some crops and trees. For example, in pepper, a reduction of phloem intake was found by *M. persicae* on a wild accession of *Capsicum baccatum*. This reduction of phloem intake by *M. persicae* was induced by callose deposition in the sieve elements (Sun et al., 2018). More recently, Sun et al. (2020b) found a quantitative trait locus (QTL) for aphid resistance in *Capsicum* species. This resistance resulted in reduced aphid survival rates and fecundity. Interestingly this QTL was fine mapped to a region with Leucine-rich repeat receptor-like kinase (LRR-RLK) analogues. RLKs are known to be involved in resistance responses against pathogens/nematodes (Fradin et al., 2009, Krol et al., 2010, Mendy et al., 2017) and in rice to the brown planthopper *Nilaparvata lugens* (Liu et al., 2015). In Arabidopsis, Kloth et al. (2017) also described a gene; small heat shock-like sieve element-lining chaperone1 (SLI1) that was found to reduce phloem intake of *M. persicae* and increase salivation during the feeding. In peach, seven QTLs were found to be linked to increased resistance via an underlying change in aphid feeding. *M. persicae* feeding on peach plants with the QTLs was salivating more in the sieve element than other peach trees (Verdugo et al., 2015, Sauge et al., 2012, Sauge et al., 2002, Lambert and

Pascal, 2011) Thus, partial resistance against *M. persicae* does exist in the germplasm of some plant species.

M. persicae genotypes and subspecies have been identified (Margaritopoulos et al., 2009, Fenton et al., 2010, Blackman, 1987). Several aphid species show within-species variations in their ability to colonise plant species. For example, Peccoud et al., (2009) found that *A. pisum* as a species consists of at least 11 separate biotypes that have evolved specialisation on different legume species. When genetic differences are found or likely to be present, the aphid clones are referred to as genotypes. When differences in biology, such as plant host preference, between aphid clones are found, these clones are referred to as biotypes. Microsatellite markers suggest that there are genotypic variations among *M. persicae* clones (Margaritopoulos et al., 2009, Fenton et al., 2010). Moreover, variation in the ability of *M. persicae* to colonise *Nicotiana benthamiana* has been identified, and there is evidence that these have diverged into the subspecies *M. persicae nicotianae* (Blackman, 1987), even though clones that belong to *M. persicae sensu strictu* have the ability to colonise divergent plants species from different families, including *N. benthamiana* (Mathers et al., 2017, Chen et al., 2020). Given these biotype and genotypic variations within *M. persicae*, the questions arise whether there is variation within *M. persicae* to colonise sugar beet and how these variations will affect the identification of sugar beet resistance to these insects.

Currently, no sugar beet resistance for *M. persicae* is identified. The sugar beet breeding company SESVanderHave attempted to identify sugar beet varieties and lines with resistance to *M. persicae*. Breeding companies generally have access to diverse germplasm that is used to cross in desirable traits into elite breeding varieties that are optimised for commercialisation. In the previous field and controlled environment fecundity assays, SESVanderHave screened part of their sugar beet germplasm resource for resistance against *M. persicae* and found evidence for the resistance of certain lines to an *M. persicae* lineage temporarily reared at their facilities. Based on these results, the company selected six sugar beet lines with variable resistance to this *M. persicae* clone. These lines were classified according to their performance as susceptible (SUS), mediocre resistant (MR), mediocre resistant double haplotype (MRDH), semi-resistant (SR), semi-resistant double haplotype (SRDH) and resistant (R). However, *M. persicae* predominantly reproduces clonally during spring and summer, and when reared in rearing facilities under controlled environmental conditions, aphids generally

reproduce via female clonal production, without male adults. Hence, the sugar beet resistance screen at SESVanderHave was probably done with a single *M. persicae* genotype. Given this, the question we need to address is whether the sugar beet resistance identified at SESVanderHave holds up to multiple other *M. persicae* clones and is thereby stable.

In addition, to understand the impact of aphid resistance to the transmission of plant viruses, sugar beet lines with aphid resistance will have to be assessed for resistance to viruses transmitted by *M. persicae* as well. In the last few decades, less effort was put into resistance breeding against viruses and their vectors, such as *M. persicae*, because both were easily controlled by treating sugar beet seed with a neonicotinoid coating. So far, it is not clear whether increased resistance to *M. persicae* is correlated with increased resistance to viruses. Given that resistance to aphids is often associated with a defence response of the phloem and that most sugar beet viruses that are transmitted by M. persicae and cause yield loss in sugar beet, i.e. BMYV, are phloem-limited viruses (See table 1.4.1 and section 1.4 of the main introduction of this thesis), it is likely that *M. persicae* and virus resistance are shared. *M. persicae* transmits over 100 plant viruses to many economically important plant crops, such as sugar beet, tomato, oilseed rape (Blackman and Eastop, 2000a). The greatest impact of *M. persicae* as an economical pest is because of its great host range and being a vector of a plethora of plant viruses (Blackman and Eastop, 2000a, Van Emden and Harrington, 2017). Therefore, it is of great interest to understand if resistance against *M. persicae* is associated with increased resistance to BMYV in sugar beet.

Here, I investigated the resistance levels of the sugar beet varieties/lines identified by SESVanderHave to four different *M. persicae* clonal populations that were reared in the JIC Entomology Facility. I found that these four clonal lineages colonise sugar beet, though with differences in fecundity and survival rates. This suggests that *M. persicae* clones vary in their ability to colonise sugar beet. I also identified variations in resistance to the *M. persicae* clones among the SESVanderHave sugar beet lines. Moreover, the resistance found to *M. persicae* in sugar beet does show similar patterns in reduced BMYV titres in virus transmission assays. The results in this chapter suggest that more information is needed about *M. persicae* genetic diversity.

3.2 Results

3.2.1 Comparative analysis of four *M. persicae* clones on plant species *Brassica rapa* and *Beta vulgaris*.

Firstly, to assess if the *M. persicae* clones O, US1L, UK_SB, 4106a survive and reproduce on *B. vulgaris*, I conducted a comparative performance analysis of them on *B. rapa* and *B. vulgaris* line MR (figure 3.2.1). Each sugar beet plant was seeded with five aphid nymphs of the same age, and the aphid colony growth was observed for two weeks. This comparative analysis showed that all aphids clones established fecund colonies two weeks after transferring to sugar beet. Overall, fewer nymphs were produced per day on *B. vulgaris* than on *B. rapa* (Student's t-test, n=24, $P<0.001$). The four clones also had significantly lower survival rates on *B. vulgaris* than on *B. rapa* (Student's t-test, n=24, $P=0.034$). However, colonies for all four *M. persicae* clones were found surviving and reproducing eight weeks post-inoculation. Interestingly, there were significant differences found in fecundity among the M. *persicae* clones on *B. rapa* (ANOVA, n=24, $P<0.001$) and *B. vulgaris* (ANOVA, n=24, $P<0.001$). The fecundity differences among the clones were explained by the lower numbers of nymphs produced by clone 4106a compared to the other three clones. This comparative analysis on *B. rapa* and *B. vulgaris* showed that all four *M. persicae* clones could be used for *B. vulgaris* fecundity assays. However, they all had lower fecundity on *B. vulgaris*.

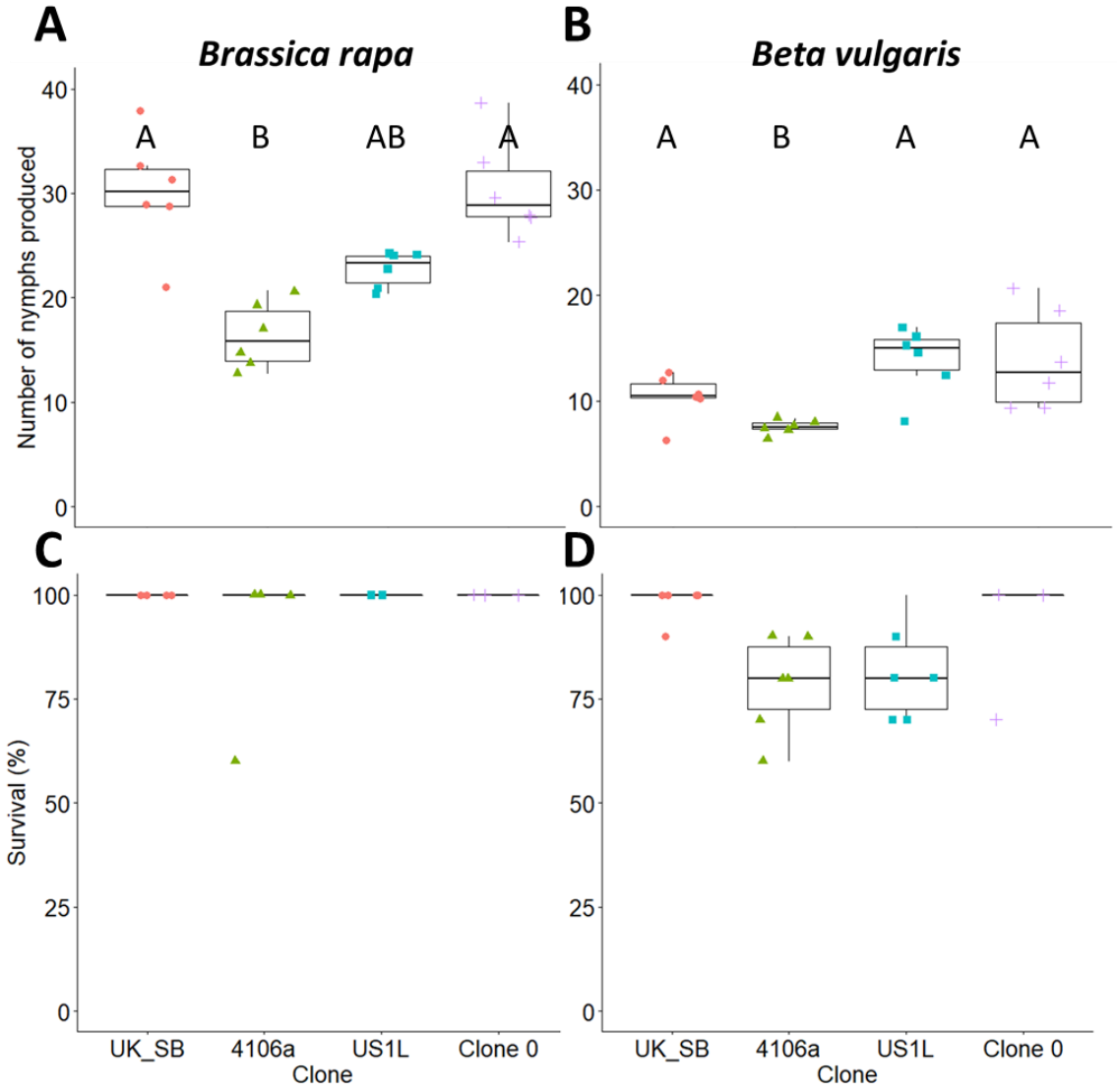


Figure 3.2.1 Comparative analysis of four *M. persicae* clonal populations on *B. rapa* and *B. vulgaris*. Survival and fecundity assays of four *M. persicae* clones (4106a, O, UK_SB, US1L). Five age-synchronised *M. persicae* nymphs were used and 14 days post-inoculation (DPI) the average survival rate and production of nymphs were calculated (N=5). Aphid fecundity on *B. rapa* **(A)** and *B. vulgaris* **(B)** and aphid survival rate on *B. rapa* **(C)** and on *B. vulgaris* **(D)**. All four *M. persicae* clones survived and reproduced on *B. vulgaris* for at least two weeks after transferring them to the sugar beet plants from *B. rapa*. Technical repeats are shown in boxplot coloured per clone. UK_SB = Red circles, 4106a = Green triangles, US1L = blue squares, Clone O = purple crosses. Letters above boxplots (A, B, AB) indicate significant differences between *M. persicae* clones a determined by ANOVA post-hoc Tukey-HSD (α=0.05).

3.2.2 Resistance levels of sugar beet lines from SESVanderHave.

In order to understand if the sugar beet germplasm classified as resistant in the test from SESVanderHave was consistent across clones, an initial test was done with *M. persicae* US1L. Six sugar beet lines were selected and provided by SESVanderHave that had variable resistance levels and had shown differences in fecundity when tested with a different *M. persicae* clone by SESVanderHave. Based on these results, these lines were labelled susceptible (SUS), semi-resistant (SR), semi-resistant double haplotype (SRDH; derived as double haplotype from SR), mediocre-resistant (MR), mediocre-resistant double haplotype (MRDH; derived as double haplotype from MR) and highly resistant (HR). These sugar beet lines are shown phenotypical structural different phenotypes. SUS plants had wide blueish leaves, SR/SRDH/MR/MRDH plants had slightly elongated stems and slightly curled leaves, and HR plants had a strong elongated stem and leaves were strong curling.

I observed significant differences (ANOVA, n=108, P<0,001) in fecundity of a single *M. persicae* clone US1L on the six sugar beet lines 14 days post-inoculation (figure 3.2.2). The US1L aphids did least well on the HR line, indicating that this line has some resistance to these aphids or that the morphological attributes of this line interfered with aphid performance. The aphids performed the best on MR and MRDH lines, and the aphids did better on these lines than on the SUS, SR and SRDH lines that were supposed to be the most susceptible to aphids according to previous data from SESVanderHave. The aphid performance tests on SR and SRDH and on MR and MRDH were similar, providing confidence in the results because the SRDH and MRDH lines were derived from SR and MR, respectively. Whereas these data are only partially consistent with those of SESVanderHave, both studies show that there are differences in aphid susceptibilities of sugar beet lines. The structural differences of the sugar beet lines may contribute to the resistance levels of the HR line to *M. persicae*.

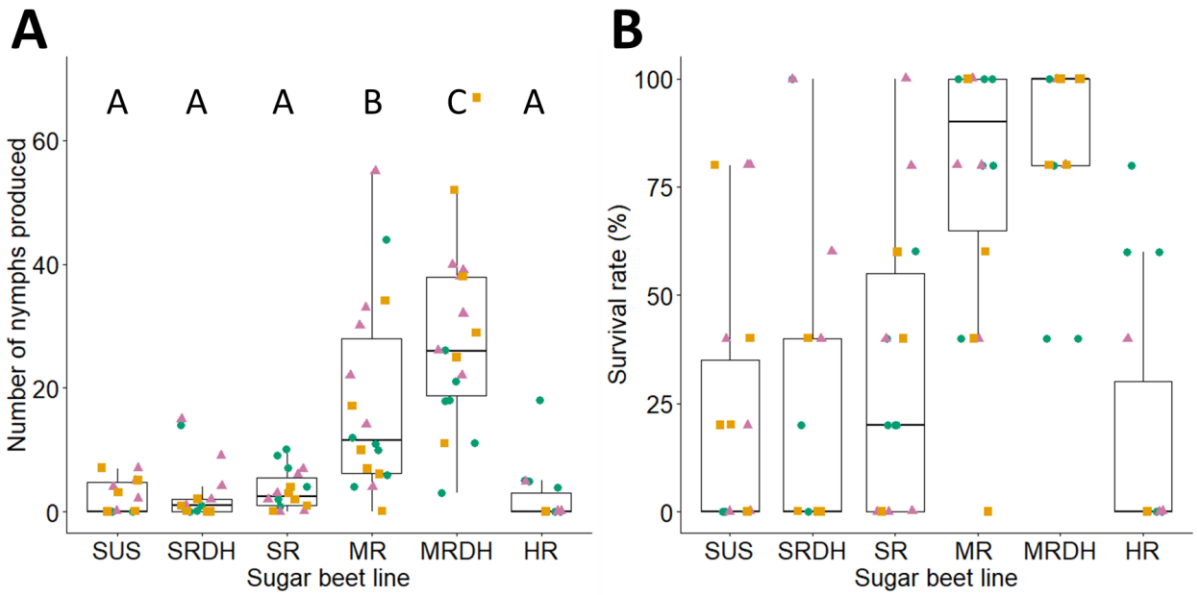


Figure 3.2.2 Fecundity assay of a single *M. persicae* clone, US1L, on six sugar beet lines show differences in fecundity levels. Six sugar beet lines with presumed variable resistance levels selected and classified by SESVanderHave were tested with a different *M. persicae* clone, US1L. **(A)** The mean number of *M. persicae* nymphs produced, and **(B)** survival rate of the age-synchronised aphids 14 days post-inoculation. In total, eighteen experimental replicates were conducted over three different biological replicates of N=6 each. Biological replicate 1 = green circles, biological replicate 2 = purple triangles and biological replicate 3 = orange squares. Letters above boxplots (A, B, C) indicate significant differences between sugar beet lines as determined by ANOVA post-hoc Tukey-HSD (α=0.05).

To test the selected six sugar beet varieties further with additional *M. persicae* clones, I tested the fecundity and survival rates of four more *M. persicae* clonal populations (4106a, O, UK_SB and US1L) on the six selected sugar beet lines (SUS, SRDH, SR, MR, MRDH, and HR) 14 days post-inoculation (figure 3.2.3). Two *M. persicae* clones did not show any differences to clone US1L in the number of nymphs produced on the six sugar beet lines. Clone 4106a did show overall a significantly lower production of nymphs on the six sugar beet lines than US1L did (Student's t-test, N=62, P =0.033).

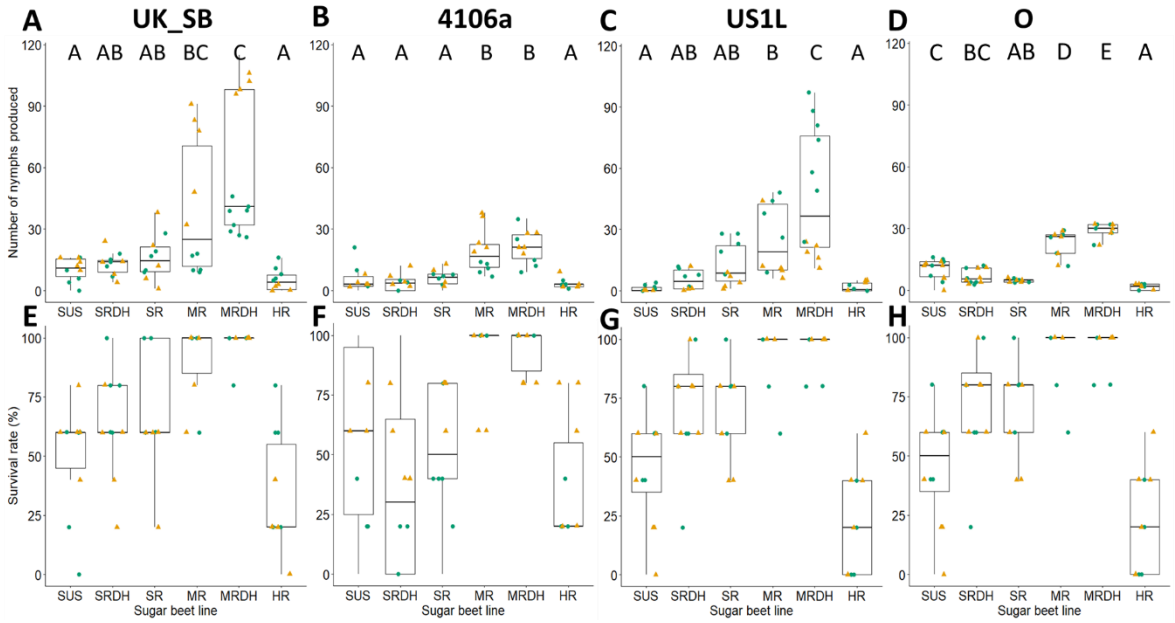


Figure 3.2.3 Fecundity assay of four *M. persicae* clonal populations on six sugar beet lines with variable presumed resistance levels. Six sugar beet lines with assumed variable resistance levels selected by SESVanderHave were tested with four different *M. persicae* clonal populations; 4106a, O, UK_SB, US1L. **(A)** The mean number of nymphs produced of *M. persicae* clone UK_SB 14 days post-inoculation (DPI), **(B)** 4106a, **(C)** clone US1L and **(D)** O **(E)** Survival rates of five age-synchronised *M. persicae* nymphs of clone UK_SB 14 DPI, **(F)** 4106a, **(G)** clone US1L, and **(H)** O. In total, ten experimental replicates were analysed over two different biological replicates (N=5 of each replicate; biological replicate 1 = orange triangle, and biological replicate 2 = green circle). Letters above boxplots (A, B, AB, C) indicate significant differences between sugar beet lines were determined using ANOVA post-hoc Tukey-HSD (α=0.05).

Few variations were found in the numbers of nymphs produced between the two biological replicates within the experiment. For two clones (US1L and UK_SB), there was a high variation in the number of nymphs produced between the biological replicates on lines MRDH and MR. For clone US1L, I observed a significant difference between the biological repeats for sugar beet lines MRDH (Student's t-test, n=5, P<0.002) but not in MR (Student's t-test, n=5, P=0.14). For clone UK_SB, I observed a significant difference between the repeats for sugar beet lines MRDH (Student's t-test, n=5, P<0.001) and MR (Student's t-test, n=5, P=0.002). No differences in survival rates of the clones UK_SB and US1L were observed between the two biological replicates on MRDH and MR.

SESVanderHave selected an additional ten sugar beet lines that showed some promising resistance levels to *M. persicae* at SESVanderHave. These 10 sugar beet lines were also examined for differences in aphid performance

using the two most distinctive *M. persicae* clones 4106a and UK_SB (figure 3.2.4).

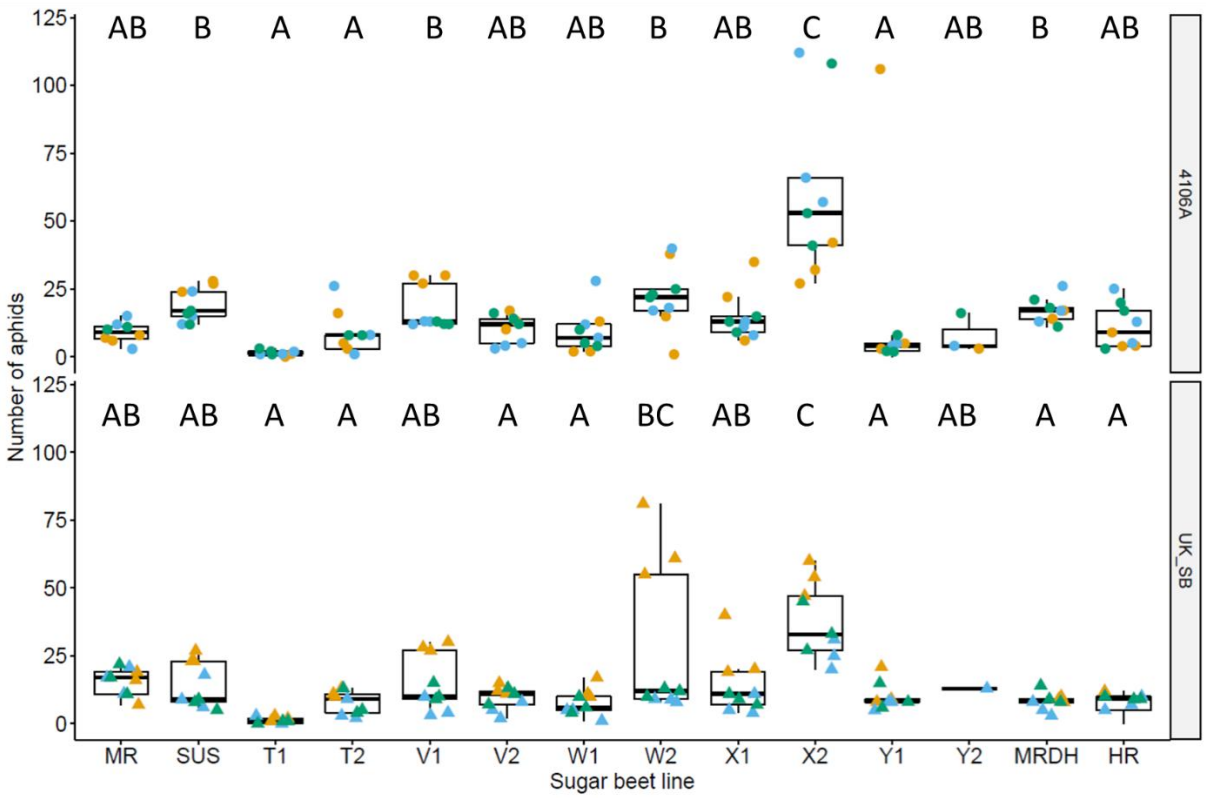


Figure 3.2.4 Fecundity assay with two *M. persicae* clones (4106a, UK_SB) on 10 wild sugar beet lines. Ten sugar beet lines with presumed resistance levels selected by SESVanderHave were tested with two different *M. persicae* clones; 4106a, and UK_SB. Mean number of, *M. persicae* clone 4106a (top) and UK_SB (bottom), aphids found per line. In total eighteen experimental replicates were used over three different biological repeats (biological replicate 1 = blue, biological replicate 2 = orange, and biological replicate 3 = green). Differences between sugar beet lines were determined by ANOVA post-hoc Tukey-HSD (α=0.05). I designed the experiment and analysed the data. The experiment was executed by insectary staff member Susannah Gill under my supervision.

On the whole, no significant differences in performances were found between the two *M. persicae* clones UK_SB and 4106a on the ten sugar beet lines (Student's t-test, N=58, P =0.51). Additionally, in general, no significant differences were found in the number of nymphs produced between the three biological replicates within the experiment. *M. persicae* clone UK_SB produced more nymphs in biological repeat 1 on sugar beet lines W2 and X2 than the other biological replicates did. Both *M. persicae* clones produced the most aphids on the sugar beet line X2. Clone UK_SB produced a significantly lower number of aphids per plant only on sugar beet line T1.

3.2.3 Resistance of sugar beet lines to *M. persicae*-transmitted viruses.

In order to investigate if there is virus resistance present in sugar beet germplasm, I used *M. persicae* clone UK_SB to inoculate eight sugar beet lines with BMYV. In this assay, four of the lines tested are F1 crosses of resistant lines tested in the fecundity assay shown in figure 3.2.4. The four F1 lines, from the previous assay were as follows: SV-JIC-1 was the F1 generation from T1; SV-JIC-2 was the F1 generation from T2; SV-JIC-3 was the F1 generation from V1, and SV-JIC4 was the F1 generation from V2. Furthermore, MRDH and HR were included as controls and to compare with previous fecundity experiments. Additionally, new sugar beet lines SV-JIC-5, SV-JIC-6, SV-JIC-7 and SV-JIC-8 were included. The eight sugar beet lines did not show obvious differences in morphology.

All plants were exposed to 5 BMYV viruliferous aphids. The aphids were removed seven days later. Subsequently, the aphids were removed, and the plants were treated with Pyrethrum to kill any aphids on the plants and moved to a greenhouse for 21 days (at which time virus symptoms became obvious) before two leaves per plant were harvested and send off to SESVanderHave for ELISA tests to examine virus titres. The sugar beet lines showed differences in viral titres. Compared to MRDH, lines SV-JIC-2 and HR had lower virus titres, whereas line SV-JIC-5 had higher virus titres (figure 3.2.5A). Plants from biological replicates one and three showed, on average lower viral titres compared to those of repeat two. Despite these differences, the lines had consistent differences in susceptibilities to the virus.

Some of the lines used in the virus resistance assays were progeny (F1) of lines used for *M. persicae* fecundity assays. These are F1 of T1, T2, V1 and (used in *M. persicae* fecundity assays) that were the V2 SV-JIC-1, SV-JIC-2, SV-JIC-3 and SV-JIC-4 (used in virus resistance assays), respectively. T1, T2, V1 and V2. SV-JIC-3 and V1 were most susceptible to aphids and virus (Fig. 3.2.5). The virus titres of the other lines are all lower than V1, consistent with the lower aphid fecundities on SV-JIC-1, SV-JIC-2 and SV-JIC-4 (Fig. 3.2.5). Many experiments have shown that aphid fecundity is higher on MRDH compared to HR (figure, 3.2.2, 3.2.3, 3.2.4), and consistent with this, MRDH is more susceptible to BMYV than HR is (Fig. 3.2.5). Overall, these data show matching aphid and BMYV resistances of sugar beet lines. Nevertheless, this will need further investigation.

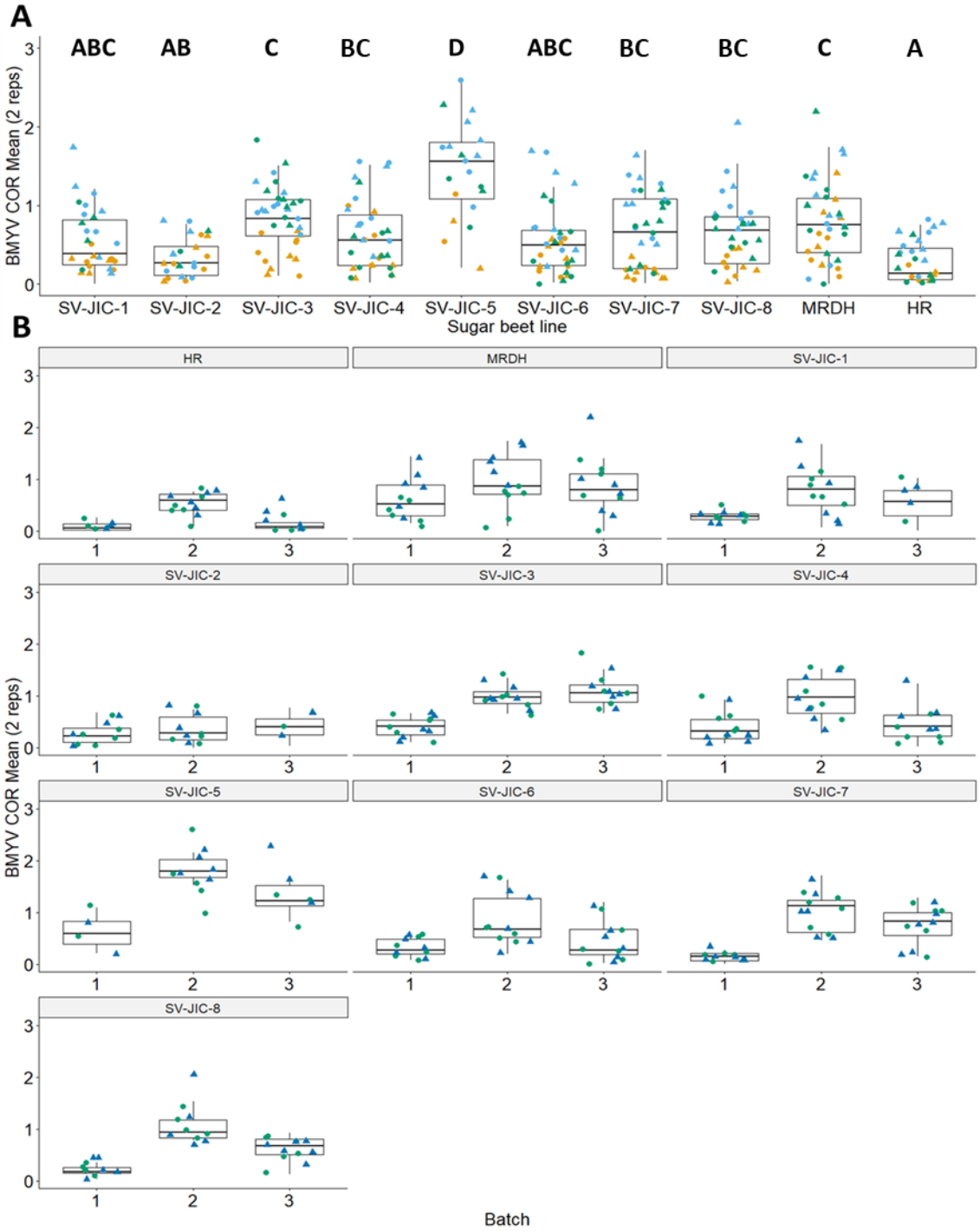


Figure 3.2.5: Virus transmission assay (titre) of BMYV by *M. persicae* UK_SB on sugar beet lines. Eight sugar beet lines with presumed *M. persicae* resistance levels selected by SESVanderHave were tested for transmission of BMYV via viral titre. **(A)** Mean viral titre of sugar beet lines four weeks post-inoculation over three different biological repeats (biological repeat 1 = blue, biological repeat 2 = orange, and biological repeat 3 = green, leaf 1= triangle, leaf 2= circle), and **(B)** Three biological replicates separated out showing mean viral titre found per leaf (leaf 1 = blue triangle, leaf 2 green circle). Differences in viral titre between sugar beet lines were determined using ANOVA post-hoc Tukey-HSD (α=0.05). The experiment was performed by insectary staff Susannah Gill under my supervision and was designed and analysed by me.

3.3 Discussion

Results shown in this chapter showed that sugar beet lines and varieties showed differences in *M. persicae* resistance levels. Aphids did consistently better on MRDH and MR lines compared to SUS, SRDH, SR and HR lines. The outcome of MRDH and MR being more susceptible than SUS and SRDH was surprising given that SESVanderHave had found previously that SUS and SRDH was more susceptible than MRDH and MR. It is possible that the differences in environments in which the experiments were done contributed to the different outcomes. For example, I did the experiments in controlled environmental rooms (CERs), and SESVanderHave did most of their experiments in greenhouse conditions. Light intensity and quality, humidity and consistency of conditions can vary dramatically between CERs and greenhouse environments. Environmental conditions such as light and humidity may affect sugar beet resistance to aphids remains tested. Nonetheless, results in this chapter and previous results from SESVanderHave both show variation in the level of aphid resistance in sugar beet germplasm.

Whereas MRDH and MR were considered susceptible, testing of additional lines showed that there are even more susceptible to sugar beet lines than MRDH and MR. For example, X2 and W2 were found to be extremely susceptible to *M. persicae*. In addition, some lines, such as T1 and Y1, were consistently more resistant to *M. persicae* compared to MRDH and had similar aphid numbers than the HR line. This is despite T1 and Y1 having no obvious morphological, phenotypic differences compared to MRDH (unlike HR). This suggests that mechanisms other than the morphological characteristics of HR contribute to the aphid resistance of sugar beet. These results confirm that there is variation in aphid resistance in sugar beet germplasm.

It was also demonstrated that *M. persicae* clones show differences in performance on sugar beet. Clone 4106a had consistently lower fecundity on sugar beet than the other three clones, and US1L and UK-SB that were originally collected from sugar beet in the field (and then reared on *B. rapa*) reproduced extremely well on sugar beet. Nonetheless, all four clones were more fecund on *B. rapa* than on sugar beet. Therefore, despite US1L and UK-SB being collected originally from sugar beet, these clones do better on *B. rapa* (which is often grown adjacent to sugar beet fields), indicating that *M. persicae* will colonise sub-optimal hosts in field conditions. These results suggest that there is variation among *M. persicae* clones in the ability to colonise sugar beet.

The higher fecundity rates on *B. rapa* in comparison to *B. vulgaris* could be the result of short-term adaptation because of the recent transfer from *B. rapa* to *B. vulgaris*. *M. persicae* does adjust to the host over time and is regulated via transcriptional plasticity (Mathers et al., 2017). A similar transcriptional plasticity mechanism is described in generalist butterfly and spider mite species (Grbic et al., 2011, de la Paz Celorio-Mancera et al., 2013). The stages of adjustment, or transcriptional plasticity, are assumed to be regulated via DNA methyltransferase dependent methylation and cathepsin B cysteine proteases (Mathers et al., 2017), and occur from after a few hours up to months after transfer from a different host (M. Gravino and Y. Chen; personal communication., 2019).

BMYV is a *Polerovirus* that is mostly restricted to the plant phloem. Plant resistance mechanisms to aphids often get apparent when aphids reach the phloem with their stylets. Therefore, it is possible that sugar beet lines that are more resistant to *M. persicae* are also more resistant to BMYV because the aphid is less likely to be able to deliver this virus into the phloem during feeding. Although the high variation between the biological repeats, still significant differences were observed in the virus transmission assay for some sugar beet lines. SV-JIC-2 showed a significant lower BMYV viral titre, while SV-JIC-5 showed a significant higher BMYV viral titre than the control MRDH. Within the virus resistance assay, four lines were F1 populations from the fecundity assay shown in figure 3.2.4. These four F1 lines showed similar patterns in virus transmission compared to the parental lines in the fecundity assay (figure 3.2.5). This could mean that the partial resistance to *M. persicae* found in the sugar beet lines might also confer in reduced viral titre. Yet, no significant correlation or patterns were observed due to high variation within samples and the smaller sample sizes.

The differences in fecundity rates per *M. persicae* clones on the selection of sugar beet lines can be a result of the genetic diversity among the *M. persicae* clones. The overall genetic diversity of *M. persicae* genotypes found worldwide is largely unknown (Margaritopoulos et al., 2009, Fenton et al., 2010). In order to obtain sugar beet lines resistance against *M. persicae,* I need to investigate the population diversity of *M. persicae* further. Obtaining durable plant resistance to aphids has been challenging. A factor contributing to this challenge is that aphids and other insects are known to have different genotypes that can overcome resistance (Arend, 2003). Examples are Greenbugs on wheat & sorghum (Curvetto and Webster, 1989, Kindler et al.,

2001); Russian wheat aphid on wheat (Basky et al., 2001); and rosy and woody apple aphids on apple (Rath-Morris et al., 1998, Young et al., 1982). Additionally, Sun et al. (2020b) showed that different *M. persicae* populations have differential fecundity and survival rates on Capsicum accessions. Hence, more investigation of the nature of genetic diversity within an aphid species is needed for breeding for long-lasting resistance in plants.

Chapter 4. Evaluating existing microsatellites markers for *Myzus persicae*

4.1 Introduction

Population geneticists study genetic variation within populations. They assess this variation by detecting and comparing frequencies of genes and alleles among individuals within a larger population and investigating how these frequencies change over space and time. *M. persicae* population genetic studies mainly used microsatellites (Sloane et al., 2001, Wilson et al., 2004) RAPDs (Random Amplification of Polymorphic DNA) for rapid screening (Criniti et al., 2006). These tools have allowed for distinguishing *M. persicae* populations worldwide, but do not relate to a phenotype, such as the ability to colonise a plant. Microsatellites were predominantly used for population genetics studies of insects (Thangaraj et al., 2016, Zhang et al., 2019, Qin et al., 2016), such as the rice weevil (*Sitophilus oryzae*) (Thangaraj et al., 2016), fruit fly (*Bactrocera correcta* (Qin et al., 2016) and tea green leafhopper (Empoasca (Matsumurasca) onukii Matsuda (Zhang et al., 2019).

Previous *M. persicae* population genetic studies were primarily based on microsatellite markers (Zhao et al., 2015, Margaritopoulos et al., 2009, Wilson et al., 2003, Wilson et al., 2004, Blackman et al., 2007, Turcotte et al., 2013, Popkin et al., 2016, Monti et al., 2016, Sloane et al., 2001, Ramsey et al., 2007). Microsatellites are simple sequence tandem repeats that consist of 2-6 nucleotide repeats in DNA sequences recurring 5-50 times. The number of repeats within these microsatellites are highly polymorphic among individuals within a population (Vieira et al., 2016). The chances of detecting mutations in tandem repeats, such as microsatellites, are tenfold higher than those in non-repeated areas (Gemayel et al., 2012). Microsatellites have been used to study, for example, natural selection rates (Nielsen, 2005), allelic fixation indexes (Slatkin, 1995), genetic diversity (Waits et al., 2000, Monti et al., 2016), bottlenecks (Spencer et al., 2000) and population sizes (Kohn et al., 1999, Fenton et al., 2005, Blackman et al., 2007). Upon next-generation sequencing becoming more affordable, microsatellites are not the preferred method for in-depth population genetics studies any longer. However, microsatellites are still useful and commonly used for rapidly screening differences among individuals (Allendorf et al., 2010).

Fourteen *M. persicae* microsatellite markers were described by Sloane et al. (2001) and Wilson et al. (2004) (Table 4.1). However, most population genetics studies of *M. persicae* have been using only maximally eight microsatellite markers (Blackman et al., 2007, Margaritopoulos et al., 2009, Sanchez et al., 2013, Popkin et al., 2016). *M. persicae* is a generalist that can colonise multiple divergent plant species and is a significant pest in numerous

economic crops (Van Emden and Harrington, 2017). However, the population dynamics and how this relates to the ability of *M. persicae* sub-populations to colonise these plant species remain poorly investigated. To address this, I wished to assess the effectiveness of microsatellites for high-throughput screening by establishing their sensitivity and resolution.

Table 4.1.1 Microsatellite markers used for *M. persicae* genotyping (Sloane et al., 2001, Wilson et al., 2004). PCR prog., PCR programme settings as previously described by Sunnucks and Hales (1996); All three programmes initiate with a 94°C denaturing step of 2 minutes, followed by one cycle of each v, w, x, and y °C for 30 seconds each, then 72°C for 45 seconds, 94°C for 15 seconds to then have 30 cycles at temperature z°C to finish then with a 72°C extension step of 2 minutes ; PMS1 (v=62, w=61, x=59, y=57 and z=55), PMS2 (v=55, w=53, x=51, y=49 and z=47), and AMS1 (2* v = 65, 2* w = 64, 2* x = 63, 2* y = 61 and z = 60).

Locus	Motif	Size range	Primer sequence	PCR prog.	GenBank no.	reference
M35	Impure $(AT)_9$~$(AC)_{22}$	178-198	F: GGCAATAAAGATTAGCGATG R: TGTGTGTATAGATAGGATTTGTG	PMS1	AF233240	Sloane et al. 2001
M37	$(AC)_{16}$	155-157	F: GTGTGAGTAAGTCGTATTG R:TTGTATTATGTACCTGTGC	PMS1	AF233241	Sloane et al. 2001
M40	$(AC)_{17}$	123-135	F: ACACGCATACAAGAATAGGG R: AGAGGAGGCAGAGGTGAAAC	PMS1	AF233242	Sloane et al. 2001
M49	$(AC)_{31}$	130-199	F: CCCATACATACCTCCAAGAC R: AGAGAGAAAATAGGTTCGTG	PMS2	AF233243	Sloane et al. 2001
M55	$(CA)_{16}CT(CA)_8$	119-129	F: TTAATCAATAACTGCTCAC R: GAAGTAGGCAGACACG	PMS1	AF233244	Sloane et al. 2001
M62	$(CA)_6$~$(CA)_{12}$	127-143	F: CGCTGGGGACGAAAAACCTG R: AACAAAAAACCGAAAACCCG	PMS1	AF233245	Sloane et al. 2001
M63	$(AC)_{29}$	163-207	F: GCGGTTTTCTTTGTATTTTCG R: GATTATGGTGCTCGGTGG	PMS2	AF233246	Sloane et al. 2001
M77	$(TA)_7$	138	F: ACACTGCAATCGTGTTATAC R: TTATATTGTATGGGCGGCGG	PMS1	AF233247	Sloane et al. 2001
M86	$(CA)_{23}$	97-141	F: TCCACTAAGACCTCAAACAC R: ATTTATTATGTCGTTCCGCC	PMS1	AF233248	Sloane et al. 2001
M107	$(AT)_{11}$	133-145	F: TAAAAAACACACAATACACA R: GACACCAATGAATGACC	PMS1	AF233239	Sloane et al. 2001
myz2	impure $(GA)_{30}$	177–207	F: TGGCGAGAGAGAAAGACCTGC R: TCGGAAGACAGAGACATCGAGA	PMS1	AY429659	Wilson et al. 2004
myz3	$(GA)_{18}$	111–125	F: GGTGTCCTGCGTTATGATTATG R: ATTCTTTTCCCGGCAGTTTAC	PMS2	AY429660	Wilson et al. 2004
myz9	impure $(GA)_{53}$	204–238	F: AACCTCACCTCGTGGAGTTCG R: CTTGGATGTGTGTGGGGTGC	AMS1	AY429661	Wilson et al. 2004
myz25	$(AG)_{24}$	119–126	F: AACCCATCTCACTCGTCAGCC R: GAATCTGGAGAGCGGTTAATGC	AMS1	AY429662	Wilson et al. 2004

In genetics studies of UK *M. persicae* populations, Margaritopoulos et al. (2009) and Fenton et al. (2010) showed predominantly five genotypes of which one was designated as a “super clone”, designated clone O at the time of the study. In total, Fenton et al. (2010) distinguished sixteen different *M. persicae* genotypes using six microsatellite markers (Myz9, M35, M40, M49, M63 and M86). During the genotyping, they monitored *M. persicae* genotype dynamics over a period of 14 years in the UK. Most *M. persicae* were being collected from suction traps near crop fields during this time, making it unclear which crop species these aphids were colonising before they ended up in the

traps. Only occasionally, the aphid was collected from the crops themselves. Hence, these data do not reveal associations of aphid genotypes and crop species in field conditions. Furthermore, Margaritopoulos et al. (2009) found three genetic clusters using six microsatellites on 197 aphid samples collected from five host plants at fifteen countries worldwide between 1982 and 2006. They found a multilocus Fst value of up to 0.154 when comparing distinct geographical populations and an overall value of 0.086.

M. persicae is presumed to have originated from Asia (Waterhouse, 1993). Two *Myzus persicae* population genetics studies were conducted in China (Zhao et al., 2015, Li et al., 2015a). Li et al. (2015a)used seven microsatellites to investigate the population genetics of *M. persicae* and showed three distinct genetic clusters and a multilocus Fst value of 0.1215 among 23 *M. persicae* populations collected from peach trees. Zhao et al. (2015) used five microsatellites on 54 Chinese populations collected from *Nicotianae* species and found two distinct genetic clusters. One genetic cluster was designated as the *M. persicae* subspecies *nicotianae* and with an Fst of 0.3353 found to have a high degree of genetic differentiation to the other cluster, as shown in most aphid population genetics studies using microsatellite markers, such as for *Sitobion avenae* (Simon et al., 1999, Llewellyn et al., 2004), *Rhopalosihum padi (Delmotte et al., 2002)*, and *M. persicae* (Margaritopoulos et al., 2009, Fenton et al., 2010, Li et al., 2015a, Guillemaud et al., 2003, Margaritopoulos et al., 2007).

In chapter 3, the four *M. persicae* clones (UK_SB, US1L, O and 4106a) used in the fecundity assays showed differences in fecundity and survival rates on a variety of sugar beet lines. These results suggest genotypic variations among the *M. persicae* clones that result in performance variations on sugar beet. However, these four clones have not yet been genotyped.

This chapter determined if six microsatellite markers could be successfully used to distinguish *M. persicae* genotypes as applied in previous population genetics studies (Fenton et al., 2005, Margaritopoulos et al., 2009, Fenton et al., 2010, Zhao et al., 2015). I first optimised the sensitivity and resolution of the microsatellite amplicon size quantification and then attempted to improve sensitivity by removing unspecific amplification. Overall, I found that the existing microsatellite marker technology generated unsatisfactory results and does not lend itself for efficient high-throughput genotyping of field-collected aphids.

4.2 Results

4.2.1 Verification of five microsatellite markers

4.2.1.1 Verification assay using a polyacrylamide gel

I used six microsatellite markers to genotype *Myzus persicae* O, US1L. FRC and 4106a reared in the insectary and assess the sensitivity and resolution of the use of a selection of microsatellite markers. The following microsatellite markers were used; Myz9, M35, M40, M49, M86, M63, (Sloane et al., 2001, Wilson et al., 2004). This selection of microsatellites was made on the usage of them in most of the published population genetics studies of *M. persicae*. Additionally, microsatellite marker sizes for genotype O were already specified and could be used as verification of the microsatellite markers. I assumed that these microsatellites markers would be an appropriate subset to test the initial strength of using microsatellite markers in a population genetics study of *M. persicae*.

In order to test the quality of the microsatellite markers (M35, M40, M49, M86 and Myz9) for *M. persicae*, a verification screen for *M. persicae* genotypes was performed. In the initial protocol verification, we observed unspecific amplification in each microsatellite marker. As *M. persicae* is a diploid, we would not expect more than two bands per microsatellite marker. Nevertheless, within the polyacrylamide gel, more than two bands were found in most samples. This suggests that more than two areas were amplified per single microsatellite. The five microsatellites were amplified with NEB TaqPolymerase with the settings as described in table 4.1.1, according to Sloane et al. (2001) and Wilson et al. (2004). Amplicons were run on a 6 % denaturing polyacrylamide electrophorese gel and visualised with silver staining (figure 4.2.1.1).

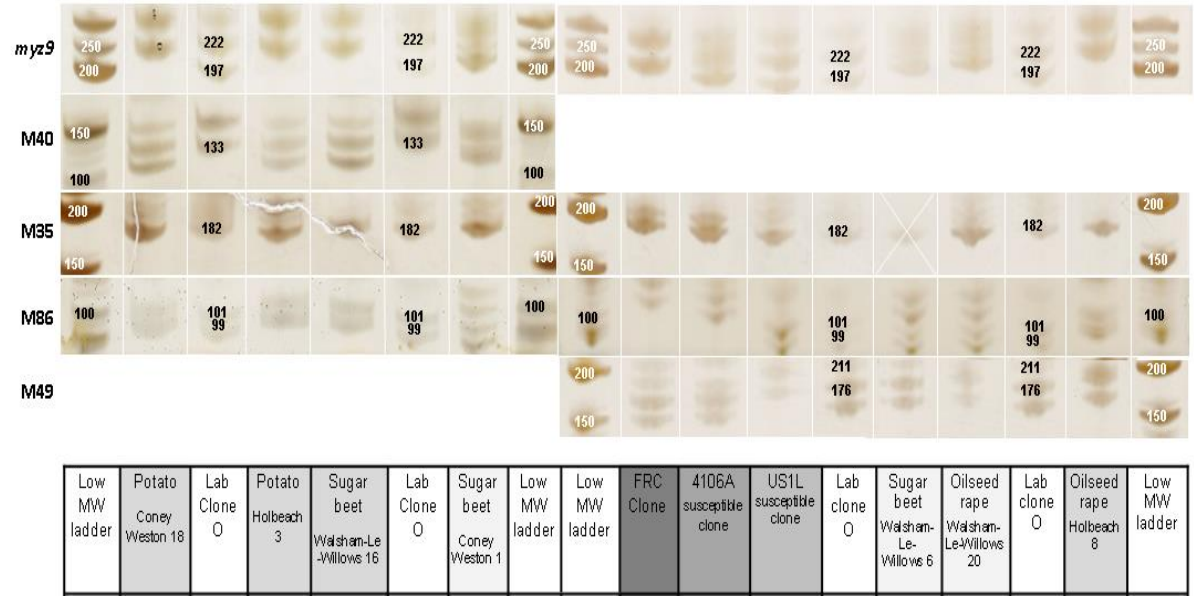


Figure 4.2.1.1 Preliminary scan of quality of microsatellite marker amplification shows unspecific amplification. (A.) PCR gel picture show amplification of five microsatellites markers (M35, M40, M49, M86 and Myz9). Low molecular weight ladders sizes are labelled within the gel and genotype O microsatellite marker sizes are labelled for comparison. **(B.)** Host and location of the collection. Danielle Goff-Leggett performed the experiments.

Due to the unspecific amplification and low resolution, I was unable to detect distinct genotypes using the five microsatellites. The known microsatellite marker sizes for genotype O were found as described for all five markers: 222 and 197 bp for Myz9, 133 bp for M40, 182 bp for M35, 101 and 99 bp for M86 and 211 and 176 for M49. Nevertheless, in all five markers, one or two additional bands were observed, which indicates that in all markers, unspecific amplification occurs. With the current protocol used for the analysis of the microsatellites, I was not able to verify the microsatellite marker sizes. The DNA ladder used did not give enough resolution to indicate the size of bp accurate. Therefore, it is unsure if the amplification of the microsatellite area was specific enough and also whether it was the correct size.

4.2.1.2 Verification assay using Sanger sequencing

To understand if the microsatellite markers were of the correct size, verification of the amplicons was conducted on reference genotype *M. persicae* clone O via sanger sequences. *M. persicae* genotype O has its genome sequenced (Chen et al., 2020, Mathers et al., 2020b) and the microsatellite sizes are well known of this genotype (Sloane et al., 2001, Wilson et al., 2004). We found specific amplification and correct size quantification, as described in Sloane et al. (2001) and Wilson et al. (2004) for the microsatellite sizes of genotype O using the five microsatellite markers. Nevertheless, this was

obtained by cutting out two bands of which are found around the right size. Microsatellite markers were amplified with similar protocols and settings, as described in table 4.1.1. Amplicons were extracted from the electrophoresis gel and sent for Sanger sequencing. I was able to verify microsatellite sizes, as described in (Fenton et al., 2010)(figure 4.2.1.2). This shows that no unspecific amplification was found in the five tested microsatellite markers, and we were able to verify the microsatellite sizes of genotype O. Therefore, another technology may be needed to be able to also illustrate the sizes of the microsatellite markers with higher sensitivity and resolution.

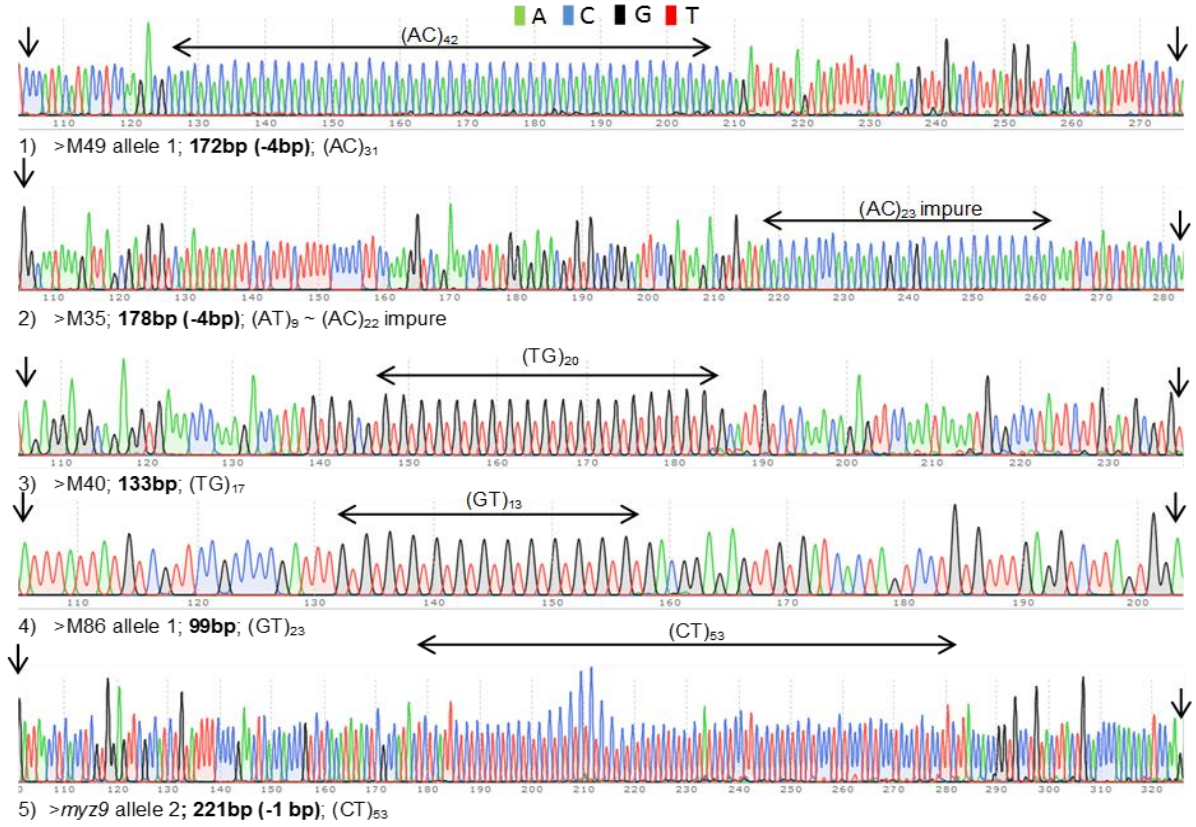


Figure 4.2.1.2 Verification of the microsatellite markers M35, M40, M49, M86 and Myz9. Sanger sequences of the five microsatellite markers. Name, size and sequence of the microsatellite were shown below the sequence. Arrows above the sequence indicate the number of repeats found per microsatellite marker. Colours represent the nucleotide (A= Green, C = blue, Black = G and red = T). Danielle Goff-Leggett performed the experiments.

4.2.1.3 Conclusion

Microsatellite markers M35, M40, M49, M86 and Myz9 were verified and do amplify the right sizes for genotype O as described in Sloane et al. (2001) and Wilson et al. (2004). However, I was not able to verify these sizes within high resolution on a polyacrylamide gel. I was only able to verify the sizes via Sanger sequencing. Additionally, one to two unspecific amplicons were found in the gel, and this could limit the usage of screening multiple

field-collected samples on a larger scale. Therefore, (1) a more sensitive technology with a resolution needs to be used to check the microsatellite marker sizes and (2) the PCR-programme needs to be optimised to reduce unspecific amplification.

4.2.2 Optimisation of microsatellite marker analysis

4.2.2.1 Verification assay using the AATI fragment Analyser.

To verify the amplification of the microsatellite markers M35, M40, M49, M63, M86 and Myz9, I tested using a more sensitive machine (AATI Fragment Analyzer) that also has a higher resolution (down to 2 bp). I used the five previously tested microsatellites and an additional sixth microsatellite with a similar PCR-programme as M49. I found that analysis of six microsatellites on four *M. persicae* genotypes (US1L, UK_SB, O and 4106a) resulted in a similar result as shown before (figure 4.2.2.1). Too many unspecific amplifications gave considerable noise that prevented conclusions regarding the genotype. To address this, I need to improve the specificity of these microsatellite markers.

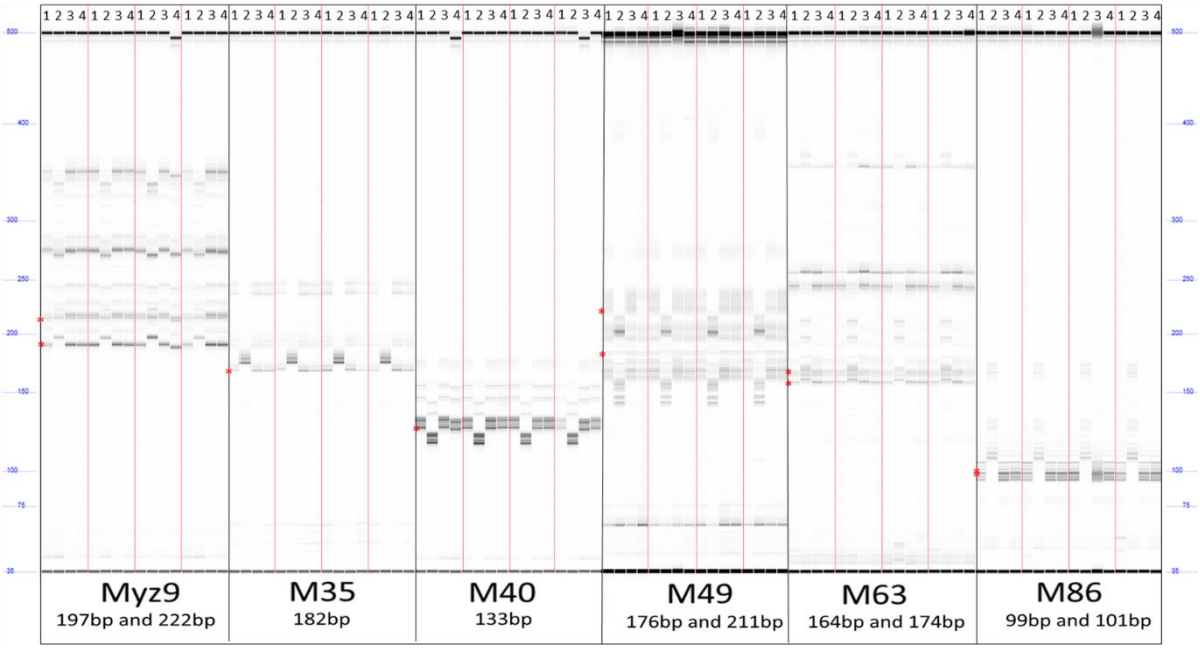


Figure 4.2.2.1 Microsatellite markers M35, M40, M49, M63, M86 and Myz9 of *Myzus persicae* genotypes O, UK_SB, US1L, FRC and 4106a. E-gel that is calibrated via a higher, lower marker and ladder. Ladder sizes are shown at left and right of the digital gel. Names of microsatellite markers M35, M40, M49, M63, M86 and Myz9 are indicated with expected sizes of alleles for clone O below. Red lines indicate technical replicates (N=4). Red stars indicate known marker sizes for Clone O. Numbers on the top indicate M. persicae genotype; 1 = O, 2 = 4106a, 3 = US1L and 4 = UK_SB.

4.2.2.2 Optimisation of the PCR-programme using M. persicae genotype O

To optimise the microsatellite markers in the Hogenhout group, I adjusted the suggested published protocols for the microsatellite markers in Sloane et al. (2001) and Wilson et al. (2004), three microsatellites markers with PMS1 PCR programme, 2 with PCR programme PMS2 and 1 with programme AMS1. I used a gradient touch-down PCR programme to assess the most efficient annealing temperatures for the specific microsatellite markers. Six different annealing temperatures were used per microsatellite marker. Annealing temperatures were shifted between 48°C and 75°C dependent on the microsatellite markers' recommended PCR-programme (Sloane et al., 2001, Wilson et al., 2004) (table 4.2.2.1).

Table 4.2.2.1 Settings touch-down gradient PCR program for microsatellite markers. (A) Location of the six different microsatellite markers in a 96-well plate. Numbers in brackets indicate the different gradient treatment. (**B)** The programme for all microsatellite markers was initiated with a 94°C denaturing step of 2 minutes, followed by two cycles of each v, w, x, and y °C for 30 seconds each, then 72°C for 45 seconds, 94°C for 15 seconds, then followed by 30 cycles at temperature z °C to finish then with a 72°C extension step of 2 minutes. For Myz9; v =55-75, M35; w = 53-73, M40 x = 51-71, y = 49-69 and z =47-67.

A

	1	2	3	4	5	6	7	8	9	10	11	12
A	MYZ9 (1)		MYZ9 (2)		MYZ9 (3)		MYZ9 (4)		MYZ9 (5)		MYZ9 (6)	
B	M35 (1)		M35 (2)		M35 (3)		M35 (4)		M35 (5)		M35 (6)	
C												
D	M40 (1)		M40 (2)		M40 (3)		M40 (4)		M40 (5)		M40 (6)	
E	M49 (1)		M49 (2)		M49 (3)		M49 (4)		M49 (5)		M49 (6)	
F												
G	M63 (1)		M63 (2)		M63 (3)		M63 (4)		M63 (5)		M63 (6)	
H	M86 (1)		M86 (2)		M86 (3)		M86 (4)		M86 (5)		M86 (6)	

B

	1	2	3	4	5	6	7	8	9	10	11	12
V	54.8	55.2	56.5	58.3	60.7	63.3	66	68.6	71	73	74.4	74.9
W	52.8	53.2	54.5	56.4	58.7	61.3	64	66.7	69	71	72.4	73
X	50.8	51.3	52.5	54.4	56.7	59.3	62	64.7	67.1	69	70.4	71
Y	48.8	49.3	50.5	52.4	54.7	57.3	60.1	62.7	65.1	67	68.4	69
Z	46.9	47.3	48.6	50.4	52.8	55.4	58.1	60.7	63.1	65.1	66.4	67

A single *M. persicae* genotype (O) was used to verify the specificity of the six microsatellite markers. The expected band sizes were found with each microsatellite marker. Yet, still, one or two more bands were observed per microsatellites. Although, I found that this approach did reduce the unspecific amplification but could not remove it completely. Warmer annealing temperatures, above 62°C reduced amplification of the microsatellite markers including the unspecific bands (Figure 4.2.2.2). Thus, a reduced amplification could give less aspecific amplification without losing the microsatellite marker amplification. However, I am not sure if this reduction of amplification with

reduced unspecific amplification would suffice for genotyping multiple *M. persicae* genotypes with unknown microsatellite marker sizes.

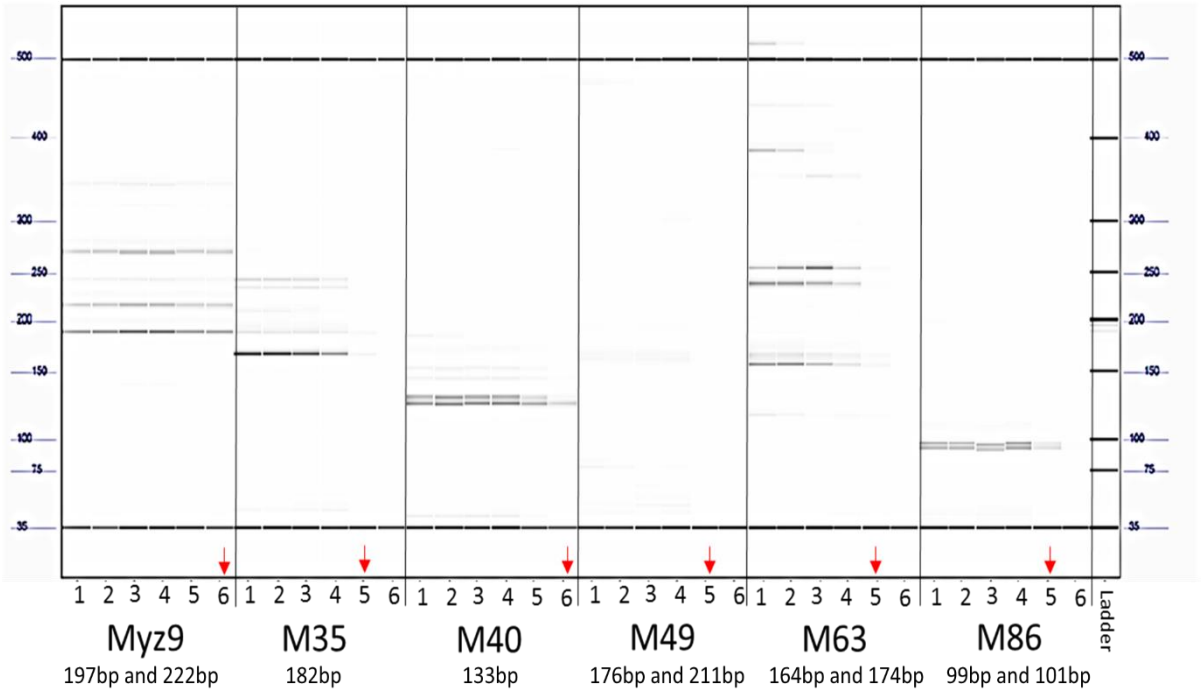


Figure 4.2.2.2 Optimisation of PCR-programme for the six microsatellite markers. Six microsatellite markers were tested and optimised on one *M. persicae* genotype; O. Microsatellites markers were amplified using six different PCR programmes and run on the AATI fragment analyser equipped with the DNF900-55 kit. Ladder sizes are shown at left and right of the digital gel. Names of microsatellite markers M35, M40, M49, M63, M86 and Myz9 are indicated with expected sizes of alleles below. Red arrows indicate PCR programme with least unspecific bands per microsatellite marker.

4.2.2.3 Further verification of PCR-programme adjustments using multiple M. persicae genotypes; O, 4106a, US1L and UK_SB.

In order to verify the microsatellite markers for multiple *M. persicae* genotypes, I used six microsatellites with the, as previously shown, adjusted PCR-programmes on four *M. persicae* genotypes: O, 4106a, US1L and UK_SB. Nevertheless, too much unspecific amplification was again acquired, and I was unable to verify the predicted microsatellite sizes per genotypes. I observed a similar pattern in microsatellite sizes between three of the four aphid genotypes tested; O, US1L and UK_SB. Due to the high amount of background noise, I was unable to predict the actual sizes of these samples (figure 4.2.2.3). Therefore, I found that using a touch-down gradient PCR for the six selected microsatellites was unsuccessful. Furthermore, overall, this indicates that microsatellites markers may be too unspecific as a tool for a population genetic study of *M. persicae*.

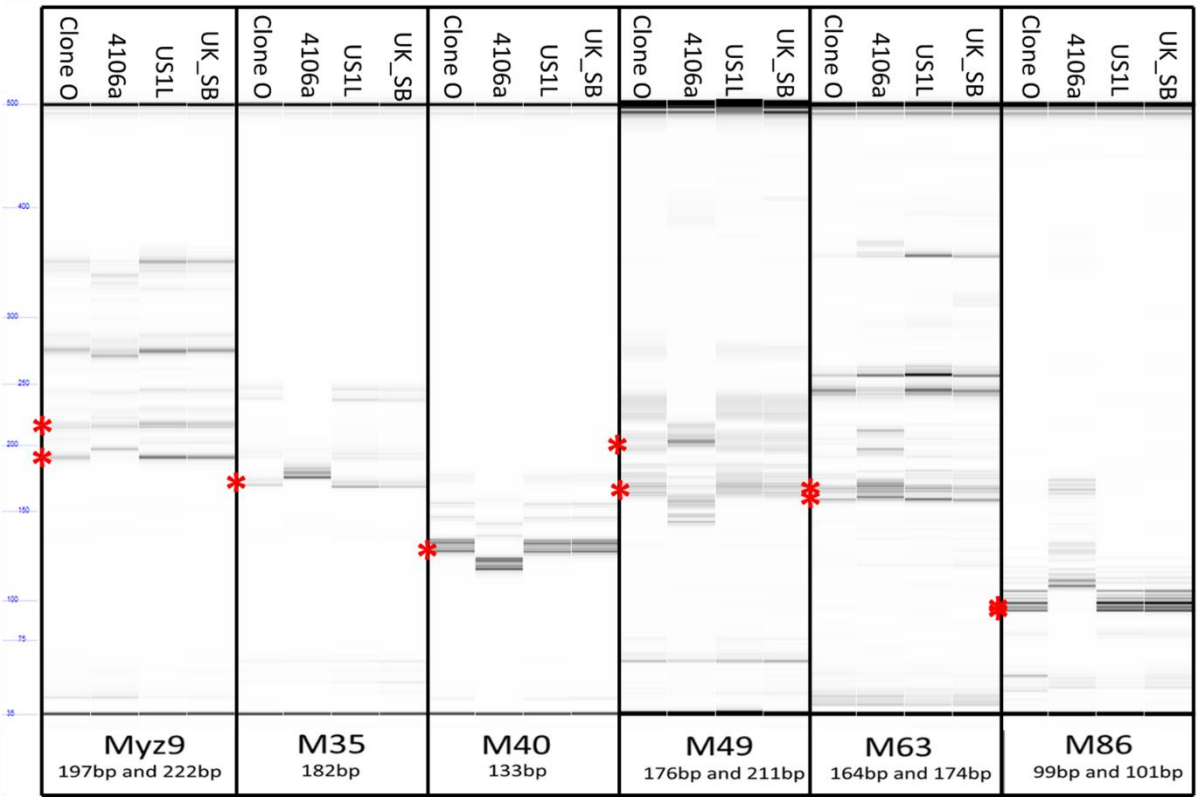


Figure 4.2.2.3 Verification of PCR-programme of microsatellite markers. Four *M. persicae* genotypes; O, 4106a, US1L and UK_SB were tested with six microsatellites markers that were amplified using three different PCR programmes adjusted according to the touch-down gradient. Ladder sizes are shown at left and right of the digital gel. Names of microsatellite markers M35, M40, M49, M63, M86 and Myz9 are indicated with expected sizes of alleles for Clone O below. Red stars indicate known marker sizes for O. Names on top indicates *M. persicae* genotypes (O, 4106a, US1L UK_SB).

4.2.2.3 Conclusion

Altogether, I was not able to optimise the microsatellite markers to get strong amplification of the six markers, without at least one unspecific band. I found that the microsatellite marker sizes for *M. persicae* genotype O were as indicated in Sloane et al. (2001) and Wilson et al. (2004). Yet, I received multiple unspecific bands that were never discussed in other *M. persicae* population genetics study using microsatellite (Wilson et al., 2003, Blackman et al., 2007, Margaritopoulos et al., 2009, Fenton et al., 2010, Zhao et al., 2015). I am not sure how sensitive the microsatellite markers are and how precise the amplification would distinguish *M. persicae* genotypes. One should first reassess the primers using the whole-genome of genotype O available (Mathers et al., 2020b). With the current reference genome, I could check the strength and specificity of the annealing position. Nevertheless, the question arises if the use of microsatellites is sensitive enough, and if it gives a suitable resolution for high throughput screening of *M. persicae* isolates collected in the field?

4.3 Discussion

The data presented suggest that the use of the previously reported microsatellite markers of *M. persicae* were too unspecific to be used to answer our question about how diverse *M. persicae* populations are collected from different locations and host. The use of either the AATI Fragment Analyzer or a polyacrylamide electrophorese gel was unable to improve the resolution and sensitivity for analysing the microsatellite markers. I was able to distinct 4106a from the other tested genotypes, but O, US1L and UK_SB remained undistinguishable.

The use of microsatellites has been the standard for many years within the population genetics studies of *M. persicae*. Nevertheless, I find that microsatellites have too low throughput and too high inaccuracy. I observed too much unspecific amplification and too low sensitivity within the microsatellite markers to analyse *M. persicae* populations genetics in-depth on a larger scale. More than two bands were shown per microsatellite marker. *M. persicae* is a diploid organism, and hence there should be a maximum of two bands found per single microsatellite marker. They used these same microsatellite markers in previous publications and specified a size range and specific microsatellite size per genotype. Nevertheless, no representation of the microsatellite sizes, accuracy and amplification were shown in any of these publications.

Microsatellites are most sensitive to evolutionary processes and can vary among loci (Vieira et al., 2016). Therefore, microsatellites are most useful to study recent evolutionary events. Interestingly, a low FST is still found in population genetics studies of *M. persicae* using microsatellites (Margaritopoulos et al., 2009, Fenton et al., 2010). The low FST values illustrate low diversity within *M. persicae*. Nevertheless, Microsatellites are well known not to be distributed randomly within the genome (Lawson and Zhang, 2006) and with many microsatellites, it is unknown if the frequency estimation was correct. Additionally, the selection of a small number of microsatellites that are highly variable between genotypes can result in ascertainment bias in population genetics studies (Petit et al., 2005, Vali et al., 2008, Guillot and Foll, 2009, Eriksson and Manica, 2012, Haasl and Payseur, 2011, Li and Kimmel, 2013, Putman and Carbone, 2014). The use of a low number of microsatellites that are highly variable will result that a high number of samples per population is needed. Nevertheless, we are currently unsure how diverse *M. persicae* is and how many populations are found within Europe. The collection around Europe could result in a low number of samples

per country. Because of the high variability within a microsatellite marker and only a few microsatellite markers used, these assays do need larger sample sizes to be accurate in estimation of allele frequencies. The high mutation rate may violate demographic model assumptions (Ziegler et al., 2009, Bhargava and Fuentes, 2010, Grover and Sharma, 2011). Thus, Microsatellites can be used for a large-scale study of recent evolutionary events, but for in-depth population genetics and dynamics studies, whole-genome sequences are needed.

Whole-genome sequencing and analysis of single nucleotide polymorphisms (SNPs) could give a more in-depth view of the diversity of *M. persicae*. In addition, whole-genome analyses have less ascertainment bias than microsatellites (Glover et al., 2010, Garke et al., 2012, Ozerov et al., 2013, Putman and Carbone, 2014) and have greater power to detect population structure compared to microsatellites. For large studies, only a few SNP loci are needed to detect structure between populations. However, when a short divergence time is found, a higher number of SNPs may be needed to investigate population structure accurately (Charlesworth, 2009). The whole genome-scale SNP analysis can also lead to the development of new SNP markers that can be used in high-throughput screening. Thus, whole-genome approaches may be used for in-depth genotyping and, in some cases, associated with phenotypes. These include whole-genome Illumina sequencing (Chewapreecha et al., 2017, Barnes and Breen, 2010) and 'MYbaits' techniques (Hopmans et al., 2014, Li et al., 2013, Jouet, 2016). Both have provided deeper insights into population dynamics (Chewapreecha et al., 2017, David et al., 2016, Harris et al., 2010, Mutreja et al., 2011, Nistelberger et al., 2016, Kollias et al., 2015).

In order to be able to investigate the genetic diversity within *M. persicae,* I need to investigate the use of other technologies and markers. The increased development high availability of next-generation sequencing technologies results in an opportunity to use whole-genome sequencing to study the population dynamics of *M. persicae*. Additionally, SNP markers can be developed and will lead to a higher resolution and more in-depth insight into the population genetics of *M. persicae* found around Europe. These resources are required to push the population genetics studies of *M. persicae* to the next era. Therefore, pre-empt the population genomic study, samples would need to be collected worldwide, and the reference genome needs to be improved to a higher resolution.

Chapter 5. Optimisation of HMW DNA extraction from aphids for long-read sequencing.

5.1 Introduction

In the previous chapter, it was found that the current genotyping markers for *M. persicae* do not satisfactorily separate individual aphids from different clones/genotypes. Therefore, it would be beneficial to develop new markers. For optimal marker development, it would be useful to access to whole-genome sequence data of the aphid species. Fortunately, whole-genome sequence data are available for M. persicae clones O and G006 (Mathers et al., 2017). Nonetheless, these draft assemblies are highly fragmented. To be able to map future genotyping markers within aphid chromosomes, including for example autosomes versus sex chromosomes or association with chromosomal regions prone to recombination, it would be better to have a chromosome level assembly of *M. persicae*.

To obtain a chromosome-level assembly, it is necessary to generate long reads that enable the scaffolding of the contigs already available *for M. persicae* clones O and G006. Several third-generation-sequencing technologies are available for generating long-read sequences. These include PacBio and Oxford Nanopore Technologies (ONT).

PacBio is developed by Pacific Biosciences of California, Incorporated. PacBio uses SMART cells to traditionally sequence reads from 10 kb up to 16 kb. Before sequencing, a SMRTbell is created. This is a closed single-stranded DNA created with ligating hairpin adaptors on each end. The SMRTbell is loaded on an SMRT cell where the SMRTbell diffuses in a nanophotonic visualisation chamber called the zero-mode waveguide (ZMW). In the ZMW, it is possible to measure the excited fluorescence particles' light pulse when freed from the base pair added to the sequence during amplification. This base pair specific fluorescence colour that is activated upon binding of the base pair to the sequenced is analysed and turned in the specific nucleotide letter (Rhoads and Au, 2015).

ONT is a long-read sequencing technology that is developed in 2014 (Ip et al., 2015), using small pores (nanopores). The sequence devices use flow cells that contain 2048 nanopores on 512 processing channels or 12000 nanopores on 3000 processing channels, depended on the sequencer. The technology is able to use electric resistance levels to sequence DNA or RNA molecules of any size. Therefore, no size selection has to be performed before sequencing on an ONT device such as the Minion. Here arises the opportunity to sequence an undefined DNA or RNA molecule size. ONT libraries do consist

of similar templates as the PacBio technology. They use hairpin adapters with a single DNA strand with its complement attached on the other end of the hairpin adapters. A consensus sequenced in developed using the template and the complement sequence by removing the adapter. ONT is a technology that is continuously under improvement. New sequence devices such as MinION, GridION, Flongle and PromethION are further developed, including their library preparation kits and flow cells.

Pacbio and ONT Minion technologies have each their advantages and disadvantages and can be used as complementary techniques. Pacbio has a long-read limitation of 10-16 kb, whereas ONT has no limit in the size of long-read sequencing. The very long reads are particularly useful for covering high-repetitive DNA regions, common in aphid genomes (Lu et al., 2016) and are particularly recombination prone and hence useful for distinguishing (closely related) aphid clones/genotypes. However, the Pacbio generates more accurate sequence data compared to the ONT. Whereas both technologies have high error rates of around 15% based on single read sequences, in the Pacbio technology, the cloned long read fragments are sequenced multiple times, and these reads can be overlaid to generate a consensus sequence that is accurately reflecting the real sequence with max 1-2% error rates. In contrast, the DNA fragments in the ONT are sequenced only twice (in both reaction of the DNA molecule) and generates a 15% error rate. Nonetheless, the ONT error rate may be overcome by sequencing the genome at higher coverage (± 40-fold often suffices). The optimal situation would be created by overlaying the Pacbio, ONT and short-read Illumina data.

For Pacbio and ONT Minion to generate useful data, it will be necessary to generate high molecular weight pure DNA from the aphids. In Pacbio, the high mol weight DNA will be sheared into the requested kbp fragments (5-30 kbp fragments), which will then be ligated the hairpins to develop SMRTbell constructs. With the presence of many small fragments (< 10-kbp), adapters used for generating the PacBio library will be ligating to these small fragments too, and in the subsequence size-selection step, all these short fragments will be filtered out. Therefore, excluding small fragments will dramatically increase the long-read yields of PacBio libraries. For the ONT Minion, the small fragments are more likely to enter the pores of the flow cells and could saturate the pores before long fragments are able to enter them. Therefore, it is essential for both technologies to avoid high numbers of small DNA fragments being present.

Past protocols in the lab have focused on the CTAB method developed by Marzachì et al. (1999) for aphid DNA isolations. These have worked well for generating Illumina-based short-read sequencing. However, the DNA isolation method may have to further optimised to get long-read data. In the chapter, I focused on obtaining pure high-molecular-weight DNA from aphids. I found that the CTAB method does not enable the isolation of high-molecular-weight (HMW) DNA. I have optimised the DNA isolation technology to acquire pure HMW DNA from aphids that resulted in obtaining long-read sequence data (Wouters et al., 2020a).

5.2 Results:

5.2.1 Evaluation of the CTAB extraction method for obtaining long-read sequence data on the ONT MinION.

5.2.1.1 Nanodrop and Qubit assessments

First, I assessed the quality of DNA samples that I obtained with the standard CTAB extraction method (Mugford et al., 2020) from groups of aphids. The purity of high molecular weight DNA after extraction is usually examined by Nanodrop and Qubit analyses.

The purity of high molecular weight (HMW) DNA after extraction is usually examined by Nanodrop and Qubit analyses. Nanodrop quantifies the DNA amount via DNA solution absorbances at spectra of 260nm and 280nm and the 260/280 ratio. Nanodrop 260/280 ratio of ~1.8 or higher indicates pure DNA samples, and anything lower than this ratio suggests contaminations with salts, proteins or DNA extraction reagents (Gallagher, 2001). Absorptions at the 230nm spectrum indicate contamination of organic compounds, such as sugars. Therefore, for a pure sample, a 260/230 ratio of ~2 (2/1) would be wanted (Liu et al., 2009). The Qubit kit used for DNA quantification uses a fluorescence dye that attaches to only double-stranded DNA strands, and the dye does not bind single DNA strands, RNAs, nor organic compounds. To better determine if preparations contain pure HMW DNA, the nanodrop and qubit values are often compared. A ratio of 1.5 - 1 Nanodrop/Qubit ratio indicates expected DNA purities, whereas ratios of > 1.5 are considered not suitable for Pacbio/ONT sequencing (Schwessinger and Rathjen, 2017).

Using this method, I obtained over 500 ng of DNA that has a Nanodrop light absorbance 260/280 ratio of ~1.8, but 260/230 ratios were lower than 1.8 (table 5.1.1). This suggests the presence of high organic compound contaminants resulting in the DNA samples failing quality controls. I then further purified the resulting DNA using the ZYMO genomic DNA purification and concentration kit and found an incremental improvement in DNA purity but similar yield and DNA fragment sizes. After DNA extraction, the 260/280 ratio was around 1.8, but the 260/230 ratio was too low in samples before purification. Following purification, the Nanodrop/Qubit ratio reduced to below 1.5. Thus, although the use of the ZYMO genomic DNA purification and concentration kit did improve DNA purity, there was a considerable loss in yield.

Table 5.1.1 DNA quantification of CTAB DNA extraction method before and after Zymo genomic DNA purification and concentration kit. DNA was extracted from ~30 aphids per sample. The sample number specifies the extracted sample, and the P with the number indicates that the sample had an additional purification step with the ZYMO genomic DNA purification and concentrator kit. After column purification, the concentration dropped around 3-fold, but the quality was improved to the approved standard for whole-genome sequence on the nanopore (samples were diluted in 50µl demineralised, sterile water (dH_20).

Sample Name	Nanodrop (ng/µL)	A260 / A280	A260 / A230	Qubit (ng/µL)	Nanodrop/ Qubit ratio
P1	10	1.8	1.1	8	1.2
1	84	1.7	1.6	28	3.0
P2	14	1.9	1.5	10	1.4
2	140	1.7	1.7	26	5.4
P3	48	1.9	2.2	36	1.3
3	283	1.7	1.8	89	3.2
P4	51	1.9	2.2	41	1.2
4	142	1.7	1.8	66	2.2
P5	44	1.9	2.2	35	1.3
5	165	1.7	1.8	54	3.1
P6	52	1.8	2.2	40	1.3
6	288	1.7	1.7	96	3.0

5.2.1.2 Evaluation of fragment sizes

Because the nanodrop/qubit values for the P1-6 samples (Table 5.1) were adequate, I decided to take these samples forward for evaluation of DNA fragment sizes. I verified the DNA fragment sizes via a Pulse-field gel electrophoresis gel of P1-P6 and two non-purified samples. Pulse-field gel electrophoresis is a technique used to separate large DNA fragments. I used a clamped homogenous electrical field DNA 4.9-120 KB ladder (Bio-Rad) to

check the sizes of the DNA molecules. Unfortunately, the ladder and the samples used in this experiment were partly degraded. Additionally, I found that the pulsed-field gel electrophoresis was not sufficiently sensitive for the DNA fragment size analysis.

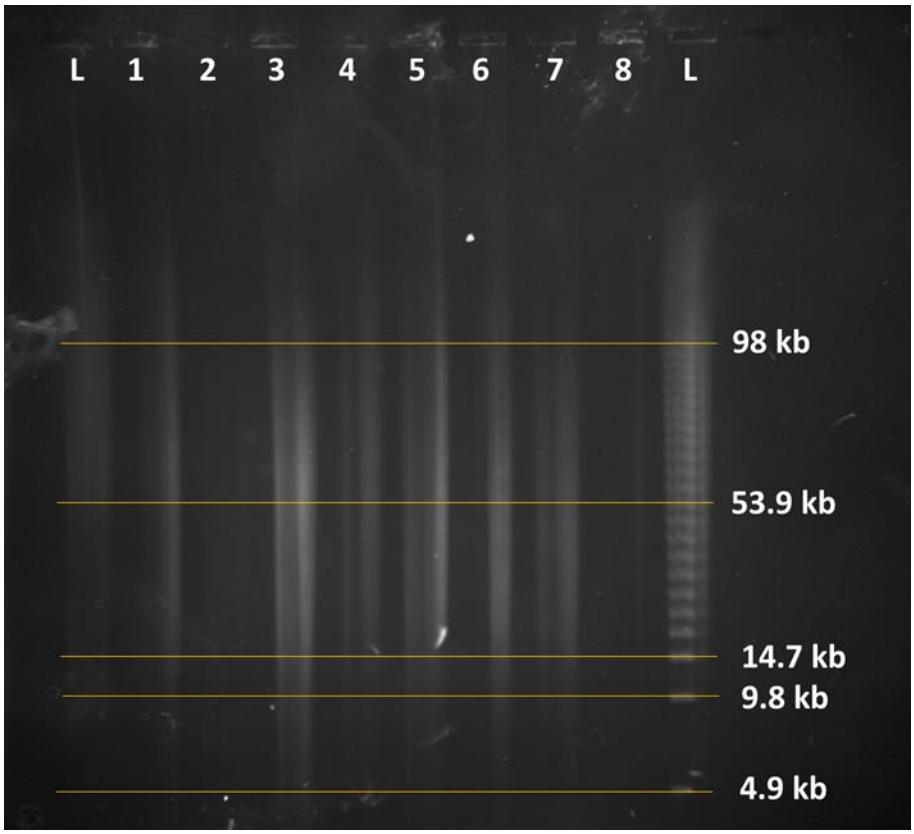


Figure 5.1.1 Field pulse gel electrophoresis of CTAB DNA extraction method shows degraded ladder and samples. Gel band number indicates the sample number, of which 1-6 were the purified samples P1-P6 and 7 and 8 represent the non-purified samples. L indicates the ladder. After long term storage of the 4.9-120 KB ladder (Bio-Rad) the high molecular bands were partly degraded and unable to distinguish in size (indicate sizes next to graph).

Next, I used a capillary system, the AATI fragment analyser Femto-pulse, for accurate size quantification of up to 165 kb DNA fragments. The femto analyser pulse is calibrated according to a bottom and top marker and a size ladder. The samples did show high relative fluorescence units (RFU) over 50kb in size (dotted line), a peak of the highest RFU above 100 kb and a low number of fragments smaller than 10kb was observed (Figure 5.1.2). Therefore, most of the fragments are above 50 kb in size. This appeared to be sufficient for ONT Minion sequencing. Hence, the samples were taken forward for library production using the LSK108 library kit and subsequent sequencing on one flowcell.

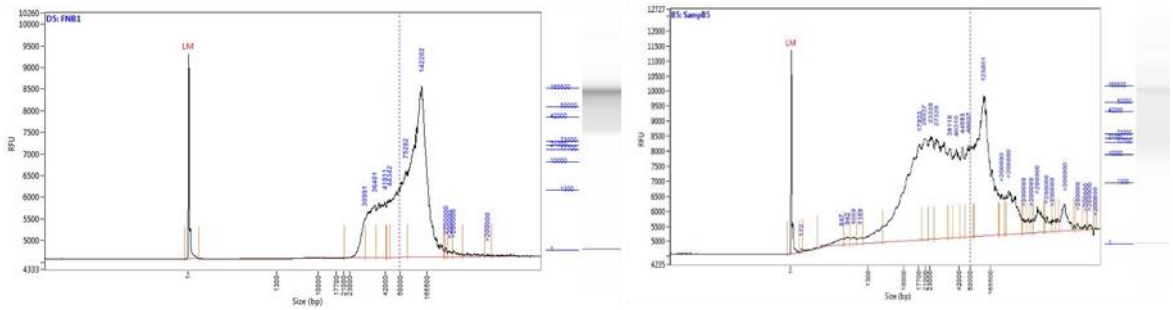


Figure 5.1.2 Femto-pulse e-gel relative fluorescence unit readouts for size indication. Fragment length estimates by the AATI Fragment analyser Femto-pulse, with fragment length on the y-axes (vertical purple dotted line indicates 50kb) and the abundance of the size fragments in relative fluorescence units on the x-axes. This shows that I have reasonable fragment sizes, mostly above 50 kb.

5.2.1.3. Obtaining ONT Minion long-read sequence data

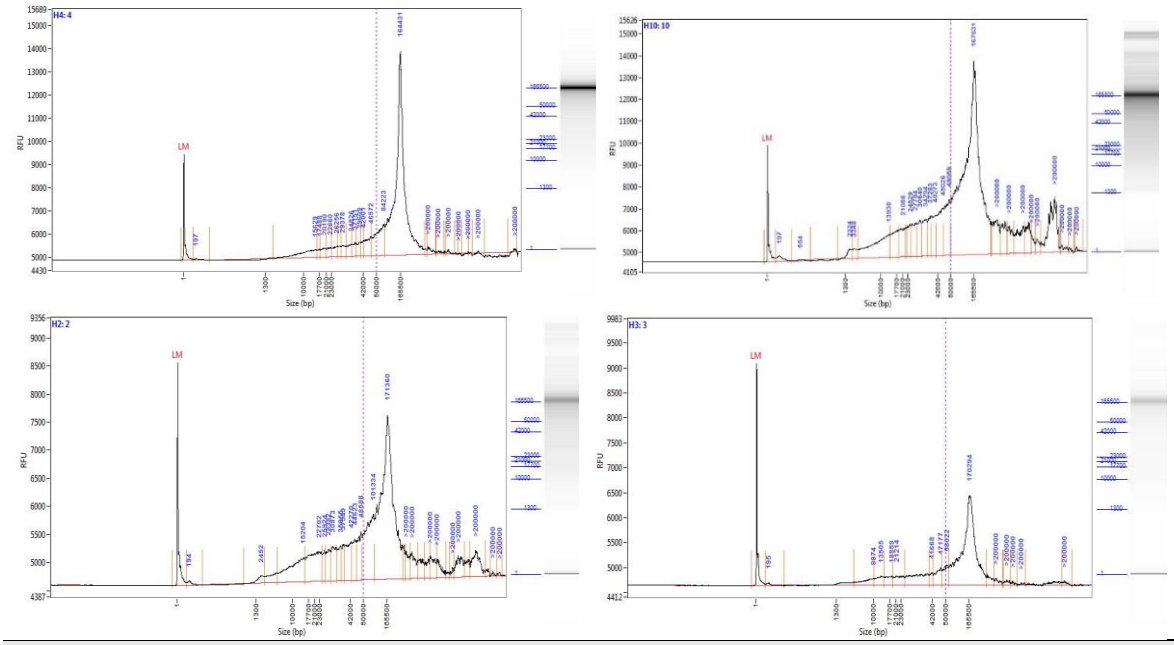
Following the Minion run on the computer, I checked the sequencing of DNA molecules in real-live. It showed that in the first 24 min/hours, six Gb of data was generated. However, the pores generated only a short sequence reads. In fact, 50% of the reads sequenced were below 3570 < bases, and only 20% of the reads sequenced were above 5120 bases had long reads. At the end, in total 10 GB data was generated in both runs [N80=1845, N50= 3570, N20= 5120] (table 5.2). The largest fragment sequenced was only 40,331 bases long, while the Femto results did show a large proportion of fragments larger than 50 kb. Thus, although we had high-molecular-weight DNA that was verified on the Femto-pulse, I was unable to get the same representation after library preparation. It is possible the DNA was sheared via the use of the G-tubes (Covaris) or via pipetting. In conclusion, whereas there were long fragments in the DNA preparations, these fragments did not end up in the pores. This may be for several reasons, and one possibility is that during the library preparation, I conducted too much shearing within the optional G-tube step.

Table 5.1.2 Summary statistics of ONT Nanopore MinION run one and run 2. Table of summary statistics including read-length scope, N50, number of reads and total yield.

Read length (bases)	Read length N50 (Kb)	Number of reads	Total amount of data (Gb of sequence)
100 - 40,331	3.570	1,553,552	2.88
100 – 67,101	3.376	3,950,007	7.98

To assess if the problem was the shearing of the DNA during library preparation, I removed the (optional) first step of the ONT ligation library prep from further library preparations, cut the pipette tip for wider openings and pipetted slower to remove the potential frictions that can shear DNA. I then

repeated the library preparation on DNA samples using the same DNA extraction and purification but was unable to achieve long-read sequencing on the ONT MinION (figure 5.1.3). The four samples did show high RFU after 50kb in size with again a peak of the highest RFU above 100 kb, and a low number of fragments smaller than 10kb was observed (Figure 5.1.3). Therefore, most of the fragments are above 50 kb in size. These fragment sizes should have been sufficient for long-read sequencing on the ONT Minion. I processed the samples with the ligation library preparation kit (LSK108) and ran them on a 1D flow cell (9.4). I stopped the flowcell prematurely after I noticed that no long fragments were sequenced, after around 0.33 gigabases of yield. Many smaller fragments (<2 kb) were sequenced, and the longer fragments, observed in the Femto-pulse e-gel, were not represented in the sequencing data (N80= 821, N50= 1,726, N20= 2,163). The max read length I achieved from the four ONT MinION runs was 24,824. Therefore, without the optional shearing step, change of the DNA extraction or pipetting style and the pipette tips, I was unable to obtain long-reads sequences on the ONT Minion.



Read length (bases)	Read length N50 (Kb)	Number of reads	Total amount of data (Gb sequenced)
100 – 24,824	1.120	451,135	0.45
100 – 7,816	3.376	346,980	0.32
100 – 8,659	1.084	320,683	0.31
100 – 16,478	1.022	793,247	0.75

Figure 5.1.3 Summary statistics of ONT Nanopore MinION run 3, run 4, run5 and run 6. Table of summary statistics including read-length scope, N50, number of reads and total yield. Illustrated are the fragment length estimates by the AATI Fragment analyser Femto-pulse, with fragment length on the y-axes (vertical purple dotted line indicates 50kb) and

Relative fluorescence units on the x-axes. This shows that I have reasonable fragment sizes but are not sequenced on the ONT Nanopore flowcell.

5.2.1.4 Conclusion

Altogether, the above data suggest that the CTAB extraction method, plus additional purification via the Zymo genomic DNA purification and concentrator kit, may not suffice to extract pure HMW DNA. Therefore, these DNA extraction methods need further improvement. A possibility is that the preps have too many organic compounds (e.g. sugars), because of the higher 230 absorbance values (low 260/230 ratio). Perhaps, these sugars have to be removed before we are able to obtain long-read sequences from HMW DNA.

5.2.2 Evaluation of the Illustra Nucleon Phytopure DNA extraction method to obtaining pure HMW DNA for long-read sequence data on the ONT Minion.

5.2.2.1. Nanodrop and Qubit assessments

The Illustra Nucleon Phytopure DNA extraction method that is used for extracting plant DNA is known to remove organic compounds. Therefore, in this next step of optimisation, I changed the DNA extraction method and used the Illustra Nucleon PhytoPure kit (GE Healthcare RPN8511) following the producer protocol and wide-bore pipette tips as described (Wouters et al., 2020a). Interestingly, during the quality check, I found that the 260/280 was around 1.8, but the 260/230 ratio was 1 to 1.5, which is on the lower side than the level required. The Qubit/nanodrop ratio was around 1.7 for all samples. Nevertheless, the lower 260/230 ratio, the 260/280 and Qubit/Nanodrop ratio was sufficient, and the genomic DNA samples were further used in Fragment length analyses.

5.2.2.2. evaluation of fragment sizes

The Femto-pulse e-gel showed a high peak at the ~100 kb fragment size with a slope back to 40 kb. The four samples showed high RFU after 50kb in size with a peak of the highest RFU above 100 kb, and a low number of fragments smaller than 40kb was observed (Figure 5.2.1). This indicates that a percentage of fragments were bigger than 100 kb, and hardly any small DNA fragments (below 40 kb) were found in the sample. Therefore, I continued processing these samples for library preparation.

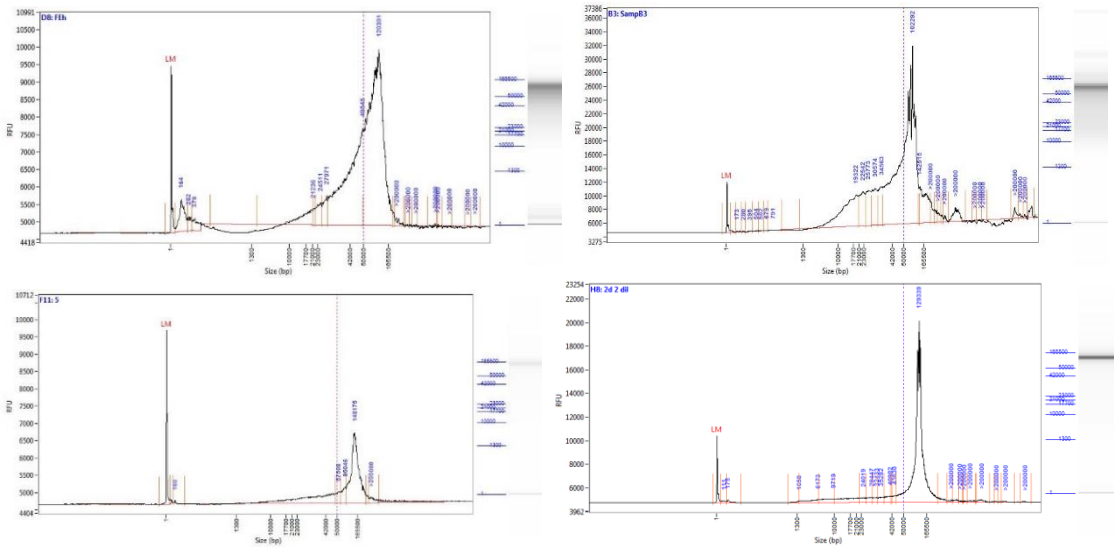


Figure 5.2.1 Femto-pulse e-gel relative fluorescence unit read outs for size indication. Fragment length estimates by the AATI Fragment analyser Femto-pulse, with fragment length on the y-axes (vertical purple dotted line indicates 50kb) and the abundance of the size fragments in relative fluorescence units on the x-axes. This shows that I have reasonable fragment sizes mostly above 50 kb.

5.2.2.3. Obtaining ONT Minion long-read sequence data

Samples were processed with the ligation library preparation kit LSK109 and run on improved flow cells (1D, 9.4 Rev-D). Long-read sequences were achieved with fewer short reads (<3kb). A max sequence read length of ~341 kb was achieved, and a reduced number of reads less than 3 kb. In total, we received an N80=6,324 N50=16,875 and N20= 41,422 as average over the four flow cells (table 5.2.1). The yield of the flow cells reduces when longer reads are sequenced, but I was still able to get an average of 5 gigabases per flow cell. In total, I obtained 20 gigabases over the four flow cells. This was together with Illumina sequences enough quality and coverage to generate a chromosome level assembly. Thomas Mathers then assembled the *M. persicae* genome in six super-scaffolds (the number of chromosomes found in karyotyping analysis). In total, 97% of the contigs were allocated across the six super scaffolds (Mathers et al., 2020b).

Table 5.2.1 Summary statistics of ONT Nanopore MinION run 7, run 8, run9 and run 10. Table of summary statistics including read-length scope, N50, number of reads and total yield.

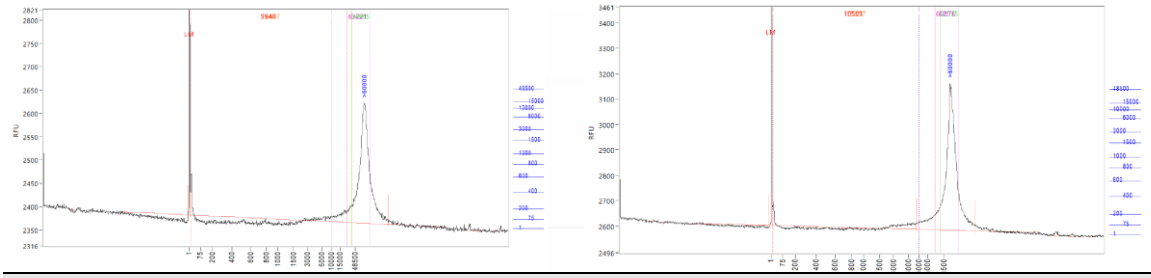
Read length (bp)	Read length N50 (Kb)	Number of reads	Total amount of data (Gb sequenced)
1,000 - 167,408	12.47	252,580	1.7
1,000 - 289,130	11.16	392,332	5.61
1,000 - 341,305	19.76	1,358,149	12.55
1,000 - 306,228	24.21	786,949	8.15

5.2.2.4 Conclusion

Thus, the above data suggest that the Illustra Nucleon Phytopure genomic DNA extraction kit is sufficing to extract pure HMW DNA of *M. persicae* for long-read sequencing on the ONT Minion. Using the DNA extraction kit from Illustra, I have shown that we managed to obtain over 27 gigabases of sequence data with an N50 of more than 16kb on the ONT Minion. We are not sure if the Illustra Nucleon Phytopure genomic DNA extraction kit would be sufficient to generate pure HMW DNA of other aphid species.

5.2.3 Verification of DNA extraction kit to obtain pure HMW DNA for long-read sequence data of another aphid species, *Acyrtosiphon pisum*.

Nevertheless, to test the functionality of this technique on other aphids and sap-feeding insects, DNA of multiple aphid species such as *Acyrtosiphon pisum* was extracted with the use of the Illustra Phytopure DNA extraction kit. As shown with the previous sample, a 260/280 ratio of ~1.8 was achieved, but a lower 260/230 ratio (between 1 and 1.6) using the Nanodrop. The Qubit/Nanodrop ratio was around 1.5 in a demanded ratio. The Femto-Pulse e-gel showed a high fluorescence peak above 100 kb, suggesting high-molecular-weight DNA with most fragments found above 100kb (figure 5.3.1). Ligation ONT library preparation LSK109 was used, and two 9.4 REV-D flow cells were run from a single library. The two flow cells generated 20 gigabases of sequence reads. Thomas Mathers made a chromosome level assembly of *A. pisum* and got 98% of the the data assembled within four super scaffolds, matching to the chromosome number of the species (Mathers et al., 2020b).



Read length (bp)	Read length N50 (Kb)	Number of reads	Total amount of data (Gb sequenced)
1000 - 326,796	33.6	752,925	10.47
1000 - 298,288	32.81	519,605	7.21

Figure 5.3.1 Summary statistics of ONT Nanopore MinION *A. pisum* run 1 and run 2. Fragment length estimates by the AATI Fragment analyser Femto-pulse, with fragment length on the y-axes and Relative fluorescence units on the x-axes. This shows that I have high-molecular weight DNA after extraction and this is sequenced on the ONT Nanopore

flow cell. On the bottom, the table shows the summary statistics, including read-length scope, N50, number of reads and total yield.

5.3 Discussion

Organic compounds are problematic even in the presence of long fragments. The organic compounds in the library preparation are unlikely to plug the ONT Minion flow cell pores because the pores were accessible to the short DNA fragments. If the pores had been plugged by the organic compounds, the yield of sequencing short fragments would also have been affected. Therefore, another explanation could be that the long fragment somehow could not access the pores, whereas the short fragments do. One scenario of how this may happen is that the compound is involved in knotting up the long DNA fragments, and these knotted molecules cannot access the pores. The likelihood of small molecules being knotted by the metabolites may be much lower. Interestingly, aphids have large quantities of sugars in their bodies (as they feed on the sugar-rich plant phloem), and these sugars may be involved in the DNA knotting. Alternatively, plant metabolites may affect the DNA structure. Both scenarios are likely given that the aphid lives on plants and that the Illustra Nucleon Phytopure DNA extraction kit which includes additional purification that was optimised for extraction of plant DNA.

The Illustra Nucleon Phytopure DNA extraction kit uses a Phytopure resin with borate molecules in a solid form that remove metabolites more efficiently during the chloroform cleaning step. In the first runs, we were unable to achieve long-read sequences. Nevertheless, we were getting a high yield of over ten gigabases per flow cell. The high yield suggests that the pores were not blocked by protein, salt, or any other contamination. If there is contamination of some sort, that will block the pores on the MinION flow cell and greatly reduced the yield. Previously, it was shown that it could be laborious to achieve long-read sequences from some organisms, of which especially some plant species due to metabolites. Aphids feed on the phloem of plants and are therefore full of plant metabolites in the body. Therefore, we changed to a plant DNA extraction kit Illustra Phytopure. This established a method to achieve long-read sequences for sap-feeding insects.

During the optimisation of the DNA extraction and library preparation, I opted for multiple changes at the same time. Half-way in the trial-error ONT, I changed the library kit from SQK-LSK108 to SQK-LSK109. The differences between SQK-LSK109 and SQK-LSK108 protocols are the combined FFPE repair

and end-repair, the ligation enzyme and a clean-up step that selects for HMW DNA. These changes lead to an increased ligation efficiency in SQK-LSK109, and with more selection for long fragments, it improved the yield of long-read sequences. Additionally, flow cells keep improving. We started with 1D (9.4) flow cells but then switched to the improved REV-D flow cells. Since ONT is continuously improving its products, I was unable to use similar library preparation kits and flow cell type consistently. Nevertheless, the lack of long-read sequences did not change with the adaption of the library preparation or the flow cells. The improved library preparation was developed to reduce the time per preparation and improve adapter ligation to increase yield. The improved REV-D flow cells were indicated as yield improving but still use the same technique for sequencing.

Because of the optimisation of DNA extraction and obtaining pure HMW DNA, it was possible to generate long-read data for M. persicae clone O and generate a high-quality chromosome-level assembly that was used for analyses in at least two scientific publications (Mathers et al., 2020b, Chen et al., 2020). This improved chromosome level assembly is now being used for population genetics studies, as explained in the next chapters. Moreover, the optimised DNA extraction protocol was also used for a range of aphid species, leafhoppers and froghoppers (spittlebugs). These samples were successfully sequenced using different sequence platforms such as ONT Minion, 10X and PacBio. It has enabled sequencing and - in most cases - chromosome-level assembly of genomes from *Myzus persicae* (Mathers et al., 2020b), *Acyrthosiphon pisum* (Mathers et al., 2020b), *Eriosoma lanigerum* (Biello et al., 2021), *Philaenus spumarius* (Biello et al., 2020) and *'Candidatus Phytoplasma asteris*' RP166, (Cho et al., 2020).

Chapter 6. Optimisation of collection, shipment, DNA extraction and species verification for aphid samples collected from fields worldwide.

6.1 Introduction

M. persicae is a generalist and found on many plant species. In laboratory settings, single clonal lineages of *M. persicae* can readily adjust to colonise Chinese cabbage and sugar beet (Chapter 3). Nonetheless, the clonal lineages show differences in performance on these plant species. Previous publications have shown evidence that, upon host change, M. persicae undergoes massive transcriptional changes, including the P450 gene family and other genes known to process plant metabolites and candidate effector genes (Mathers et al., 2017, Chen et al., 2020). Moreover, there is evidence that *M. persicae nicotinae* appears to be better at colonizing tobacco species than *M. persicae sensu stricto* (Field et al., 1994), and this better colonization ability is associated with the amplification of P450 gene family (Singh et al., 2020a). Taken together, this raises the possibility that a degree of cryptic host specialisation may occur in *M. persicae*, resulting in populations that are structured by host-plant species (i.e., the presence of host-adapted lineages), and this may include *M. persicae* subpopulations that are more likely to be captured from sugar beet in the field. Assessing the potential of the presence of a *M. persicae* population structure that is associated with plant host species is important, because this will lead to improved predictions of how aphids introduce viruses into sugar beet, such as whether viruses are most likely carried by neighbouring sugar beet, other crops or weeds to newly planted sugar beets.

Currently, the population diversity and dynamics of *M. persicae* is largely unknown. Moreover, past and current attempts to assess *M. persicae* diversity in field settings do not take plant hosts into account (Margaritopoulos et al., 2009). Indeed, most aphid capture strategies involve flight traps that capture flying aphids, and it is not known what plant host species these capture aphids were colonizing before they flew into the traps. The work described in this chapter is focused on the collection of aphids directly from sugar beet and other plant species in experimental field plots across Europe. To do this, several methods required optimization, including strategies of aphid collection and preservation of single aphids for DNA purification and genotyping.

Various strategies have been developed to collect aphids from the field, though these are mostly adapted for the taxonomic identification of aphids. A commonly used method is to capture winged individuals using suction-traps

that drop the aphid in water basins when collected from 10/15 meters high (Bell et al., 2020, Grabener et al., 2020). This method captures lots of aphids and is not labour intensive. However, the DNA quality of the aphids and other insects in water reduces rapidly over time (Moreau et al., 2013). Moreover, the plant species the insects were feeding on before capture is not known. Storage of collected insects in ethanol is another common technique (Moreau et al., 2013, King and Porter, 2004). However, DNA leaks out into the ethanol reducing the yield and quality of DNA that may be recovered from a single aphid. Additionally, where small insects have been maintained in ethanol, many individuals were often combined in a single tube. Given that DNA may leak into the ethanol, there is a risk of contamination if samples preserved in this way are used for high throughput whole-genome sequencing (WGS)(Moreau et al., 2013). To obtain pure high molecular weight DNA, snap freezing of living individual insects in liquid nitrogen is the best option. However, access to resources such as liquid nitrogen is limited, and this is not feasible in a remote field location. Moreover, field stations or farmers do not have access to liquid nitrogen nor have the licences in place to handle or ship samples in liquid nitrogen. Altogether, strategies for in-field collections of individuals aphids suitable for good quality DNA extraction and WGS require optimisation.

Maddison (2013) described a storage method used for DNA isolation from field-collected beetles. This method involves putting individual beetles inside a tube with silica gel. The silica gel removes humidity from the beetle sample, thereby preserving the tissue for an extended time at room temperature. DNA extracted from these beetle samples were found to be good quality for sequencing and other analysis. This method might work for aphids as well. However aphids have a softer external exoskeletal and therefore, silica gel can crush the aphid. Hence, the collection tube will have to be adapted for aphids to prevent the crushing of the aphid. Dried insect samples from the museum collection were found to be suitable for DNA extraction (Castalanelli et al., 2010, Gilbert et al., 2007, Thomsen et al., 2009, Chen et al., 2010). Hence, samples stored in silica gel tubes are likely suitable for DNA extraction and genome sequencing.

To obtain enough DNA for genome sequencing, sufficient DNA needs to be extracted from single insects. At least 10 ng DNA yield is necessary for Illumina (low-input) WGS. Aphids are typically around 3 mm in size, with *M. persicae* about 2 mm and thus smaller than most beetles. Therefore, it is not clear if sufficient DNA can be obtained for WGS from such a small insect.

Recently, Kingan et al. (2019) described a method showing that sufficient DNA can be obtained from a single mosquito for WGS. However, the mosquitoes were stored in liquid nitrogen. It remains to be determined if sufficient high-quality DNA can be extracted from single aphids stored in silica gel tubes.

Accurate species-level identification of field-collected samples is required to efficiently carry out genomic surveys of diversity for any given species. *M. persicae* individuals can be distinguished from co-occurring allospecific species by the following characteristics: an oval body shape of a length of 1.2 to 2.1 mm, convergent antennal tubercles in dorsal view and clavate siphunculi with dark to black tips (figure 6.1.1). Although named the green peach aphid, adult wingless females have variable colouring such as yellow, pink, red and from light green to dark green (Blackman and Eastop, 2000b). However, *M.* persicae is often confused by other small green aphids that occur in fields. Moreover, the aphids are collected by field workers that are non-experts in aphid taxonomy, do not have a microscope or hand lenses in the field, and the dried aphids in the silica gel tubes shrink and lose most of their unique characteristics of the species. Therefore, aphid samples that have been shipped need to be genotyped at the species level before being used for WGS.

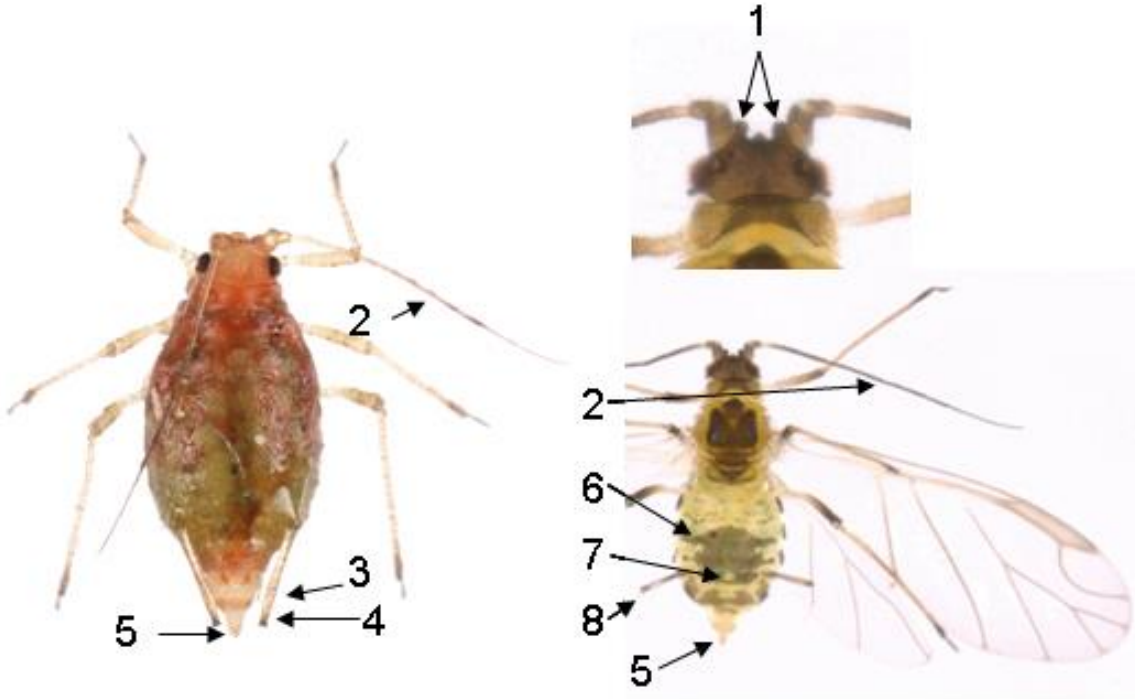


Figure 6.1.1 Annotated photograph of *M. persicae* figure for taxonomical identification. (**1.**) Out-sticking convergent antennal tubercles, (**2.**) Antennae are as long body, (**3.**) The siphunculi are pale, straight to marginally clavated inwards and around two times the length as the cauda, (**4.**) the siphunculi have dark to black tips and are longer in the aptere form. (**5.**) pointed short Cauda, (**6.**) Aptere forms have a shredded dark dorsal patch with

(**7.**) a lighter spot in the middle. (**8.**) siphunculi are darker than the alate form and are an even colour. Reprinted by permission from Encyclop'Aphid (Hullé et al., 2020). Copyright 2020.

A common marker for species identification is the mitochondrial gene cytochrome oxidase subunit I (COI) (Martin and Tooley, 2017, Tobe et al., 2010, Robideau et al., 2011, Tocko-Marabena et al., 2017). COI is commonly used for species-level identification because (1) It is a conserved mitochondrial gene that has a within-species divergence of less than 2% (Hebert et al., 2003). (2) COI barcoding is straightforward to implement in many organisms as the Folmer primers have previously been shown to be successfully amplifying the DNA of many species, including aphids. (3) There is an existing database of sequences available that can be used for Species identification. COI barcoding has been used in many arthropods and species are well characterised online databases such as the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/nucleotide/). The COI based species identification is usually conducted on isolated genomic DNA with the universal arthropod primers LCO1490 (GGTCAACAAATCATAAAGATATTGG) and HCO2198 (TAAACTTCAGGCTGACCAAAAAATCA) (Folmer et al., 1994). These primers amplify a 710-bp COI fragment from aphids (Lee et al., 2011).

The goals of this chapter were: (1) to develop the *M. persicae* silica gel tube collection strategy, (2) to optimise DNA extraction for single aphids to obtain enough yield for whole-genome Illumina sequences, (3) to develop a collection initiative with capture plants around Europe and (4) to identify and select approximately 100 *M. persicae* individuals from various hosts and locations worldwide using COI genotyping.

6.2 Results

6.2.1 Optimisation of sample collection and DNA yield from extractions.

6.2.1.1 Development of storage vesicle for mid-term storage at room temperature

In order to collect samples from rural fields around the world and allow shipment without the use of hazardous substances, collected samples need to be stored at room temperature for at least two weeks without losing yield. To be able to obtain samples with enough quality and quantity for WGS, I developed an affordable method for field collection and storage at room

temperature. I developed a silica storage tube that dries out the aphid and protects the aphid from being crushed by the silica gel. This tube is made from easily accessible lab equipment. The storage tubes include a 20ul filter pipette tip to prevent silica gel from entering (Wouters et al., 2020b) (figure 6.2.1.1). These storage tubes did not crush the aphid within the tube and were able to be sent in large quantities all around the world without any hazardous substances.

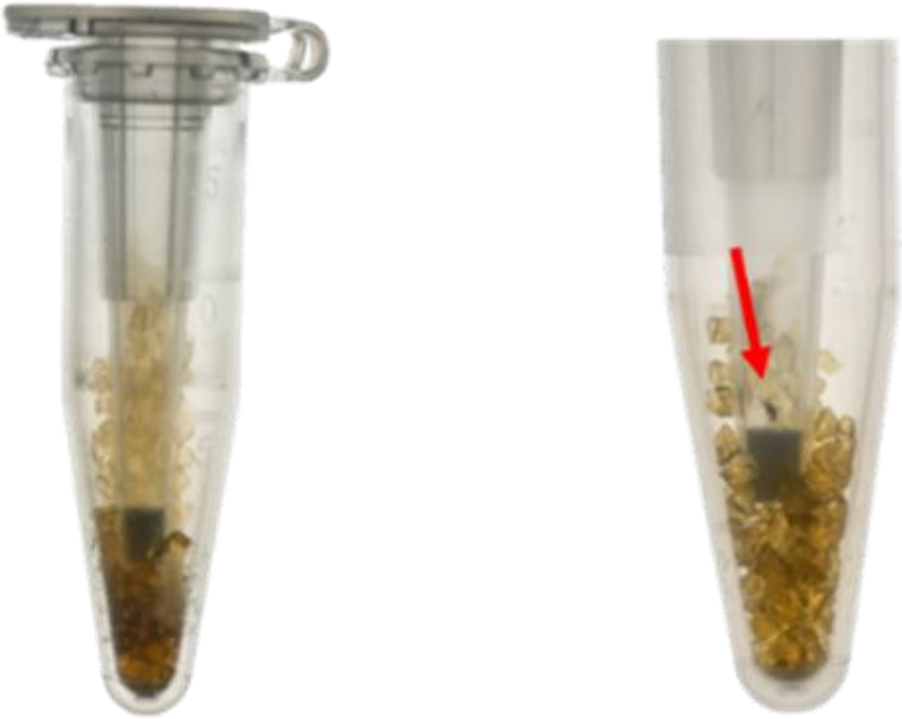


Figure 6.2.1.1 Insect collection tube for mid-term storage on room temperature. Eppendorf tubes with orange silica gel on the bottom. 20µl filter pipette point is put in the middle of the 1.5 ml Eppendorf after cutting approximately the first centimetre of the point. The aphid will then be collected in the pipette point to avoid crushing of the aphid during transport. The red arrow indicates the collected aphid in the middle of the filtered pipette tip.

To assess if aphids stored in the silica tubes were preserved for extraction of high-quality DNA, I collected single aphids from our clone O lab population and either extracted DNA immediately from snap-frozen fresh tissue or stored samples in tubes either with or without silica gel for up to eight weeks. Eight repeats of single aphids were collected in pipette points within tubes either with or without silica gel (N=8). First, samples for eight weeks of storage were collected and stored, with the six-week stored samples collected two weeks later, etcetera. After eight weeks, I snap-froze and extracted DNA from the stored aphids. DNA was extracted from all 104 samples on the same day in two batches and quantified using fluorometry

(Qubit) and spectrophotometry (Nanodrop). Fifty-two samples were extracted at the same time, with four samples from each treatment. On average, I recovered 141 ng of DNA (+/- 10 ng standard error) from freshly collected snap-frozen single aphids based on quantification with Qubit and find that yield diminishes within three weeks of storage (figure 6.2.1.2A). After eight weeks of storage with silica gel at room temperature, I recovered an average of 14.3 ng DNA (+/- 2.2 ng) based on Qubit quantification. In contrast, 5.1 ng (+/- 0.75 ng) of DNA was recovered from samples stored without silica gel after only three weeks based on Qubit quantification. Similar trends were also observed using quantification with Nanodrop, although this method reported consistently higher concentrations of DNA (figure 6.2.1.2B). Thus, I was able to extract high-quality DNA after storing single aphids with silica gel for over eight weeks at room temperature. Additionally, using the same method, I succeeded in obtaining sufficient DNA for Illumina 150PE PCR-based WGS from samples stored for over a year at room temperature.

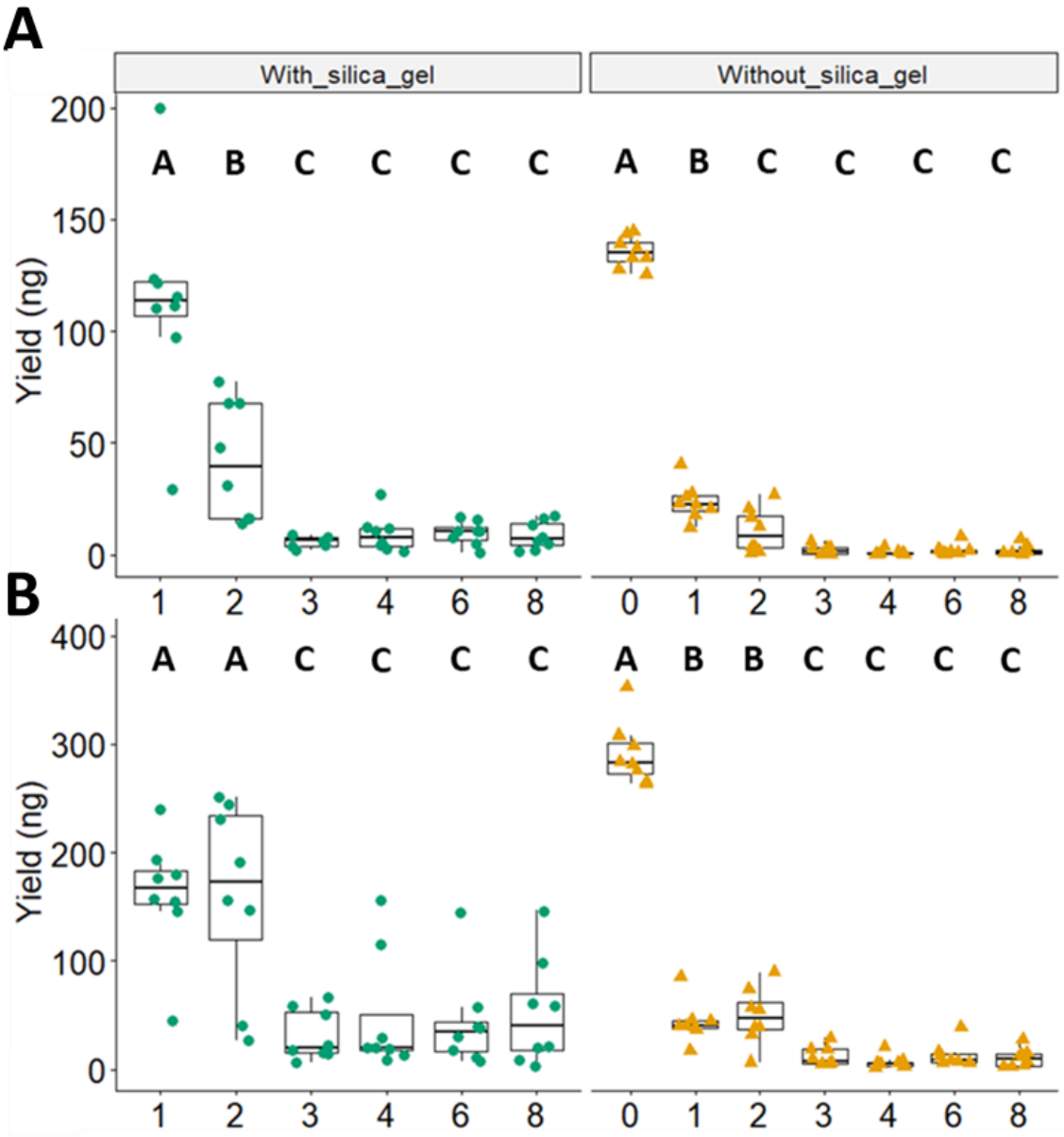


Figure 6.2.1.2 Storage experiment of collection tubes with and without silica gel for up to 8 weeks at room temperature. DNA extraction yields of single *M. persicae* aphids were calculated with either **(A)** Qubit and **(B)** Nanodrop quantification of the yield. The scatterplot shows repeats per storage week. Green circles are samples stored with silica gel, and the yellow triangles indicate storage without silica gel. Boxplot indicated median and interquartile range. Letters above boxplots (A, B, C) indicate significant differences between storage weeks were determined using ANOVA post-hoc Tukey-HSD (α=0.05).

6.2.1.2 DNA yield improvements by optimisation of the DNA extraction from single aphids

In order to obtain DNA samples from single aphids with sufficient quality and yield to construct Illumina libraries for WGS, further optimisation of DNA extraction from individual dried aphids was required. I, therefore, made multiple modifications to the DNA extraction protocol and measured

the effect of these changes on DNA yield. First, I tried to increase the yield by changing the lysis buffer (Figure 1a and 1b). Second, I tried to remove or add steps within the extraction for improved yield and quality of the DNA (Figure 1c and 1d). I found that different lysis buffers did not improve the DNA yield. The most optimised buffer in our lab was found to be the CTAB buffer that was used previously in the lab (CTAB C) (Marzachì et al., 1999). When optimising the extraction protocol by adding and removing steps, only removing the centrifuge step before the chloroform (clean-up step) did increase the amount of yield per extraction (treatment C). Therefore, we used the same lysis buffer as before but changed some steps within the DNA extraction. This was published by Mugford et al. (2020) and used in multiple projects within the Hogenhout group (such as a population genetics study for *Aphis fabae*; unpublished).

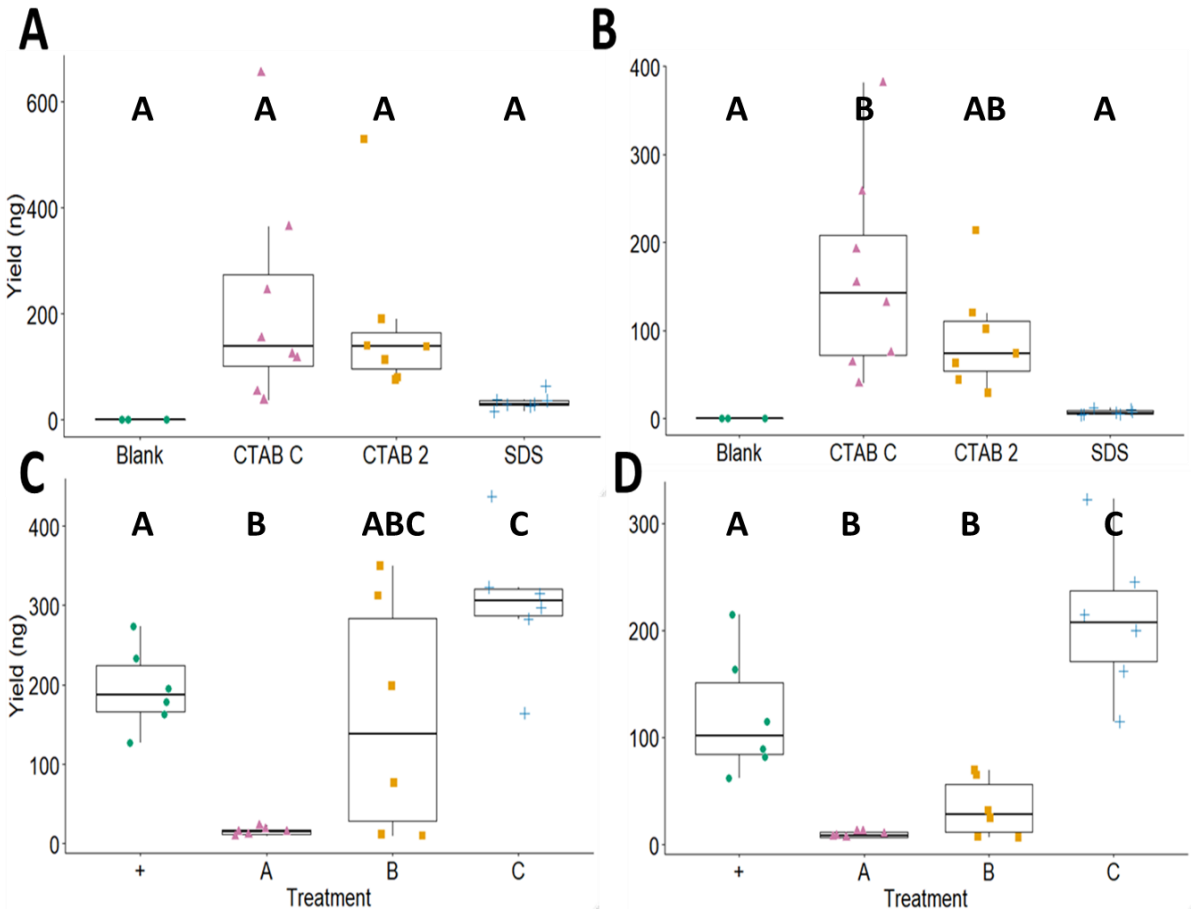


Figure 6.2.1.3 Optimisation of DNA extraction of single aphids changing lysis buffers or by altering DNA extraction steps. DNA extraction optimisation by **(A, B)** changing the lysis buffer and **(C, D)** adding or removing steps within the DNA extraction protocol. **(A, B)** Total yield in ng after using the standard DNA extraction protocol (Marzachì et al., 1999) and using different lysis buffers; Blank = blue circles, CTAB C (Positive control) (Marzachì et al., 1999) = Red triangle, CTAB 2 (Schwessinger, 2017) = orange square and SDS (Mayjonade et al., 2016)= Blue cross. **(A)** is yield quantified by Nanodrop and **(B)** by Qubit. The CTAB buffers had the highest yield recovery, and the two different CTAB buffers do not make a

significant difference (α=0.05) in total yield. **(C, D)** Total yield in ng after altering single steps within the DNA extraction protocol; + (no changes) = Blue circles, A (pipette closer to chloroform layer) = red triangle, B (add additional CTAB after with the Chloroform step) and C (remove the first centrifuge step) = blue cross. The higher total yield is obtained when the first centrifuge step is removed (treatment C) in both quantifications with **(C)** Nanodrop and **(D)** Qubit. Letters above boxplots (A, B, AB, ABC, C) indicate significant differences between treatments were determined using ANOVA post-hoc Tukey-HSD (α=0.05).

6.2.1.3 Conclusion

Optimisation of storage and DNA extraction was successful. Sufficient DNA was obtained from aphid samples stored in the storage vesicles (figure 6.2.1.1) for eight weeks or longer. On average, 14.3 ng DNA +/- 2.2 ng was received from samples stored for eight weeks in the storage vesicles at room temperature. Moreover, further optimisation of the DNA extraction protocol led to an increased yield recovery from an average of 103.3ng +/- 9.4 to 172.6ng +/- 14.2. Both the optimised DNA extraction protocol (Mugford et al., 2020) and the storage vesicles (Wouters et al., 2020b) were widely used in the Hogenhout group. The optimised DNA extraction protocol was used for single aphid DNA extractions. The storage vesicles were used for communal country-wide initiatives such as for BRIGIT (https://www.jic.ac.uk/brigit/) where *Philaenus spumarius* spittlebugs froghoppers were collected from field locations.

6.2.2 Sample collection initiative of aphid samples focussed on *Myzus persicae*

Myzus persicae samples were collected worldwide, with a focus on the European continent. I used multiple strategies to receive samples from around the world. (**1.**) Research labs from multiple countries on different continents collected samples in 2017. (**2.**). The sugar beet breeding company SESVanderHave (SV) planted capture plots in nine of their field stations around Europe. I designed 20 m^2 capture plots that were grown in these SESVanderHave field station locations within Europe. These 20 m^2 plots consist of 5 m^2 of each Canola, Chinese cabbage, potato and sugar beet. Every week of the peak growing season (from May up to August) the plants were checked for aphids, and these were collected from the different host within the capture plots or the surrounding area. A table with the number of samples collected of what host at what date is found in the appendix (table S1).

6.2.2.1 Sample collection of Myzus persicae by research labs

Myzus persicae samples were collected from 2017 through to 2019 from several locations around the world: South Africa (Stellenbosch region), Kenya (Nyeri region), China (Zhejiang region), Australia (Mallee region and Wimmera region, Australia). Additionally, some lab colonies were received from different companies and research institutes. These include Switzerland (S2B01), Argentina (S2196G), France (Les Crespys, Generac, MG1107, FRC09), UK (S1200Q), Spain (ES01, Lierida), Netherlands (NL_WUR and NL_IRS) and Israel (I1). In all cases, samples were collected in silica gel containing collection tubes as mentioned above (figure 6.2.1.1; additional information is found in chapter 2.3) and stored/shipped at ambient temperature and stored at -70°C after arrival in the John Innes Centre.

6.2.2.2 Sample collection of aphids within Europe via capture plots

In order to collect aphid samples from specific crops over 3 years from different regions in Europe, I arranged a collection initiative with SESVanderHave and developed 20m^2 collection plots. Aphid samples were collected near field stations from SESVanderHave and Florimond Desprez around Europe between spring 2017 and summer 2019 (figure 6.2.2.1). SESVanderHave collected samples at ten locations in and around sugar beet fields for three consecutive years, and Florimond Desprez collected aphid samples in and around rose orchards at one location in 2018. In all cases, samples were collected in the above-mentioned silica gel containing collection tubes (figure 6.2.1.1; additional information is found in chapter 2.3), stored/shipped at ambient temperature, and stored at -70°C after arrival in the John Innes Centre.

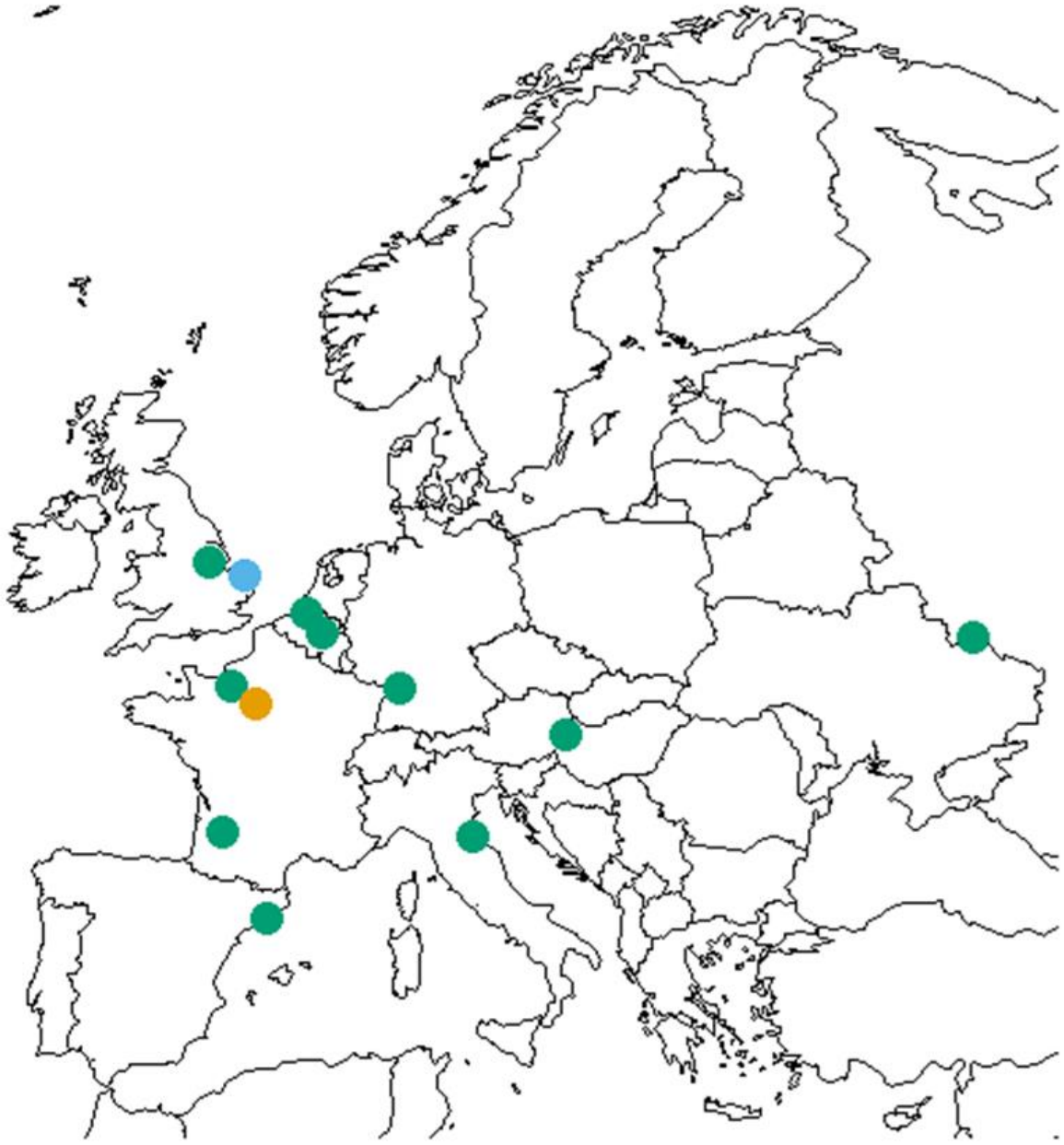


Figure 6.2.2.1 Locations of collection sites from SESVanderHave and Florimond Desprez around Europe. A map of Europe with circles indicates ten field stations where aphids were collected from the 20m^2 plots with four crop species of SESVanderHave (green), Florimond Desprez (orange) and John Innes Centre (blue). Aphid were also collected in the areas around these field stations.

At all collection sites from SESVanderHave, capture plots were used in 2018 and 2019, with the exception of Russia. Furthermore, a similar capture plot was also grown at John Innes Centre's field station in Bawburgh in 2019. I designed plots of 20m^2 with blocks of 5m^2 with each either *Brassica rapa*, *Brassica napus*, *Solanum tuberosum* or *Beta vulgaris* plants grown (figure 6.2.2.2). Yellow insect sticky traps were used at the capture plots to indicate

the abundance of alatae aphids around the season. The capture plots were monitored for aphid abundance every 1 to 2 weeks from April until August. Three single apterous aphids were collected from the host plants in separate silica gel containing collection tubes, stored/shipped on ambient temperature, and stored at -70°C after arrival in the John Innes Centre. Additionally, aphids were collected from fields and weedy plants near the capture plots.

Figure 6.2.2.2 Aphid capture plots at SESVanderHave field stations. Capture plots of 20m^2 with blocks of 5m^2 with each *Brassica rapa*, *Brassica napus*, *Solanum tuberosum* or *Beta vulgaris* plants were grown in rows next to each other. The capture plots were sown/transplanted within the field in early March. From April up to August, the capture plots were inspected for apterous aphids on a bi-weekly base. Three tubes of single aphids of the apterous aphid colonies were collected from colonies found on the plants. Additionally, yellow insect sticky traps were used for an abundance of alatae aphids around the capture plots.

6.2.3 Aphid species verification with mitochondrial gene COI

6.2.3.1 species verification using BLASTn

Prior to genome sequencing, I first carried out species-level identification of samples based on Sanger sequencing of the conserved mitochondrial gene encoding COI. As I used universal arthropod primers (Folmer et al., 1994) for the PCR amplification of COI before WGS, this process also allowed me to identify and discard samples that were likely contaminated with parasitic wasp DNA. After amplification, I sanger sequenced the amplicons. These reads were searched against the NCBI non-redundant (nr) nucleotide database ((MD), 1988) using BLASTn with default settings. The species with the top hit of taxonomic annotation that had the most complete and least polymorphism was identified as the organisms' species. Therefore, predictions of aphids' species were made on the quality of the taxonomic annotation.

I found that most samples were not *M. persicae* based on the top BLAST hit. Most samples were found to be *A. fabae*, other aphid species, or inoculated by a parasitoid wasp (contaminated). Overall, I screened 547 samples, from which I successfully received the COI sequences of 464 samples. Around a quarter of the successfully sequenced samples were found to be *M. persicae*, and half of the samples were other aphid species such as *Hyalopterus pruni*, *Aphis fabae, Macrosiphum euphorbiae* and *Phorodon humuli* (figure 6.2.3.1). Additionally, 56 (11%) samples were identified as a parasitoid wasp. A summary of the blast results for each sample is shown in appendix (table S2).

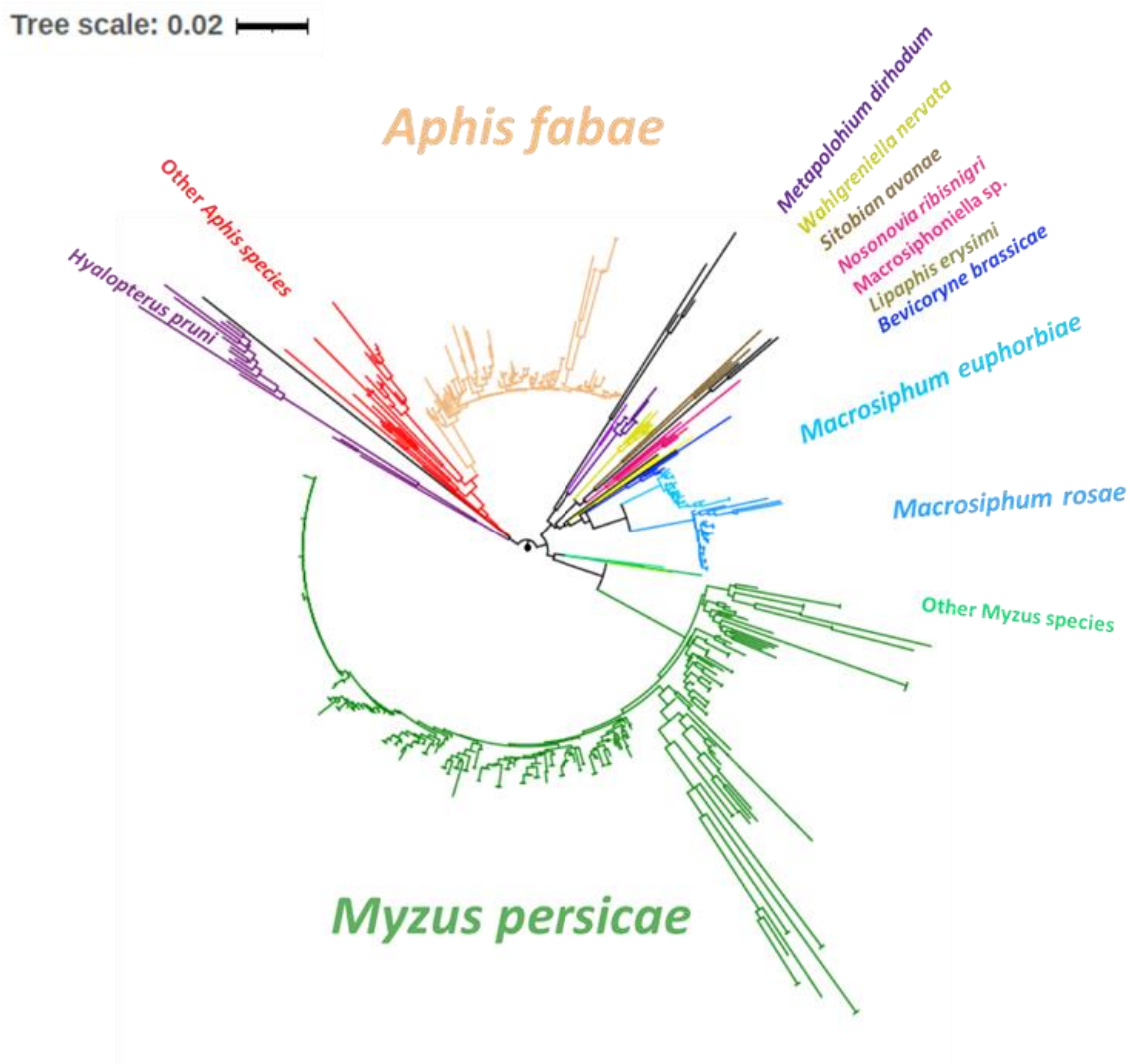


Figure 6.2.3.1 Neighbour-joining phylogenetic tree constructed COI sequences used for species verification. Neighbour-joining phylogenetic tree illustrated via the web-based programme the interactive Tree Of Life tree (ITOL) (Letunic and Bork, 2019). Species names written besides the tree branches identified as the species.

6.2.3.2 species verification using maximum likelihood phylogeny

To confirm the taxonomic assignment of samples, I constructed a maximum likelihood phylogeny of the COI sequences and selected outgroup taxa from Genbank. The outgroup taxa were composed of data derived from *Myzus ligustri, Myzus cerasi, Phorodon humuli, Aphis fabae* and *Aphis craccivora*. This analysis revealed that some samples that were initially identified as *M. persicae* via BLASTn were not correctly classified. After I aligned and constructed a maximum-likelihood phylogenetic tree, these samples grouped more distantly away from *M. persicae* in a phylogenetic tree than *M. ligustri* (figure 6.3.2.2). Which indicates that they are more distinct than *M. ligustri* (the closest known relative of M. persicae). To rectify this, I only retained samples that formed a monophyletic group with known *M.*

persicae samples and discarded those that cluster with *M. ligustri* or other outgroup taxa. Therefore, after screening, based on the phylogenetic analysis, including the non-COI verified lab-collected samples, I retained 119 samples for Illumina WGS.

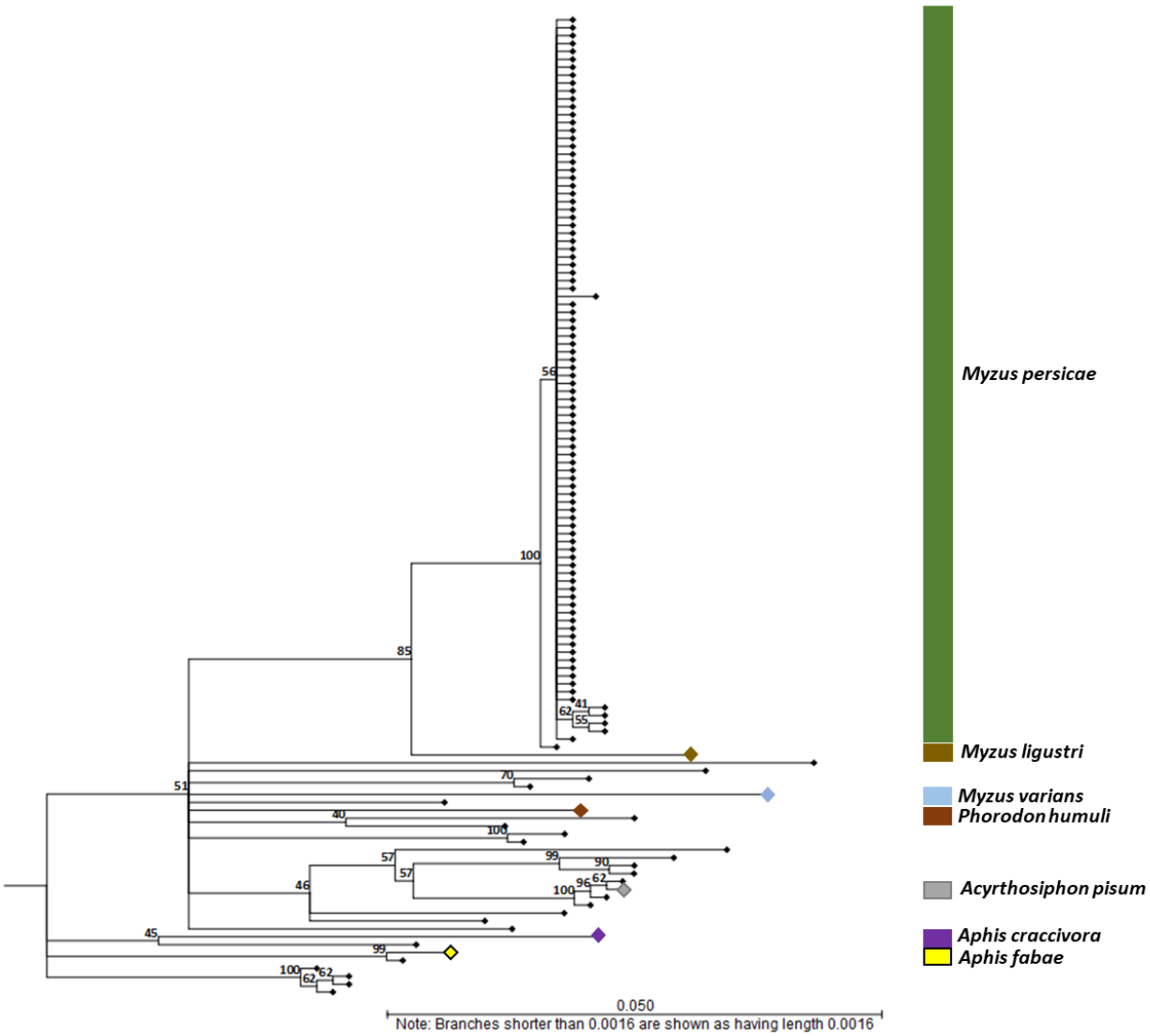


Figure 6.3.2.2 Maximum-likelihood phylogenetic tree constructed using COI sequences obtained from *M. persicae* and verified in comparison to closely related species. A subset of sequences checked for aphid species with phylogenetic analysis. I used one thousand bootstraps and collapsed nodes with less than 40% bootstrap support.

6.2.3.3 Pairwise divergence among M. persicae

To gain an initial overview of the level of genetic diversity found within *M. persicae* across my collection sites, I examined COI sequence divergence among *M. persicae* samples and compared this to levels of COI divergence in samples identified as *Aphis fabae* - another widely distributed generalist aphid (Blackman and Eastop, 2000b) that was co-collected with *M. persicae*. The

average intra-specific pairwise divergence of 0.2% (Foottit et al., 2008) *M. persicae* samples was below 0.01%. In contrast, an average pairwise divergence between 36 *A. fabae* samples was 1.62% (figure 6.3.2.3). This shows that the average intra-species divergence of *M. persicae* was found to be significantly lower than within *A. fabae*. This indicates that there is less diversity found within *M. persicae* in comparison to *A. fabae*. In both *M. persicae* and *A. fabae*, maximal intra-species divergence was substantially lower than inter-species divergence with close relatives identified based on phylogenetic analysis of COI (Figure 6.2.3.1). Divergence between *M. persicae* and *M. ligustri* was 3.74%, and divergence between *A. fabae* and *A. gossypii* was 7.41%. As such, samples I have identified as *M. persicae* and carried forward for WGS are likely not biased by the inclusion of divergent (mis-identified) cryptic species.

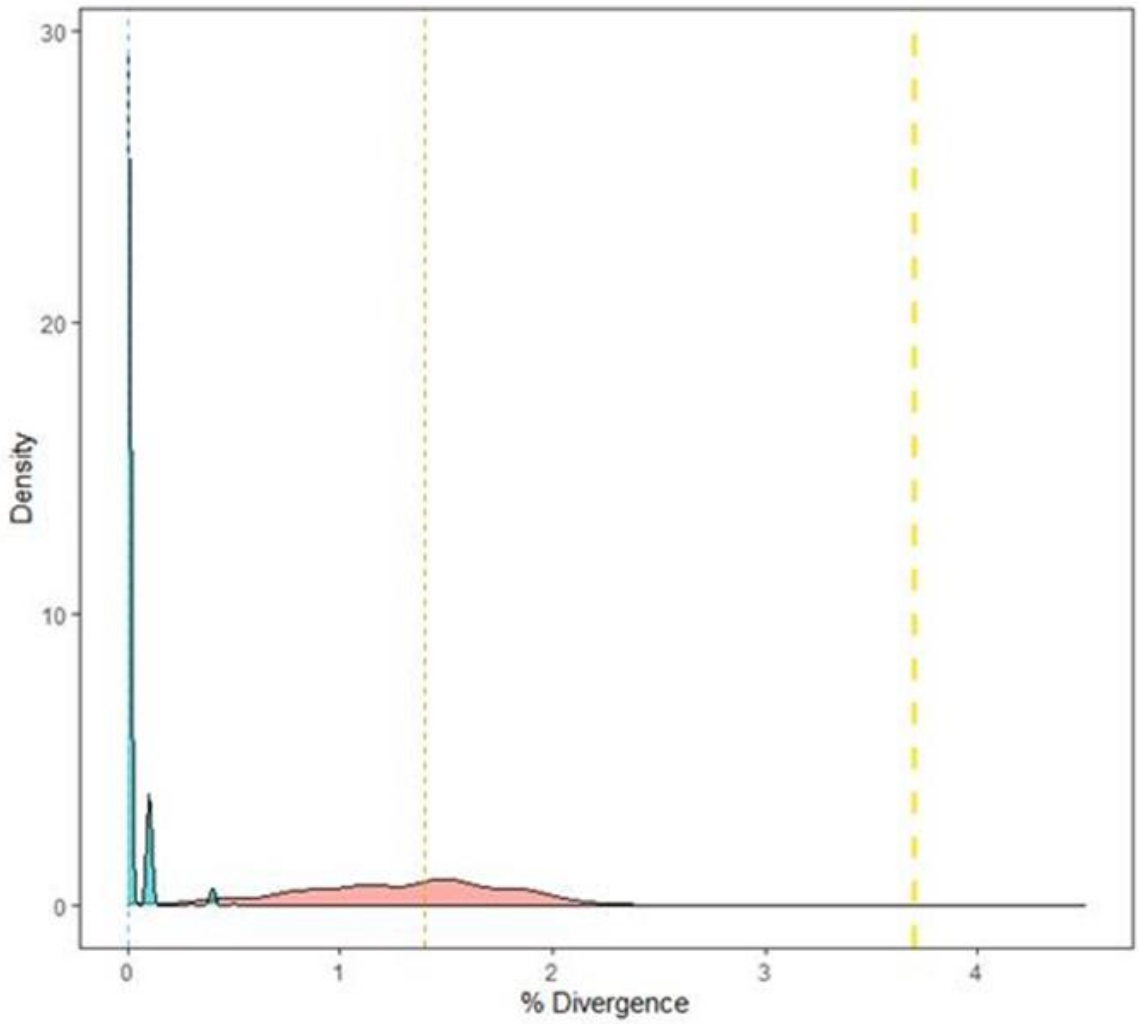


Figure 6.3.2.3 Pairwise divergence of COI sequences from *M. persicae* and *A. fabae*. *M. persicae* has lower intra-species divergence than *A.* fabae. Blue field = *M. persicae*, red field = *A. fabae*. Dotted lines in coloured fields indicate medians. Blue line = *M. persicae*, red line

= *A. fabae*. The interrupted yellow line shows the inter-species divergence between *M. persicae* and its closest relative, *M. ligustri*.

6.2.3.4 Conclusion

DNA was extracted of 547 aphid samples collected around the world and sent for Sanger sequencing of the COI sequence. In total 464 sequences were of enough quality to find best-BLAST hits for the aphid or parasitic wasp species. Around a quarter of these samples were found to be *M. persicae*, while over half of the samples found to be other aphid species and 11% parasitic wasp. Nevertheless, a few samples that were identified as *M. persicae* samples using BLASTn were found other species when plotted on a maximum-likelihood phylogenetic tree with other known aphid species obtained from NCBI. Therefore overall, 119 identified samples were taken forward for divergence analysis to find the presence of cryptic species complexes. The intra-specific pairwise divergence was used as the last verification step that the COI sequence from 119 *M. persicae* presumed samples was indeed the same species. I found that the intra-specific divergence of 0.01% was lower than in comparison to *A. fabae* with an intra-specific divergence of 1.62%. Additionally, the inter-specific divergence between *M. persicae* and the closest know relative *M. ligustri* was 3.74%. Therefore, I verified that the 119 samples were highly likely to be *M.* persicae, without the chance of bias by the presence of cryptic species complexes and were taken forward for WGS.

6.3 Discussion

In this chapter, I adjusted the protocol for DNA extraction from small insect samples (Mugford et al., 2020), established a sample storage system of field-collected insect samples for genomics (Wouters et al., 2020b), developed a collection strategy for ten different locations within Europe and identified and selected over 100 samples for WGS to obtain an initial assessment of *M. persicae* global diversity. Taken together, these advances provide a platform to conduct high throughput genome-wide diversity analysis on small insects.

The modified storage vessels from (Maddison, 2013) have developed an opportunity to obtain samples from around the world by reducing logistic problems (Wouters et al., 2020b). These storage vessels proved to maintain a sufficient amount of genomic DNA within the dried aphids after storage at ambient temperature for over eight weeks. In contrast, samples stored without silica gel had lost practically all genomic DNA after three weeks of storage at room temperature. Silica gel removes water from the samples that will help to retain the DNA within a stable environment for storage at room temperature. Because of the thinner exoskeleton of aphids compared to beetles, as the tubes from Maddison (2013) were developed for beetles, additional protection of the aphid was needed within the tubes. The filtered pipette tip within the middle of the storage tube protects the aphid from crushing by the silica gel within the tube. The separation between the silica gel and the insect did also reduce the chances of contamination during the DNA extractions. The silica gel storage tubes give opportunities as a cheap method to collect samples for WGS from rural areas without needing to freeze samples or use chemicals such as ethanol. No exceptional packaging or declaration is needed for dried samples, and there is no risk of thawing or damage because of temperature fluctuations. They can be used on a large scale, and after sample collection can be stored for over 8-weeks, which gave an abundance of time to collect the sample in rural fields around the field and shipment to the research lab in the United Kingdom. In some cases, I successfully extracted genomic DNA for WGS of samples stored for over a year at room temperature. This storage method was successfully implied in big communal country-wide collection projects such as the BRIGIT project (https://www.jic.ac.uk/brigit/).

Ample genomic DNA from a single dried aphid was obtained for Illumina 150PE PCR-based WGS. The lysis buffer used in the adjusted protocol from Marzachì et al. (1999) worked out to be the most efficient for DNA extraction

of aphid samples, and by removal of a centrifuge and transfer step, it increased the overall yield from a single dried aphid. DNA extraction of small dried insect samples is a growing interest with the improvements of next-generation sequencing techniques. Dried insect samples maintained in museums are of greater interest to study within population genetics studies to understand the mechanism and features of rapid evolution (Suarez and Tsutsui, 2004, Pelissie et al., 2018). The optimisation of a DNA extraction protocol for dried insect samples could be applied within dried insect samples: (i) collected from around the world from rural places and places with a lack of facilities for freezing samples. (ii) stored for a long duration at room temperature, such as samples from museums. (iii) obtaining more DNA yield from fresh insect samples. Therefore, this optimised protocol could improve yield in fresh samples and generate an opportunity to recover enough yield from small dried material for WGS.

In the collection initiative to collect *Myzus persicae* from crops and trees, most collected samples were either colonised by parasitoid wasps or a different aphid species. Communal collection by field workers of samples did turn out to include a high number of samples from other aphid species. *M. persicae* abundance is most elevated in springtime and reduced over the summertime within the UK (The Rothamsted Insect Survey, https://insectsurvey.com/archive0). Temperature affects aphids and other insects survivability (Bale, 1987, Sheppard et al., 2015) *Myzus persicae* optimal temperature for reproduction is around 25°C, with fluctuation depending on the generational acclimation (Alford et al., 2012). The collection years 2017, 2018 and 2019 had mild winters and relatively warm and dry summers. These conditions might have affected the *M. persicae* population densities of regions within Europe, as temperatures were exceeding above 25°C from the offset in early spring. According to field workers who checked fields weekly for aphids, *M. persicae* were found at lower abundances on crop plants in summer compared to springtime, while *A. fabae* was less abundant in the spring and gained greater population densities in the summertime. These differences in *M. persicae* and *A. fabae* abundances in spring versus summer are also apparent from the suction trap data of the Rothamsted insect survey (https://insectsurvey.com/archive0). Altogether, over 450 samples needed to be screened to obtain 119 *M. persicae* samples to take forward for whole genome resequencing by Illumina.

Most of the samples collected from sugar beet were found to be species other than *M. persicae*. In total, *Macrosiphon euphorbiae* and *A. fabae* were

present in high numbers on sugar beet during the growing season. Nonetheless, *M. persicae* is seen as the most important aphid pest in sugar beet due to its ability to transmit most sugar beet yellows viruses (Watson, 1940). When the sugar beet crop is infected with yellows viruses early in the season, the farmer can lose up to 49% of the yield, rendering a loss of profit (Smith and Hallsworth, 1990). Because *M. persicae* is found on sugar beet predominantly early in the growing season, this aphid is primarily responsible for the introduction of beet yellows viruses into sugar beet fields. However, *A. fabae* can also transmit viruses and may be responsible for the further spread of the beet yellows viruses in the sugar beet fields during summer *(Karasev, 2000)*. This spreading over the entire field gave rise to higher infection within fields. The dynamic of aphid vectors of viruses is of great importance to understand the prevention of sugar beet viruses. Previously, sugar beets were sown with a neonicotinoid (thiamethoxam, clothianidin or Imidacloprid) within the seed coat, which protected the sugar beet the first 12 weeks of the growing season (Dewar et al., 1996). These 12 weeks of protection against insects was sufficient to prevent damaging yield losses. Thus, though the low abundance of *M. persicae* found in the sugar beet field over the season, their early colonisation in the growing season and proficiency of vectoring of the virus makes *M. persicae* the most significant insect pest in sugar beet.

Sanger sequencing of the COI gene was used to verify the aphid species of the collected samples. Most samples checked were found not to be *M. persicae*, although enough samples were obtained for the population genomics study. Nevertheless, twenty initially presumed *M. persicae* samples were used for WGS after verification via BLASTn were found to have either contamination with other insects/pathogens or were found to be different aphid species (further described in chapter 7). Some *M. persicae* COI gene Sanger sequences found in NCBI were incorrectly specified. Via phylogeny, I showed that these twenty samples are genetically more diverse than the closest relative *M. ligustri* and therefore, falsely identified as *M. persicae.*

With an intra-species divergence of lower than 0.01% in *M persicae*, this seems to be low compared to most other aphids such as *A. fabae (Foottit et al., 2008)*. Additionally, Foottit et al. (2008) had found similar low divergence for *M. persicae* but found a COI sequence divergence of 0.16% for *A. fabae* when they compared ten samples. The divergence is lower than the *A. fabae* divergence I found (1.62%). This high divergence in *A. fabae* could be because of the presence of multiple cryptic taxa within the current species within *A. fabae* (Gauffre and D'Acier, 2006), as shown in other aphid species

Nippolachnus piri (Kanturski et al., 2018). The low divergence in *M. persicae* could have multiple explanations, such as (1.) a recent expansion after a bottleneck (possibly due to extensive use of insecticides within fields, e.g. neonicotinoids), (2.) No physical barriers due to human interference, (3.) sample size or sequence region were too small to investigate genetic divergence. Therefore, to examine these explanations, further investigation of population genetics is needed on a genomic scale to understand the diversity of *M. persicae*.

Chapter 7. Population genomic analysis of *Myzus persicae*.

7.1 Introduction

Insect pests pose a substantial threat to agriculture through direct feeding damage and because of their role as vectors of plant disease agents. As such, considerable effort has been put into control of these pests. This has been primarily achieved through insecticides and the breeding of resistant crop varieties (Sun et al., 2018) and the use of genetically modified crops (Yu et al., 2014, Bates et al., 2005, Wang et al., 2018). However, many insect pests have large populations and can rapidly evolve to subvert these efforts. For example, *Helicoverpa zea* (Tabashnik et al., 2008) and *Plutella xylostella* (Baxter et al., 2005) have developed resistance mechanisms against *Bacillus thuringiensis* (Bt) toxin family single-crystalline 1 (Cry1). Another example is *Anopheles gambiae* that developed resistance mechanisms against pyrethroids and Dichlorodiphenyltrichloroethane (DDT) (Ranson et al., 2000, Chandre et al., 1999). Furthermore, in some cases, introgression (the transfer of genes between species) has hastened the spread of insecticide resistance, for instance a Bt resistance allele that was introgressed from *H. zea* into *H. amigera* (Valencia-Montoya et al., 2020). As such, control of insect pests remains challenging. Furthermore, in some cases, introgression has enabled pest species to colonise new plant host species, such as *Albugo candida* (McMullan et al., *2015),* multiple examples in the fungi order Hypoclealean (Zhang et al., 2018) and the aphid genus *Cinara* (Jousselin et al., 2013, Larson et al., 2019).

The need to combat rapidly evolving insect pests whilst minimising the environmental impact of control measures drives a need to build stable, long-lasting resistance into crops (Sandhu and Kang, 2017, Mundt, 2014). Stable resistance is a level of resistance that would be able to be long-lasting without the presence of resistance breaking biotypes. When plant resistance is introduced into the field and a biotype is found with a variant to evade the resistance, the resistance is broken, and this gives once more plant susceptibility to a *M. persicae* population. Therefore, to achieve stable plant resistance a thorough understanding of the diversity and the evolutionary potential of pest species is required.

The majority of insect pests are specialists that colonise one or a few plant species (Forister et al., 2015, Bernays and Graham, 1988). However, some species are generalists and able to colonise a wide range of plant species. These generalists are able to rapidly spread via colonizing multiple plant species and have the potential to introduce viruses from weeds into

crops. *M. persicae* is a key generalist and is a significant crop pest that colonises over 400 hosts, including the majority of crops, and transmits over 100 plant viruses (Blackman and Eastop, 2000b, Weber, 1985). Moreover, *M. persicae* is known to rapidly adapt to external factors, such as pesticides (Puinean et al., 2010, Bass et al., 2011, Margaritopoulos et al., 2021, Devonshire et al., 1998, Mingeot et al., 2021), temperature (Alford et al., 2012) and host swaps (Mathers et al., 2020a, Chen et al., 2020).

Although often referred to as a generalist, it is not yet clear whether *M. persicae* host range may include a genetic component. Recently, Chen et al., (2020) showed that a single *M. persicae* genotype is able to colonise nine different plant species from five families, indicating that *M. persicae* is a true generalist. Yet, differences were found in the ability of *M. persicae* to colonise specific plant species (Hong et al., 2019). In chapter 3, I showed that *M. persicae* genotypes have different performance on different hosts and that this difference could be genotype-specific within *M. persicae*. Furthermore, a cryptic tobacco-specialising sub-species has previously been described and defined as *M. persicae nicotianae* (Weber, 1985, Kati et al., 2014, Nikolakakis et al., 2003). Adaptation to tobacco has recently been shown to involve overcoming the tobacco chemical defences by a specific P450 gene (CYP6CY3) that has been duplicated in the *nicotianae* sub-species (Singh et al., 2020a) and does efficiently detoxify nicotine (Bass et al., 2014). This raises the possibility that *M. persicae* generalism may, at least in part, be due to the presence of cryptic host-associated lineages such as described in other aphid species. For example, in *A. pisum* it has been shown that multiple biotypes are found that colonise specific plant species, albeit the *A. pisum* host range at the species level is mostly restricted to legumes of the Fabaceae (Via, 1999, Via and Hawthorne, 2002, Ferrari et al., 2006, Ferrari et al., 2008, Peccoud et al., 2009).

Aphids, including *M. persicae*, have an unusual life-cycle with both sexual and asexual stages. Aphids mostly reproduce clonally/asexually in the summer on their secondary host, whereas they reproduce sexually once a year in the autumn and overwinter as eggs on their primary hosts (Blackman and Eastop, 2000b). *M. persicae* is abundant in many countries and reproduces sexually once a year on *Prunus* species (Blackman, 1974). *M. persicae* initiation to a sexual stage is dependent on temperature and day-length (Blackman, 1974). However, so far, the extent to which single clones of *M. persicae* can spread to different countries and to what extent these insects reproduce

sexually in these countries has remained unclear (Vorburger et al., 2003). The unusual life-cycle of aphids may influence patterns of their diversity and play roles in adaptation rates to new hosts or in their abilities to overcome specific pesticides or crop resistance genes. For example, sexual reproduction may facilitate the transfer of beneficial alleles into a new genomic background, and successful combinations may reach very high frequencies through clonal reproduction. Previous studies have shown that *M. persicae* populations can be facultatively (or in some cases are obligately) asexual in temperate regions such as in England (Fenton et al., 2010), France (Guillemaud et al., 2003) and Australia (Vorburger et al., 2003).

So far, a thorough understanding of *M. persicae* global diversity has been lacking. Previous *M. persicae* population genetics studies have used a small number of markers (low resolution) and have primarily focussed on narrow geographic regions or specific countries (Zhao et al., 2015, Margaritopoulos et al., 2009, Wilson et al., 2004, Wilson et al., 2003, Blackman et al., 2007, Turcotte et al., 2013, Popkin et al., 2016, Monti et al., 2016, Sloane et al., 2001, Ramsey et al., 2007). However, these studies have revealed that low diversity was found within and between countries in Europe and Australia, with some indication of genotypes or clonal populations that dominate the population named “super clones” (Vorburger et al., 2003, Fenton et al., 2010, Margaritopoulos et al., 2009, Vorburger, 2006). Fenton et al. (2010) have decribed the diversity and genetic distance of *M. persicae* in the United Kingdom over a period of 14 years (from 1995 up to 2009) using five different microsatellites. Interestingly, before 2000 three different genotypes were predominantly found. Yet, around 2006, a genotype (O) was discovered and dominated the population in the next few years, according to the “super clone” proposition (Margaritopoulos et al., 2009). Moreover, Margaritopoulos et al. (2009) investigated 215 genotypes from 14 countries from four continents with six microsatellites. They distinguished between three population clusters and revealed a low global differentiation (FST = 0.086). In contrast, higher differentiation was found between the tobacco-associated population from Europe and tobacco-free regions (FST=0.154) (Margaritopoulos et al., 2009). Additionally, a higher Fst was observed location-dependent between Australasia and European samples. Nevertheless, it remains under-investigated if *M. persicae* populations have host preferences and whether high genetic distances are present among populations based on plant host species preference or locations.

Although found on all continents, *M. persicae* is thought to originate from Asia (Waterhouse, 1993, Wieczorek and Chłond, 2019). Consistent with this, Asian populations have higher diversity than the European and Australian populations (Li et al., 2015a, Zhao et al., 2015). Li et al. (2015a) described pairwise genetic differentiation estimates FST of 0.1215 among populations collected from many regions around China. Moreover, three population clusters were predicted within the samples collected in China. Yet, another study described a high degree of genetic differentiation (FST =0.335) when *M. persicae nicotianae* genotypes were collected around similar areas in China. Only two *M. persicae* clusters were predicted within this subset (Zhao et al., 2015). Nevertheless, they indicated that they found that the five microsatellites were significantly linked and had heterozygosity excess. This may possibly be a result solely from asexual reproduction of one of the predicted clusters.

Here, I used whole-genome re-sequencing to investigate *M. persicae* diversity globally, and particularly in Europe. I investigated hypotheses relating to how *M. persicae* populations may be structured. These are that populations could be structured in four potential manners: **1.** per region, **2.** per plant host species, **3.** per plant host species within specific regions, or **4.** no clear clustering by plant host species or region (Figure 7.1.1). Additionally, given that I have generated high-resolution genome data for a large collection of *M. persicae* samples, I can test the hypothesis that genomic areas have been affected by recent positive selection pressures among *M. persicae* populations. Taken together, these analyses will provide a detailed overview of *M. persicae* diversity that, in the future, may be used to assist the development of stable crop resistance to this important pest species.

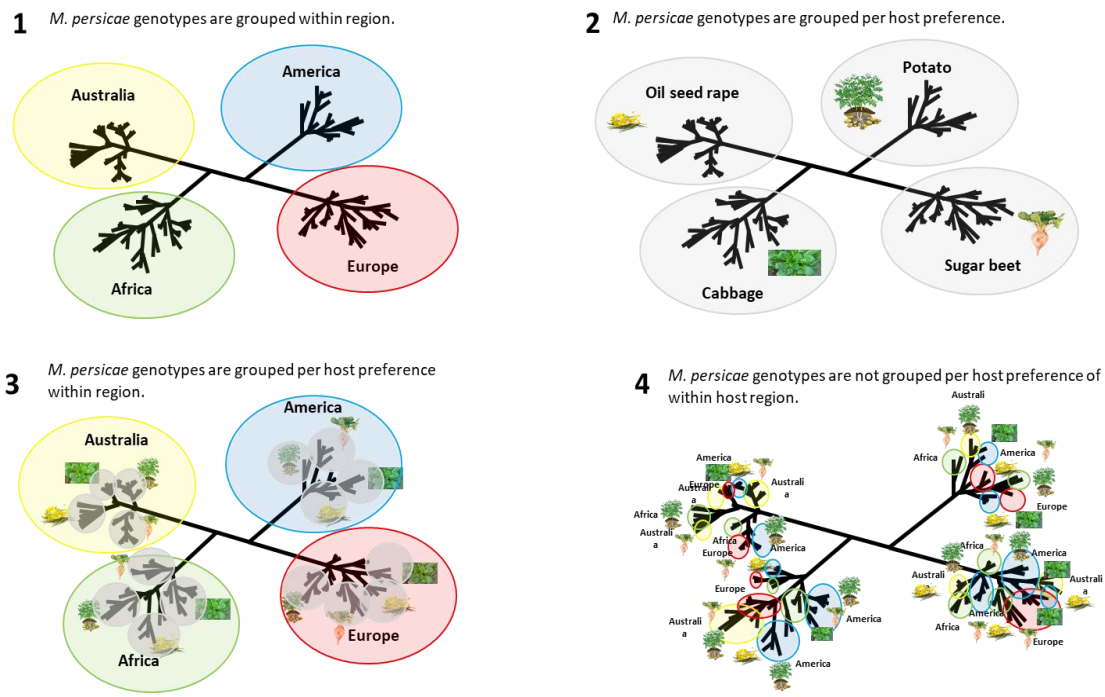


Figure 7.1.1 Hypothesis for population dynamics. Four scenarios are possible in the population dynamics of *M. persicae*. In scenario **1**, the genetic diversity and population structure of the generalist aphid species *M. persicae* is explained on a regional base. In scenario **2**, it is clarified by the plant host. In scenario **3**, it is described by both the region and the plant host. Finally, in scenario **4**, the genetic diversity of *M. persicae* is not explained on either host or region.

7.2 Results

7.2.1 Whole-genome re-sequencing

To investigate global *M. persicae* diversity and population structures, I selected 119 samples for whole-genome sequencing. Samples were collected from fourteen countries and six continents. For each sample, I generated between 8 Gb and 40 Gb of sequence data. After trimming for sequencing adapters and low-quality bases (example found infigure 7.2.1.1 A and B), I retained between 5 and 37 Gb of data per sample after this filtering. I subsampled all samples to between 5 Gb to 8 Gb of sequence data. All samples had sufficient quantities of data, at approximately 10x to 20x estimated coverage based on the *M. persicae* genome assembly size of 395Mb (Mathers et al., 2020b), to be carried forward for downstream quality control.

To ensure a reliable set of samples to carry out population genomic analysis, I developed a robust quality control pipeline that included sample analyes for coverage depths, GC proportions, taxonomic annotations at the genus level and sequence read proportions mapped. Based on these quality

control asssessments, I identified 18 samples that were likely either low-quality sequence data or suffered from contaminations. The steps taken to identify these 18 samples are decribed below.

As a first pass check for contamination, I investigated the GC content of the reads generated for each sample. The *M. persicae* genome has a GC content of around 27% (Mathers et al., 2020b) and was found in most samples (example given in figure 7.2.1.1 C). Any samples that deviate strongly from this are likely to be contaminated. For instance, some samples had two GC content peaks, one at 27% and another at 60% figure 7.2.1.1 D, whilst a few had only a single peak at 60% (figure 7.2.1.1 E). Therefore, some samples were flagged as suspicious, but further analysis was needed before I opted to discard these samples.

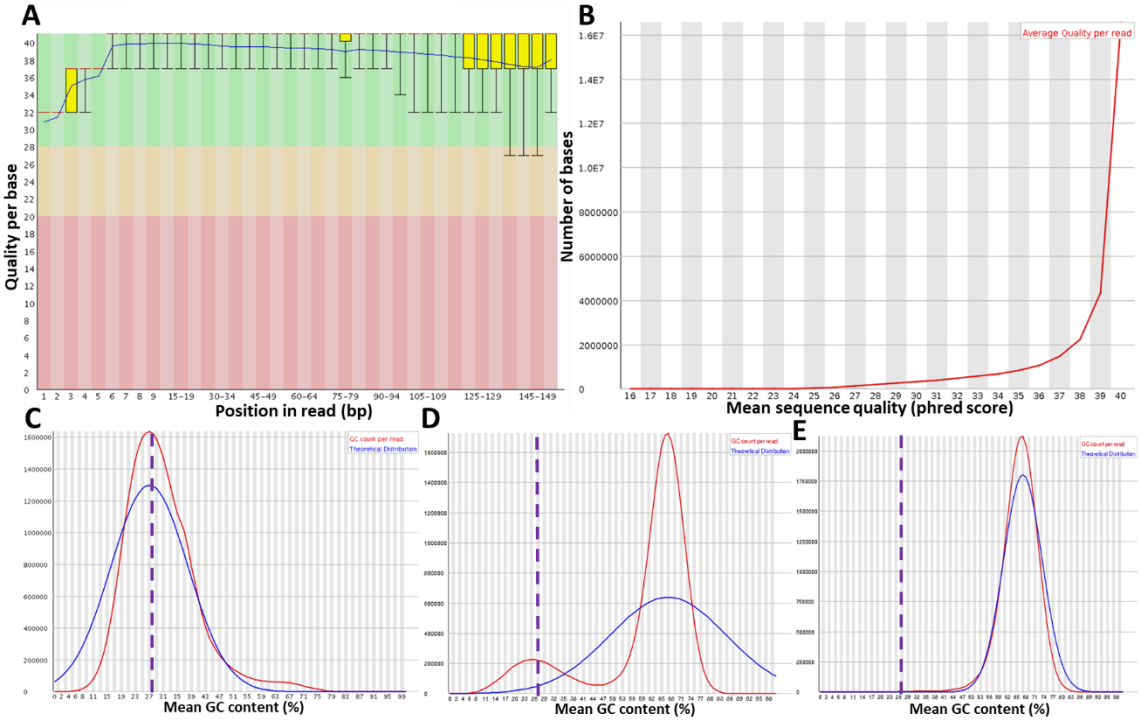


Figure 7.2.1.1 Trimming report and GC content analysis. The quality of the sequence data after trimming was analysed using Trimgalore FastQC. The quality of the sequence reads was checked via the quality report using (**A**) quality per base, (**B**) Overall mean sequence quality and (**C**, **D**, **E**) proportion of GC content. (**A**) The quality in a read per base was checked for quality above 30 (99.9% accurate). Yellow bars indicate quality score per base pair with standard deviation. The green field showed that the quality score is as wanted above 30, while orange indicates a poor-quality score between 20 and 30, and red indicates a failed sequence with a quality score below 20 (Less than 99% accurate). (**B**) Mean quality per read was on average Q38 for all bases, which indicates an average accuracy of more than 99.9% per base was found in the sequences. (**C, D, E**) Mean GC content percentage of the genome sequences. The red line shows the GC percentage per read. The blue line indicates a theoretical distribution (a normal distribution from the average). The interrupted purple line shows the expected average GC count for *M. persicae* (27%). (**C**) Shows an expected GC count distribution with an average of 27%, (**D**) represent a

'suspicious' sample with most of the reads having a GC content of around 60%, but some reads show similar GC content as expected, (**E**) Had a sharp peak at around 60%, which is more than double than the expected GC content per read within *M. persicae*.

To identify samples that are likely affected by contamination, I used the *de novo* BlobTools workflow (Laetsch et al., 2017). I generated a *de novo* assembly for each sample and used Blobltools to calculate and plot the average coverage vs average GC content of each contig in the *de novo* assembly. I also blasted each contig against the NCBI database and used Blobtools to assign taxonomy information based on the blast hits. Therefore, this tool allowed quick screening of quality by assessing coverage depth, GC proportion, taxonomic annotation on the genus level and the proportion mapped.

Aphids have one primary symbiont, *Buchnera aphidicola*, that is found in most aphids, including *M. persicae* (Xu et al., 2020). This endosymbiont provides essential amino acids and vitamins for its hosts and is necessary for aphid survival (Douglas, 1998, Nakabachi and Ishikawa, 1999). Because of the high abundance of *B. aphidicola* within an aphid, read coverage of this symbiont genome is typically found double the amount to that of the aphid (e.g. figure 7.2.1.2a). When multiple aphids are present, it is likely that the assembly will have contigs from two or more *Buchnera* strains, and these will be identified by BlobTools. I observed the presence of two different *Buchnera* symbionts within one of the aphid samples of half the coverage (figure 7.2.1.3a).

Both the GC proportion and the percentage mapped were similar in the majority of the samples. The average GC proportion within the genomes was found around 27 % (example is given in figure 7.2.1.2a), and over 80 % was mappable to their own reads (figure 7.2.1.2b, c). Twenty-one samples were found to either have a higher GC proportion or a lower mappability than 80 %. Additionally, in the 21 samples, less than 50% of the reads were identified as *Myzus* at the taxonomic annotation (example given in figure 7.2.1.3b, c). Consequently, 21 of the WGS samples were judged to be contaminated or had poor mappability. This could mean that some WGS were not *M. persicae* but other aphid species. In order to ensure that all these samples are *M. persicae*, I verified the aphid species using the whole genome sequences (appendix table S3). Therefore, these samples were classified as contaminated and not used in further population genomics analyses.

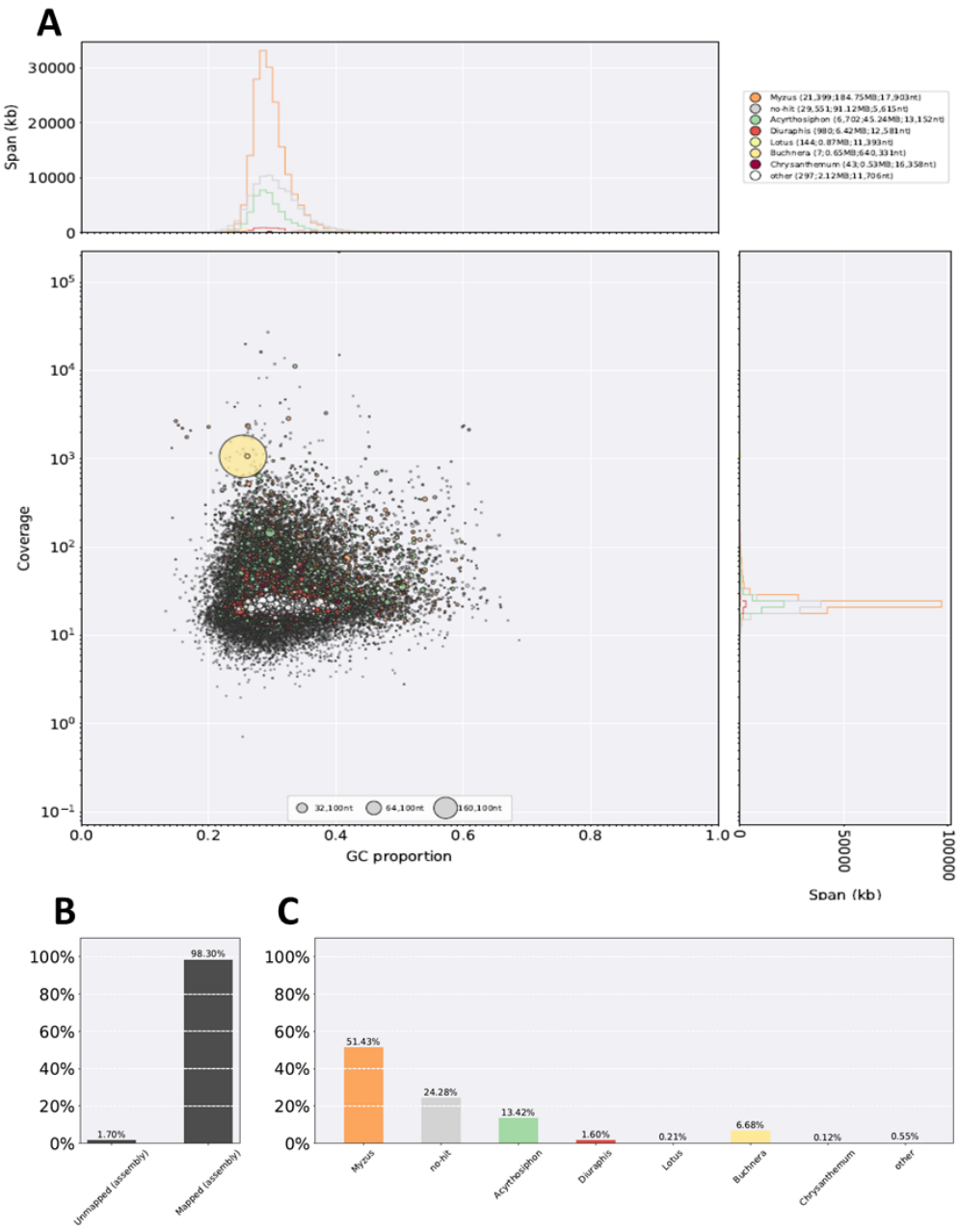


Figure 7.2.1.2 *de novo* assembly of re-sequenced aphid samples for verification of *M. persicae* samples without contamination (example of pure *M. persicae* sample without contamination). Example of BlobTools output of *de novo* assembly and best-BLAST hits from contigs from a pure *M. persicae* sample 4106a, without contamination. (**A**) 3- part plot illustrating the coverage, GC proportion, and taxonomic annotation of assembly sequence reads. In the central, a Blobplot is shown with circles sizes shown proportional to the sequence length. The colours indicate the taxonomic annotation on the genus level, as shown in the legend on the right top. Circles are placed based on their GC proportion (X-axes) and on the sum of coverage across both libraries of read 1 and read 2. (**B**) Percentage of reads mapped in assembly. (**C**) The proportion of mapped reads allocated according to taxonomic annotation.

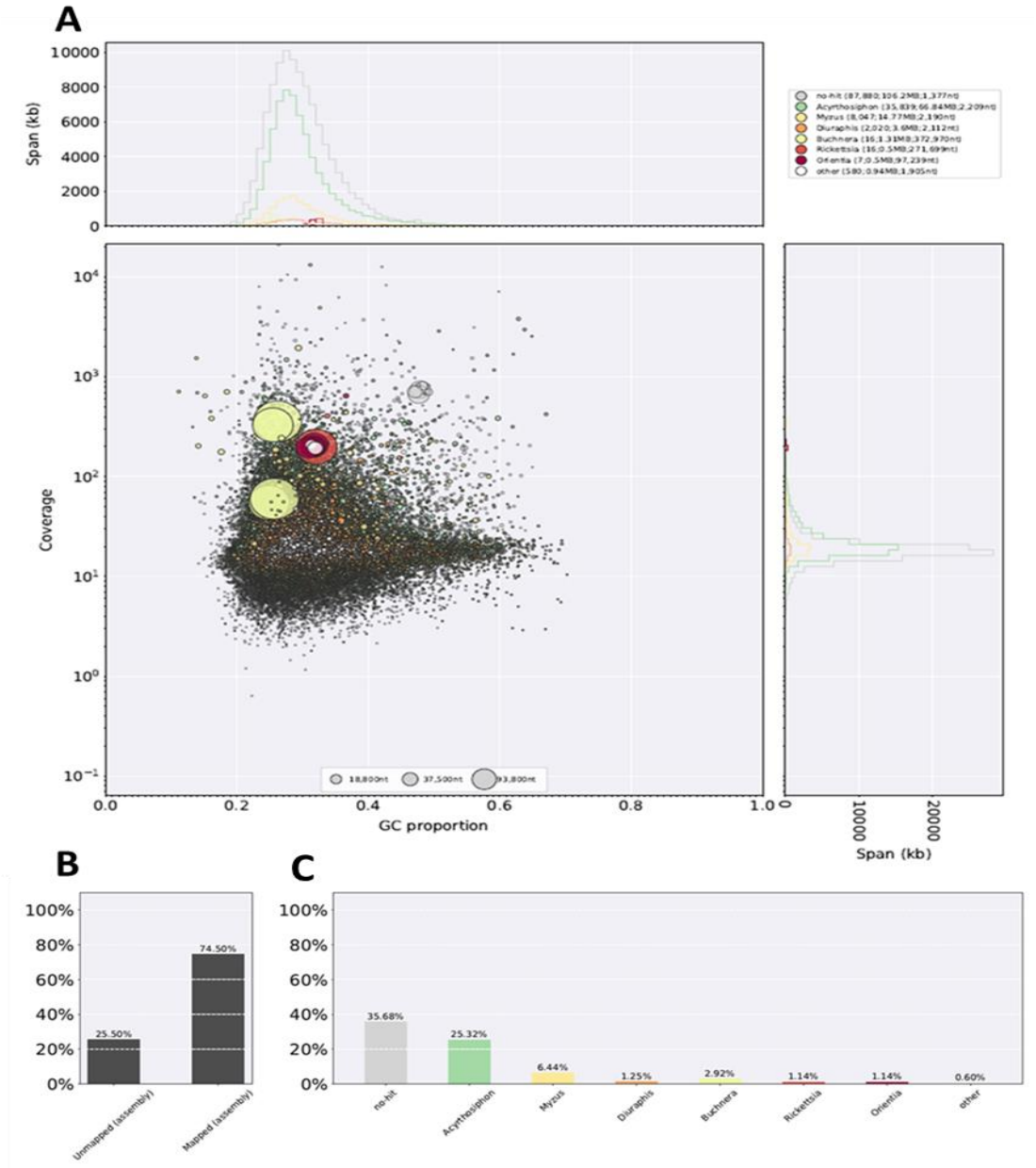


Figure 7.2.1.3 *de novo* assembly of re-sequenced aphid samples to verify *M. persicae* samples without contamination (example contaminated sample). Example of BlobTools output of *de novo* assembly and best-BLAST hits from contigs from HUN12, a contaminated sample. (**A**) 3- part plot illustrating the coverage, GC proportion and taxonomic annotation of assembly sequence read. In the central a Blobplot is shown with circles sizes shown proportional to the sequence length—colours indicating according to the taxonomic annotation on genus level, as shown in the legend on the right top. Circles are placed based on their GC proportion (X-axes) and on the sum of coverage across both libraries read 1 and read 2. (**B**) Percentage of reads mapped in assembly. (**C**) The proportion of mapped reads allocated according to taxonomic annotation.

To conclude, in total, 89 of the 119 sequenced genomes were taken forward to analyse *M*. *persicae* diversity. Twenty-seven WGS were discarded due to contamination or verified as a non-*M. persicae* species. These samples

were verified as contaminated using analytical tools such as BlobTools and based on the *de novo* assemblies of the sequences and the taxonomic annotation tool Mash screen for verification of the species. In addition to the 87 samples, 12 *M. persicae* WGS were obtained from another study (Singh et al., 2020) and two samples from research labs in the Netherlands. Thus, the 101 samples that are taken forward are likely to be high quality and free from contamination.

7.2.2: Variant calling and filtration

WGS reads of the 101 samples were mapped to the chromosome-scale *M. persicae* clone O reference genome (Mathers et al., 2020b) and variants called using BCFTools Mpileup. In total, 11.53 million variants were identified. To obtain a high-quality nucleotide polymorphisms (SNP) set, I selected variants that had a quality of above 30 (Q30), only biallelic SNPs, no indels, and total depths per sample of above 4 per site. This retained 8.94 million SNPs. However, many of these SNPs were only found in a single individual (private allele) either as a heterozygous (singleton) or a homozygous (private doubleton) site (Table 7.2.2.1). Strikingly, 6.05 million variants (68% of the total filtered SNP set) were derived from only two individuals – FRC09 (from France) and C25 (from China). Therefore, it is likely that these two samples represent closely related (sub)species or are possibly hybrids. Of these two possibilities, the latter is most likely as both samples have high heterozygosity – C25 had a heterozygosity of 63% and FRC09 25%. I, therefore, discarded FRC09 and C25 from further analysis to assess *M. persicae* diversity. Additionally, fifteen other samples were found to have over 5000 unique SNPs either as heterozygotic or homozygotic specific to the genotype. Nevertheless, these SNPs could represent true genotype-specific variances within *M. persicae* genotypes and so I retained these samples. Three samples were found to have a missing rate of more than 10 per cent of the filtered SNPs (figure 7.2.2.1 A); these were MISC47 (54% missing SNPs), A156 (24%) and A151 (15%). These three samples were kept for further analysis, but it was noted that a high percentage of the bases were not mapped to the reference genome. This left 99 samples in my analysis, all of which had (low) proportion of heterozygous sites (Figure 7.2.2.1). Following the removal of FR09 and C25, I re-filtered the SNP set to retain variable sites among the remaining 99 samples. This produced a final set of 2.80 million SNPs.

Table 7.2.2.1 Variances found specific to genotypes. The green marking indicates the samples with high missing rates, as shown in figure 7.2.2.1. The dark red marking indicates the samples with high heterozygosity proportion, as shown in figure 7.2.2.1. Additionally,

both of these samples had over a million Singleton/doubleton. The orange marked samples had either a genotype-specific variance or homozygous variance specific to the genotype of over 5000. The blue marking indicates *nicotianae* subspecies received from Singh et al. (2020a).

Sample	Singletons/doubletons	Homozygous variance specific to genotype	all homozygous variance
4106a	932	538	15543
A102	4831	1471	54982
A138	291	510	94548
A151	127	1312	80473
A156	256	1579	46176
A161	349	691	95470
A166	10191	3682	55425
BE1	146	174	23325
BE23	120	57	7940
BE2_B2	128	154	23491
BE33A	94	93	18990
BE49	140	137	14577
BE6	201	162	47467
C25	4563137	8251	40127
Crespys	39750	8245	93106
ES01	210	152	43711
ES146	638	644	31563
ES149	367	661	30808
ES88	432	520	89366
ES92	302	498	89146
FR15	940	664	86552
FR433	30792	12714	108421
FRC	13887	2130	7988
FRC09	1486446	22703	114655
Generac	464	749	103769
I1	12298	9507	107802
K16	384	325	84771
K40	913	796	97617
K43	798	725	97497
K66	441	362	83043
K87	361	450	83149
K88	151	402	83921
Lierida	4811	2696	54285
MG1107	10629	12338	130362
MISC10	118	184	24584
MISC14	146	48	5513
MISC27	62	118	4512
MISC28	118	157	13514
MISC30	106	326	22483
MISC31	897	326	47184

MISC33	57	74	6882
MISC34	186	620	46877
MISC38	188	100	4520
MISC4	194	202	16701
MISC42	166	164	14958
MISC46	69	182	23501
MISC47	331	1353	12357
MISC48	95	328	30599
MISC5	147	186	14171
MISC51	138	248	47341
MISC53	101	201	23646
MISC56	47	80	5011
MISC60	137	36	3413
MISC64	165	196	37349
MISC65	121	229	9207
MISC67	97	56	4841
MISC69	87	115	14058
MISC76	118	251	47129
MISC78	69	51	3480
MISC79	50	83	3954
MISC81	6021	1192	25314
NIC410G	25559	10516	96179
NIC_23	12608	16710	147589
NIC_410R	34153	2725	15984
NIC_5191A	1563	1039	98675
NIC_57	14127	9896	103041
NIC_8124	2414	4155	106759
NIC_926B	12873	18975	132688
NL1	265	295	25742
NL21	202	243	24463
NL5	279	153	37241
NL93	566	942	25466
NL_IRS	29990	7992	86452
NL_WUR	287	2134	50564
O	1447	515	12315
Q1200	127	234	56769
S121	1742	858	48225
S152	10769	15278	146199
S17	559	789	88876
S196	163	353	77663
S204	107	264	30369
S2196G	4703	12111	128117
S232	144	344	77025
S2B01	4807	9269	119909
S472	514	364	76697
S481	450	575	29157

SUS_1X	15422	2322	10562
SUS_4106a	1284	426	55480
SUS_4255A	5209	2485	63417
SUS_NS	1796	615	76694
SUS_US1L	1202	315	58199
UK1	135	1352	29925
UK19	501	316	7929
UK2	201	321	50021
UK20	631	1287	24773
UK21	134	307	49655
UK25	129	218	23457
UKSB	740	50	5313
UKW3	289	1781	42080
US1L	700	117	7005
a456BE	10	211	22631

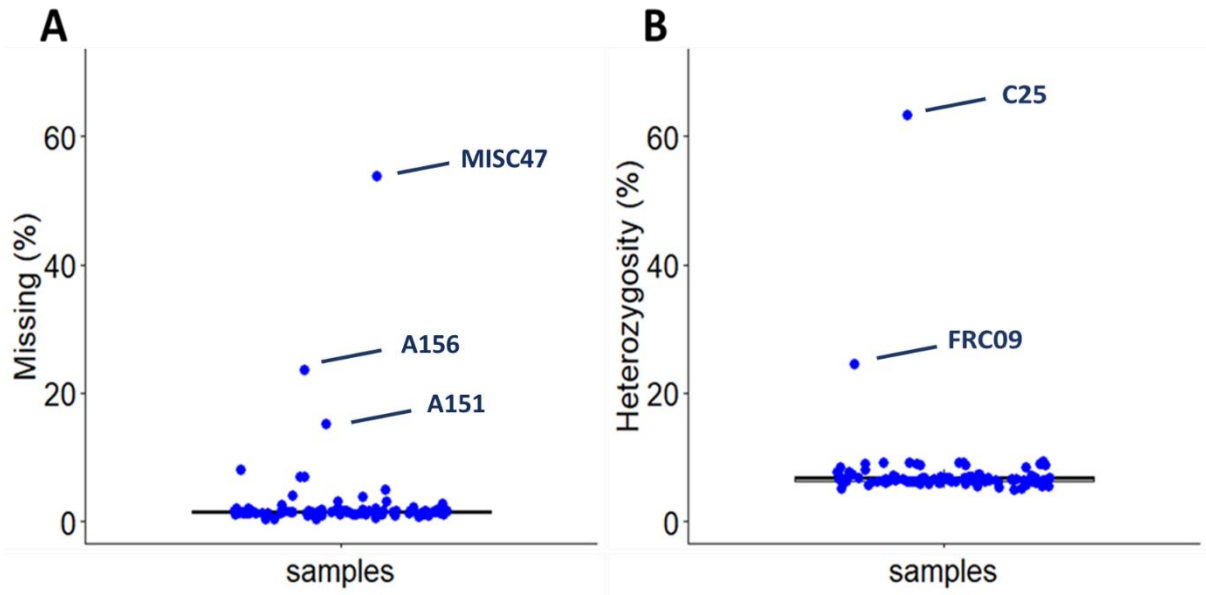


Figure 7.2.2.1 Missing rate and heterozygosity proportion of the filtered SNPs. Boxplot with scatterplot on top. Each dot represents a single WGS sample. Both (**A**) missing rate and (**B**) heterozygosity are shown in the percentage of the filtered variances file of the 101 WGS.

Finally, to assess the remaining variances between samples, I performed a similarity analysis on the 99 WGS *M. persicae* samples. Overall, high similarities were found among the WGS *M. persicae* samples (figure 7.2.2.2). All samples had at least >90% similarity, even when the VCF file was filtered for only variant sites within the genome. Some samples were very similar, for instance, clone O, US1L, UK_SB that had a similarity of over 97%. The dendrogram indicated that there are possibly multiple populations found

among the subset. Thus, high similarities were found among the samples, but differences among them were noticeable.

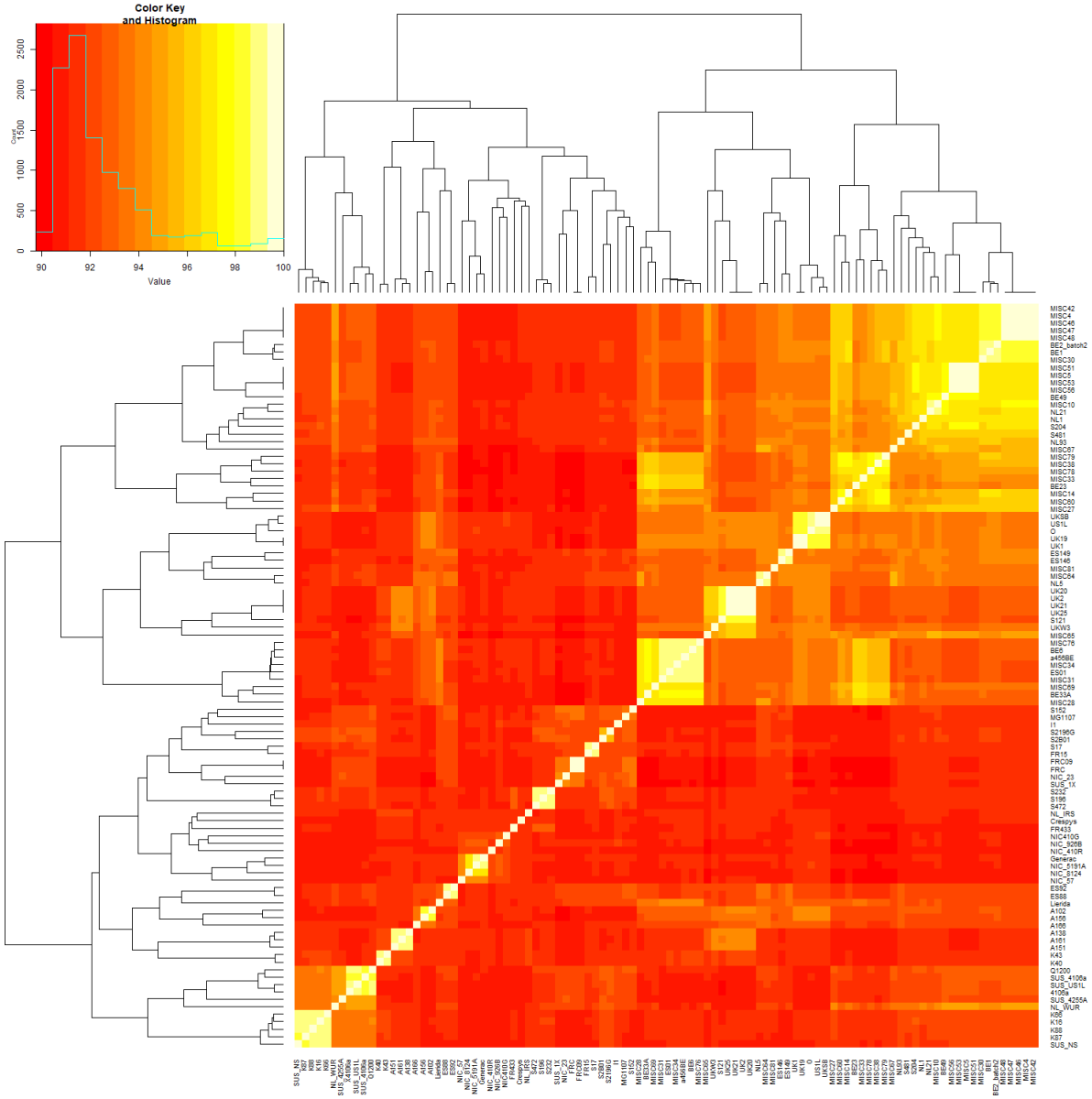


Figure 7.2.2.2 Similarity matrix of 99 WGS *M. persicae* samples. Colour indicate the similarity between the samples, red = low similarity, yellow = high similarity. The dendrogram on horizontal and vertical axes show the difference in the level of similarity of the samples. The similarity matrix is calculated from all used SNPs after removal of sample C25 and FRC09. filtering.

7.2.3 *M. persicae* population structure and genome-wide patterns of linkage disequilibrium

In order to understand if *M. persicae* populations structure according to plant host and/or based on geography, I performed principal component

analysis (PCA) using the SNPs identified in section 7.2.2 and all the 99 retained samples. Four principal components together explained 15.5% of the variation (PC1=4.2%, PC2=3.9%, PC3=3.8% and PC4=3.6%). This analysis revealed limited clustering by country or even by continent with samples that are separated by large geographic distances being close to each other in the PCA, for example the Argentinian vs European samples and the Australian vs European samples in PC 2,3 and 4 (Figure 7.2.3.1). Additionally, there appeared to be little effect of plant host on how samples clustered in the PCA with the exception of samples collected from tobacco which form a distinct group differentiated in PC1 and PC3 (figure 7.2.3.1 A and B**)**, supporting the sub-species classification of *M. persicae nicotianae* (Blackman et al., 2007, Eastop and Blackman, 2005).

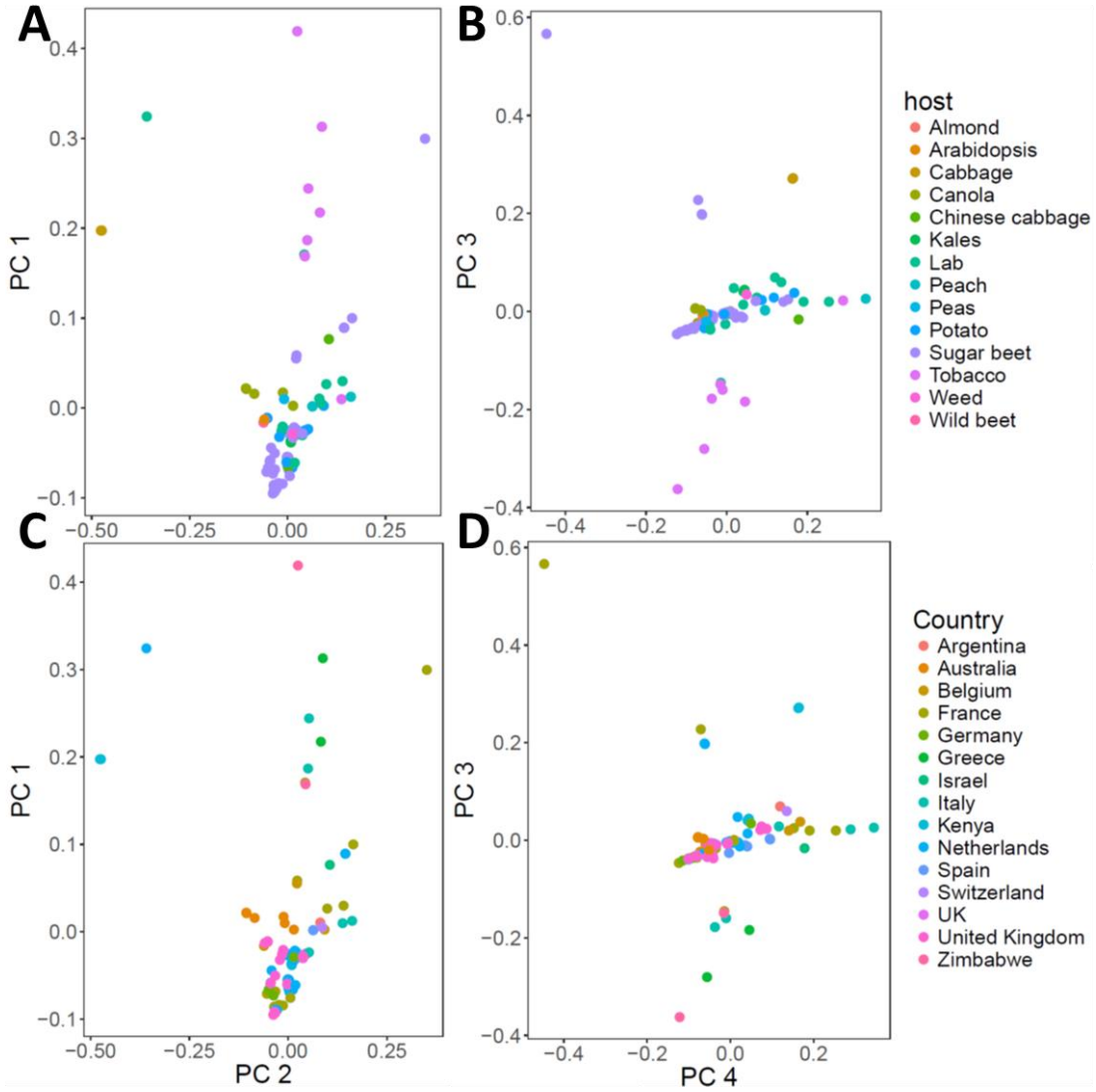


Figure 7.2.3.1 Principal component analysis (PC1-PC2, PC3-PC4). PCA that shown around 15.5% of the variance between the genomic sample set of 2.8 million SNPs. **A. and B.** Shown samples colour coded per host. (**A**) presents PC1 and PC2, (**B**) presents PC3 and PC4. **C. and**

D. represents samples colour coded per country. (**C**) presents PC1 and PC2, (**D**) presents PC3 and PC4. PCs explain the following percentage of variation in genetic subset; PC1=4.2%, PC2=3.9%, PC3=3.8% and PC4=3.6% respectively. The PCA with genotype labels are found in the appendix (figure S1)

To further investigate potential *M. persicae* population structure, I built a Neighbour-Net network using all 2.8 million SNPs. As the SNP data are unphased I selected a single randomised haplotype to represent each sample. The network gives an overview of the phylogenetic relatedness of samples with reticulations indicating possible recombination between the samples or groups. The network analysis confirmed our previous finding that there is no strong clustering based on plant species or country. Yet, two main groups were observed within the network (figure 7.2.3.2 A; appendix figure S2). One group with potentially 4/5 subgroups was found (on the left side of figure 7.2.3.2). This 'left-hand group' had more reticulation (linking lines), suggesting recent recombination (i.e. sex) of these aphids. In contrast, the other group (at the right side of figure 7.2.3.2) was more star-like and had long branches, suggesting that this 'right-hand group' may consist of many diverse clonal lineages. I selected five groups, 3 groups from the left-hand group and two from the right-hand group for further analyses. These groups were selected based on the network analysis (figure 7.2.3.2) and the Bayesian hierarchical clustering analysis (figure 7.2.3.4).

In order to understand the genetic distance among the five selected groups, pairwise genetic differentiation was calculated (table 7.2.3.1). The pairwise FST values ranged from 0.218 (between group 1 and group 3) to 0.436 (between group 3 and group 5). Group 5 showed the highest FST values among all the tested groups. The higher FST between population 5 and other groups suggested high genetic differentiation between group 5 and the other groups. Moreover, between the left and the right group, the genetic distance was small (0.090) compared to that of the five selected groups in the analysis. This could either imply that low genetic differentiation was found among the two groups, or too much variation was found within the groups. The latter possibility is more likely given the high genetic distance among the five groups.

Table 7.2.3.1 Weighted Average FST between five different *Myzus persicae* populations (as selected in figure 7.2.3.2).

	Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**
Group 1	-				
Group 2	0.303	-			
Group 3	0.218	0.307	-		

Group 4	0.274	0.293	0.245	-	
Group 5	0.324	0.428	0.436	0.308	-

To further investigate population structures among *M. persicae* genotypes based on host and/or location, I used an admixture analysis on the 2.8 million variable sites within the subset to firstly reveal the potential number of populations (K) and secondly, how these populations were segregated within the 99 WGS genotypes over location and host.

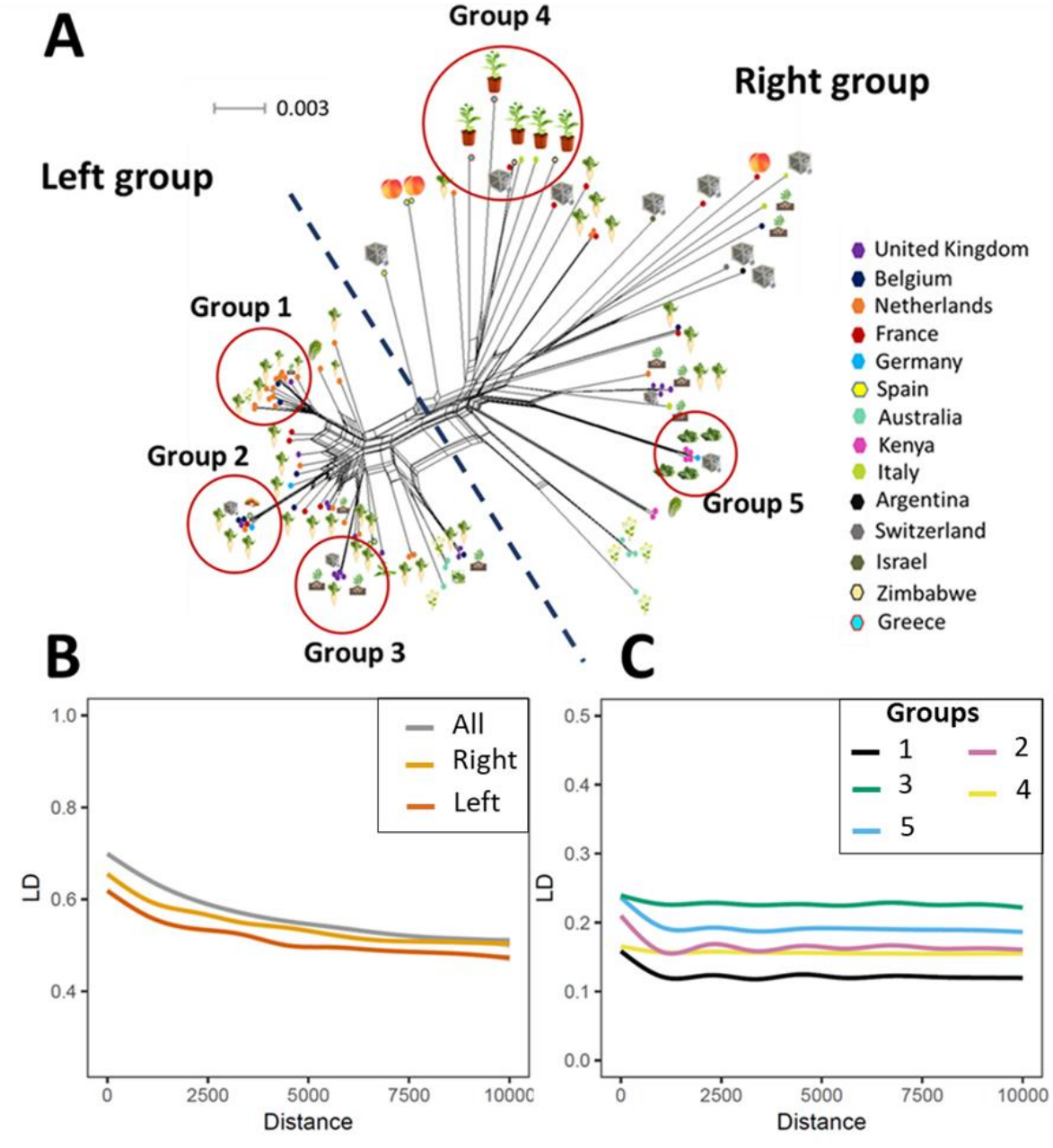


Figure 7.2.3.2 Neighbour-Net network based on the WGS variance file from 2.8 million SNPs of 99 *M. persicae* genotypes and linkage disequilibrium plots of 5 subgroups. (**A**) Neighbour-Net network analysis of 99 WGS samples. The scale represents distances estimated using the uncorrected p-distance. The samples were colour coded according to the country they were collected on, and the illustration depicts the host the sample was collected from, respectively. The squared insect cage/box indicated lab-reared samples and, therefore, mostly forced on a different host upon collection. Labelled circles in the network indicated the referred group used in the linkage disequilibrium plots (group 1-5). The blue interrupted line infers the cut off between the 'right' and 'left' group shown in the LD plot in B. The network analysis with genotype label are found in the appendix (figure S2) (**B and C**) Linkage disequilibrium (LD; R^2) plotted over X-axes over distance in bases plotted over the Y-axes.(**B**) The linkage disequilibrium was plotted over the genome. The lines were categorised according to grouping as indicated with the interrupted line within the network analysis. grey represents the LD of all samples included. The orange (Right group) and red (left group) are the two big clusters (without *nicotianae*). (**C**) The linkage disequilibrium was plotted over the genome of selected WGS samples, categorised according to grouping within the network analysis. Black = group 1, pink = group 2, Green = group 3, Yellow= group 4 and blue = group 5.

To investigate the contribution of sex (ie. recombination) to patterns of *M. persicae* diversity within the subsampled WGS samples, linkage disequilibrium (LD) decay with r^2, the square correlation coefficient between two loci found within the genotypes, was checked (Hill and Robertson, 1968). LD or co-segregation of alleles of two or more loci, is non-random and therefore likely to be different from the expected random segregation (Rockman and Kruglyak, 2009). Co-segregation can be identified within a close perimeter or on different chromosomes (Rockman and Kruglyak, 2009) and be the consequence of linkage drag (Lewontin, 1964), mutations (Cutter et al., 2006), or clonal populations (Singh et al., 2020b, Bradic and Carlton, 2018). Moreover, LD ratios give an indication of the recombination frequency with high linkage and no decay indicating lack of recombination (or a recent bottleneck) and recombination promoting random genetic shuffling of alleles and therefore LD decay. Given this, LD ratios give indications of whether recombination have taken place, possibly as a result of sex, in the recent past, which a higher LD ratios indicating low recombination rates, a recent bottleneck, or stronger selection (Cutter et al., 2006, Lewontin, 1964, Hill and Robertson, 1968). I found that the LD decay within the five selected groups (figure 7.2.3.2 C) was lower compared to that of the entire dataset of 99 (figure 7.2.3.2B). Groups 1 and 2 showed decay within 1 kb distance from the compared SNP. Remarkably, groups 3, 4 and 5 showed hardly any LD decay, with group 3 showing the highest LD on average compared to the two other groups. The samples in group 3 are lab genotypes that likely do not have a sexual stage, and all four of these were collected in the UK. Therefore, the lack

of LD decay might be a result of clonal propagation without a sexual stage. Furthermore, the network analysis observation of the left and the right group found that the LD is higher in the right-hand group than the left-hand group, which indicates less recombination or stronger selection in the latter group.

I used an admixture model implemented in the Admixture-1.3.0 software to explore different numbers of K to the population structure based on 2.8 million SNPs. Runs were carried out for the entire data set and values of K between 1 and 40. The population number predictions were based on the number of Bayesian information criteria and cross-validation errors (Pritchard et al., 2000). The fewer criteria that are being used to predict the number of populations the less error are found within the dataset. This indication is important to avoid overestimating the number of putative ancestral populations. In the subsample set, I found that K=8 and K=10 had the lowest error score (figure 7.2.3.3) and therefore had the highest chance to be correct without overfitting the data. Both K=8 and K=10 had similar cross validation error scores. In this case, it was best to select the lowest number of K (K=8) to avoid baseless splitting and overfitting the data.

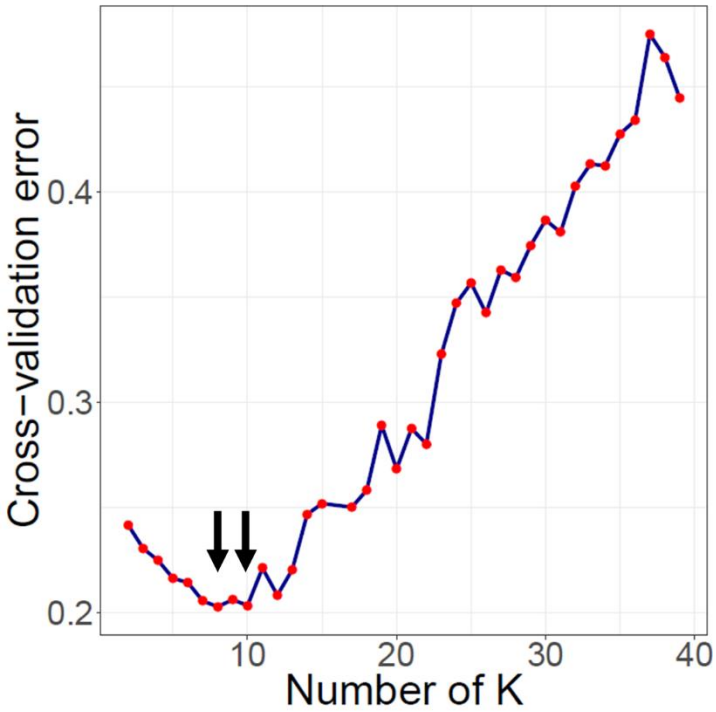


Figure 7.2.3.3 Admixture prediction of the number of populations (K) for the hierarchical genetic structure that was found within the 99 *M. persicae* WGS subset.

The Bayesian clustering of 99 WGS samples using K=8, K=9 or K= 10 were grouped by host (figure 7.2.3.4 A) and by country (figure 7.2.3.4 B). Bayesian clustering was independently performed for K=2 up to K=10 as

shown in the appendix (figure S3 and figure S4). Population separation in the Bayesian clustering did show a separation of *M. persicae* for a few plant host species; these were canola, cabbage and tobacco that had unique populations per host. However, the aphids on canola and cabbage in these clusters were solely collected from either Australia or Kenya. Therefore, it remains unsure whether this clustering was host or country dependent or both. Some genotypes show a mixture of two ancestral populations indicating gene flow (either ongoing or historical) between the identified clusters. Some clustering based on country was observed within the field-collected samples, including some samples from Kenya that clustered together and found to have the highest genetic differentiation among the tested groups (FST=0.324). Only the *M. persicae* subspecies *nicotianae* received from Singh et al., 2020, including two lab-reared colonies, clustered together and had *Nicotiana* species as a shared plant host (group 4; Figure 7.2.3.2A).

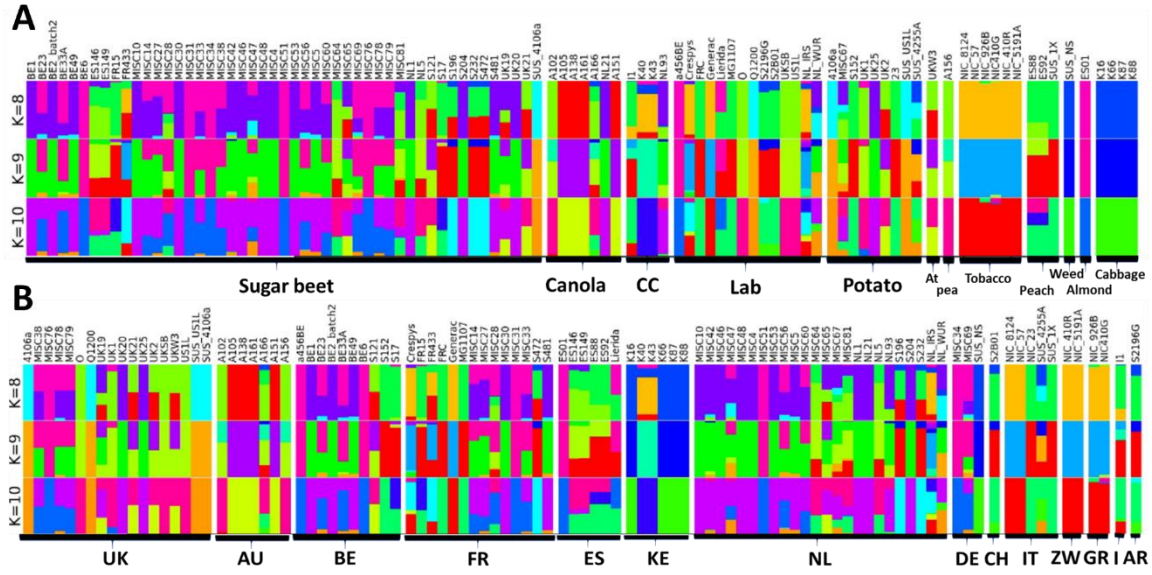


Figure 7.2.3.4 Bayesian hierarchical clustering plots per predicted population number from K=8 to K= 10, grouped per host or per country designed with Admixture. Genotypes were colour coded by predicted population genetic ancestor (K). **A** grouped per host (CC = Chinese cabbage, At = *Arabidopsis thaliana*). **B** grouped per country (country ISO codes used for each country, except Israel that is represented as "I"). The additional admixture plots from K=2 up to K=7 with genotype label are found in the appendix (grouped per host: S3 and grouped per country: S4).

7.2.4: Selective sweep analysis of five selected groups.

Next, I searched the *M. periscae* genomes for signatures of recent selection by scanning for selective sweeps. A selective sweep occurs when a

beneficial mutation spreads through a population (Barton, 2000, McVean, 2007, Begun and Aquadro, 1992, Smith and Haigh, 2009). This process causes flanking neutral regions to be dragged along with the selected mutation due to linkage – a process known as "hitchhiking" (Smith and Haigh, 2009). As a result, selective sweeps create regions of locally low diversity around the selected mutation, the size of which depends on the local rate of recombination and the strength of selection (Barton, 2000, Smith and Haigh, 2009). Various tools and models were developed to identify selected genomic regions using these signatures (Nielsen et al., 2005, Pavlidis and Alachiotis, 2017, Akbari et al., 2018). Here I used SweeD version 3.3.2 (Pavlidis et al., 2013) to identify candidate selective sweep regions in the *M. persicae* genomes. SweeD calculates a composite likelihood ratio (CLR) via a model according to Kim and Stephan (2002), to predict the likelihood of a selective sweep within the given area. *M. persicae* genomic regions involved in selective sweeps may be associated with host specialisation (for example, in the nicotianae sub-group) or insecticide resistance or other unidentified processes.

First, I looked for potential selective sweep areas within the two large groups identified by my population structure anaysis: "group left" and "group right" (figure 7.2.6.1). Secondly, I looked further for selective sweep areas within the five smaller sub-groups (numbered 1-5; figure 7.2.6.3), including the *nicotianae* cluster (group 4). Regions putatively involved in selective sweeps in these sub-groups may be associated with host adaptations (particularly for sweeps associated with the *nicotianae* sub group) and provide targets for future functional analysis.

Putative sweep regions were flagged for follow-up analysis if they had CLR >200. To investigate the signatures of selection within the areas likely for a selective sweep, I compared the following three different summary statistics across 5 Mb windows surrounding the potential selective sweep area: population differentiation (FST); nucleotide diversity (Pi); and Tajima's D. For FST calculations, I compared the focal group to all other genomes within the sample set.

7.2.4.1 Selective sweep and Tajima D analysis of the left and the right group

The searching of selective sweep signatures between the "right group" and "left group" (figure 7.2.4.1) revealed no strong composite likelihood ratios (CLR) (figure 7.2.4.1). The likelihood was checked in 100000 locations per one of the six super scaffolds consisting of 2.8 million SNPs. No likelihood scores

above 200 were observed within either the right or the left group. Consistent with a lack of a strong signal in the CLR statistic, no obvious strong dips in Tajima's D were observed in both the right and the left groups (figure 7.2.4.2). Nevertheless, different Tajima's D statistics were observed between these groups. While the "right group" had a Tajima's D of 0, which means that no evidence of selection was found in this group and variation was observed as expected, the 'left group' had a Tajima's D of around 1, suggesting that the latter group has less pairwise differences and might have undergone a sudden population contraction. This is consistent with the network analyses (figure 7.2.5.2), because the branch lengths of the 'left group' were shorter than those of the 'right group'.

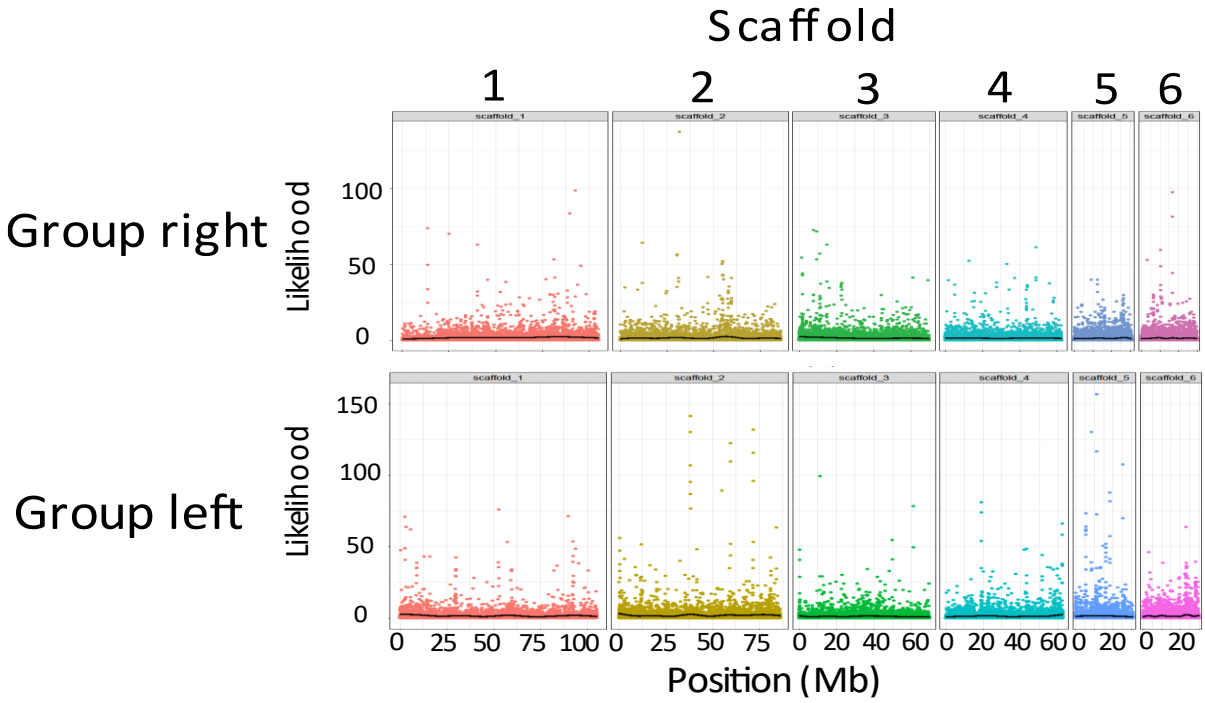


Figure 7.2.4.1 Selective sweep analysis of left and right groups. Composite likelihood ratios of selective sweep areas were plotted over the six main super scaffolds. Colours indicate the different scaffolds, and the black line indicates the trend.

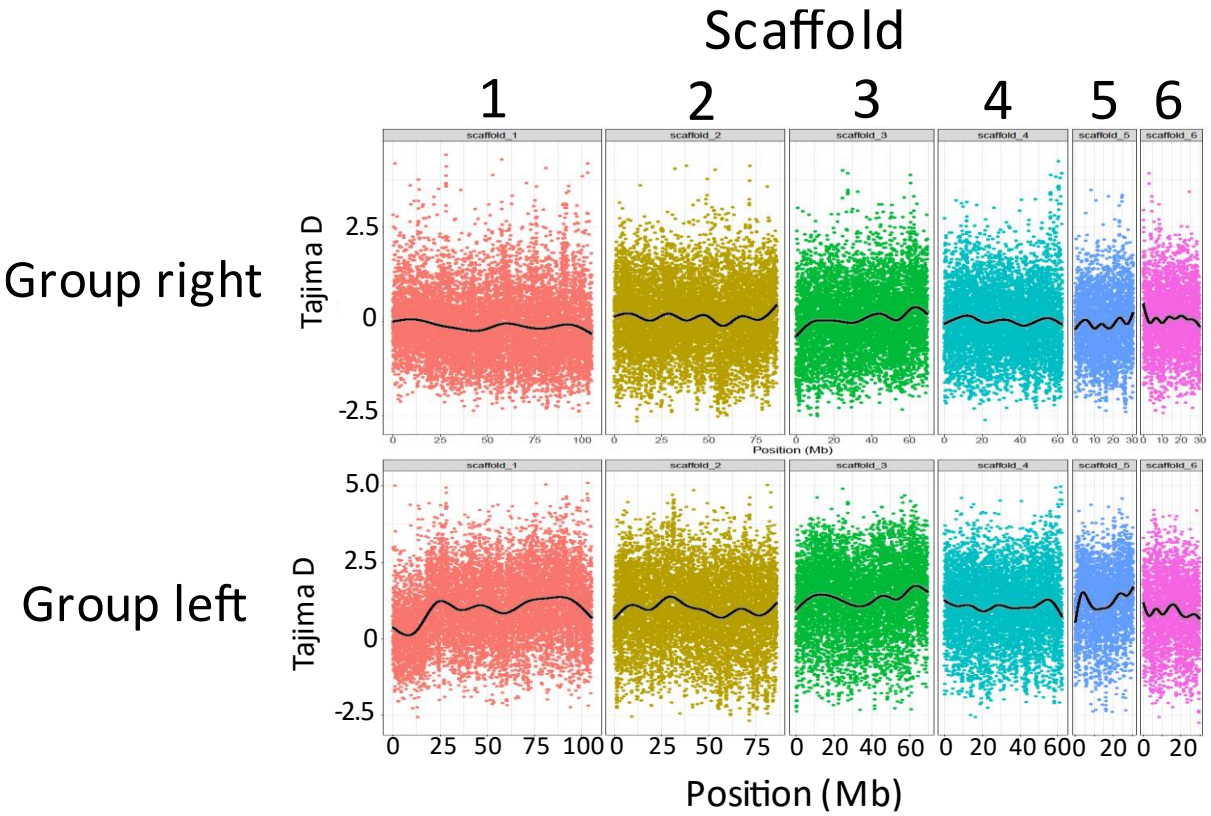


Figure 7.2.4.2 Tajima's D analysis of left and right groups. Tajima D scores were calculated using 10 kb windows per 5 kb slides. Colours indicate different scaffolds, and the black line indicates the trend using the locally estimated scatterplot smoothing (LOESS) method.

7.2.4.2 Selective sweep and Tajima D analysis of the five selected groups

In contrast to the large groups investigated above a total of 33 potential selective sweep regions were found within the five smaller groups (figure 7.2.4.3). Group 1 had four CLR peaks, one in scaffold 1, one in scaffold 2, one in scaffold 4 and one in scaffold 6. Group 2 had five CLR peaks, of which three on scaffold 1 and one on each of the scaffolds 2 and 4. Group 3 had ten CLR peaks, of which four on scaffold 1, three on scaffold 2 and one on each of the scaffolds 3, 4 and 5. Group 4 had nine CLR peaks with four on scaffold 1, two on scaffold 2 and one on each of the scaffolds 3, 4 and 5. Group 5 had five CLR peaks, with three peaks found on scaffold 1 and one on each of the scaffolds 2 and 5.

Identical regions of high CLR were found in multiple groups. These included one region with an increased likelihood of a selective sweep on scaffold 1 in four distinct groups (group 1, 2, 3 and 5). This selective sweep region was found between 27 million bases and 32 million bases. I found that 16 genes were found underlying this potential selective sweep area. Functional annotations of this region revealed that the genes in this area were annotated as non-coding RNA (ncRNA). Another region of high CLR was shared among three groups (group 2, 3 and 5) and lies within the first five million

bases of scaffold 2. I found 37 genes underlying this region, of which 18 were predicted as protein-coding genes. Finally, in the first 5 million bases of scaffold five, a peak of high likelihood for a selective sweep was shared between groups 3 and 5. I found 6 genes within this region and these were predicted as protein-coding genes. Therefore, multiple regions for selective sweep were found in similar regions among the groups. Most of these overlapping regions were shared among groups 2, 3 and 5.

Considering all 33 potential selective sweep regions, a total of 975 genes were found. Group 3 had the highest number of genes underlying a selective sweep region, with 460 genes found within 5 million bases, whereas group 1 had the lowest number of genes found (33 genes) in a selective sweep region. Altogether, selective sweep regions of groups 2, 4 and 5 had between 100 and 200 genes. Only group 4 had two putative effector genes within one of the CLR peaks. This was the CLR peak on scaffold 2 at region 29.5 Mb up to 34.5 Mb. As mentioned above, group 4 was the only group that had some association with host preference (*Nicotiana sp*.).

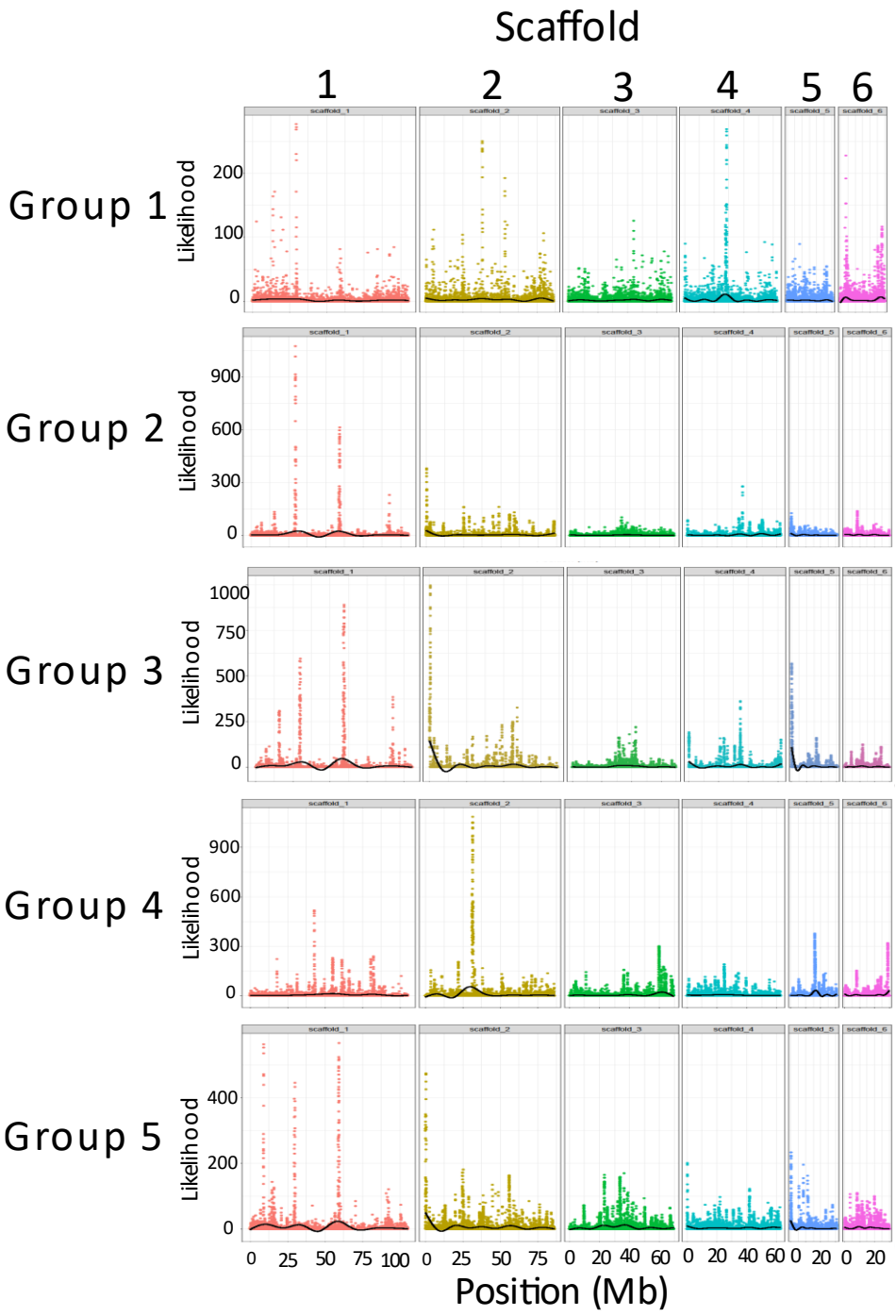


Figure 7.2.4.3 Selective sweep analysis of 5 selected groups. Composite likelihood ratios of selective sweep areas were plotted over the six main super scaffolds. Colours indicate different scaffolds, and the black line indicates the trend. The groups were selected according to population structure differences and presented in figure 7.2.5.2.

The average Tajima's D scores vary per group (figure 7.2.4.4). Whereas group 1 and group 4 had a low Tajima's D of on average near 0 , the other groups had Tajima's D scores of between 1 and 2. Low Tajima's D scores suggest a sudden population contraction or a smaller samples size.

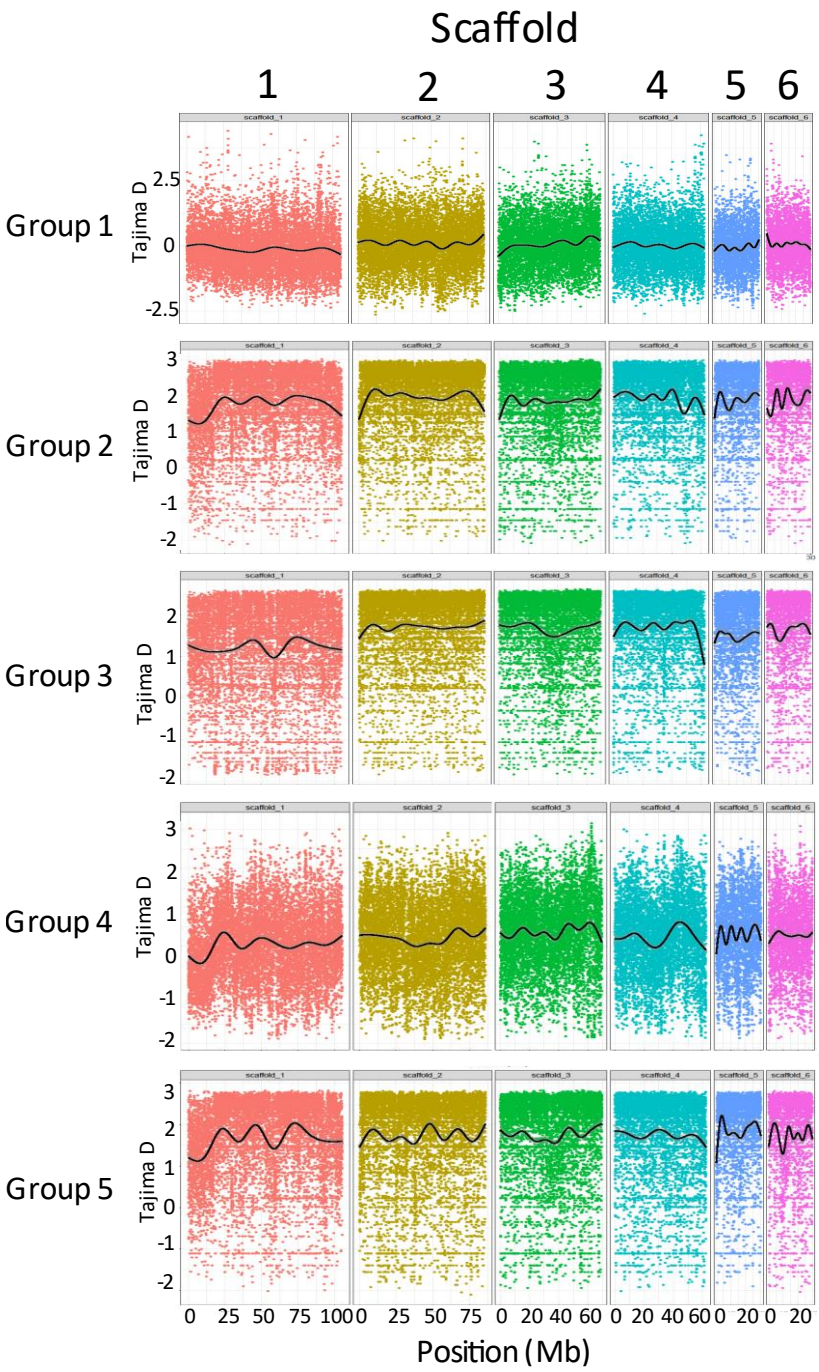


Figure 7.2.4.4 Tajima's D analysis of 5 selected groups. Tajima D scores were calculated using 10 kb windows per 5 kb slides. Colours indicate different scaffolds, and the black line indicates the trend (calculated via the LOESS method). The groups were selected according to population structure differences and presented in figure 7.2.5.2.

7.2.4.3 Selective sweep regions within group 1

In total, four putative selective sweep areas were found across the six chromosomes in group 1. I selected the selective sweep region four on scaffold 6 that had the most support using statistical analyses for further analysis (figure 7.2.4.5). In this region, the FST did not fluctuate more than 0.1 over the 5 Mb window (figure 7.2.4.5B), which indicated a low population

differentiation for this region. Moreover, the nucleotide diversity analyses showed a flattened line (figure 7.2.4.5C) and a drop in Tajima's D (figure 7.2.4.5D). These results indicated that the region has a low nucleotide diversity and a potentially abundance of rare alleles. Furthermore, from the gene annotation, I found only a single gene within this region (gene ID: 0346230). This was predicted as a protein coding gene as a Dynein heavy chain protein family member (GO:0003777, GO:0005524, GO:0007018, GO:0030286). These proteins are known to function as translocator motor proteins on microtubules (Asai and Wilkes, 2004).

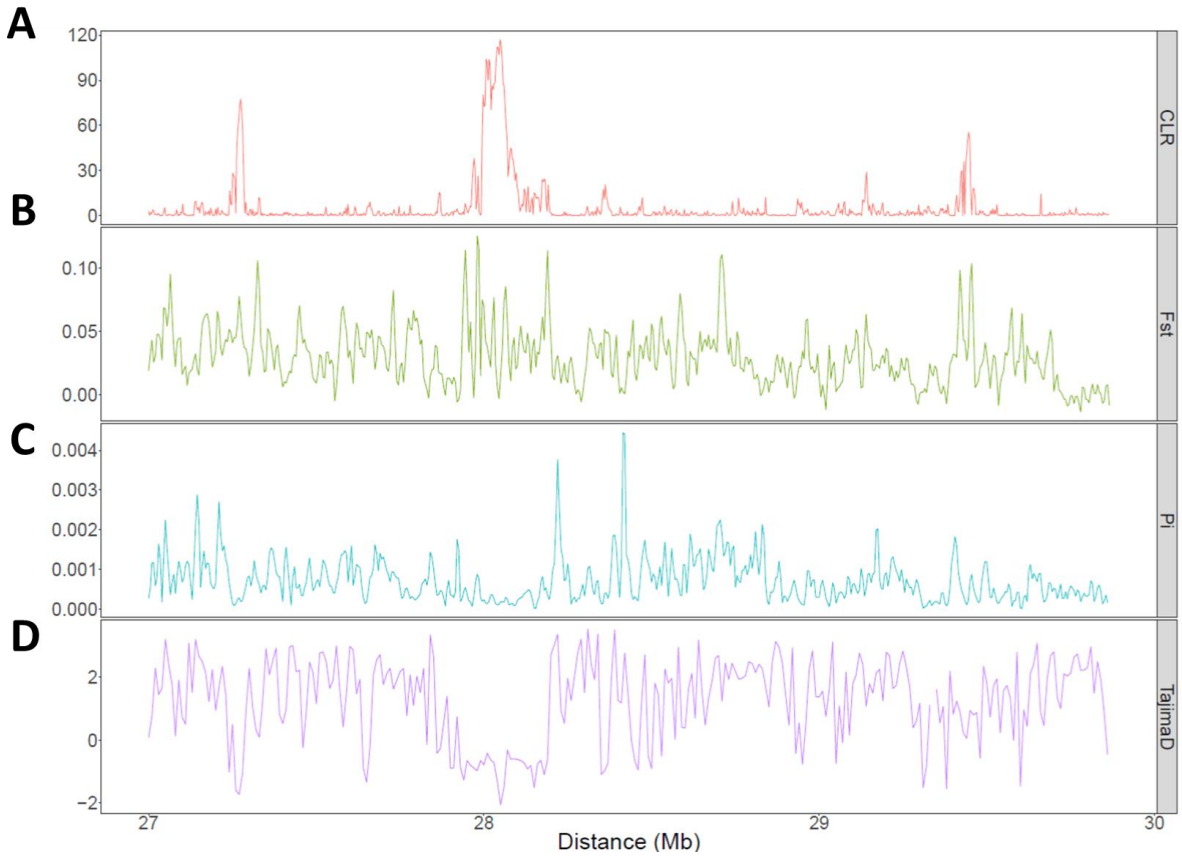


Figure 7.2.4.5 Summary statistics of putative selective sweep area on scaffold 6 (region 4 from group 1). Colours indicate the different scores. (**A**) Selective sweep likelihood quantified in CLR. (**B**) weighted population differentiation score in FST, between group 1 and all other genomic samples. (**C**)Nucleotide diversity (Pi). (**D**) Tajima's D statistic.

7.2.4.4 Selective sweep regions within group 2

In total five putative selective sweep areas were found within the 6 scaffolds of group 2. Additional statistical analyses showed that two selective sweep regions looked most promising. These were selective sweep regions one and two that were both found on scaffolds 1 and 2, respectively. Of these, region one was also shared in other groups. FST of this region was reduced relatively to surrounding regions (figure 7.2.4.6B), whereas the nucleotide

diversity was higher (figure 7.2.4.6C) and Tajima's D lower (figure 7.2.4.6D). I found 16 genes underlying this region (Gene ID: 0044760, 0044770, …, up to 0044910). All these 16 genes were annotated as ncRNAs.

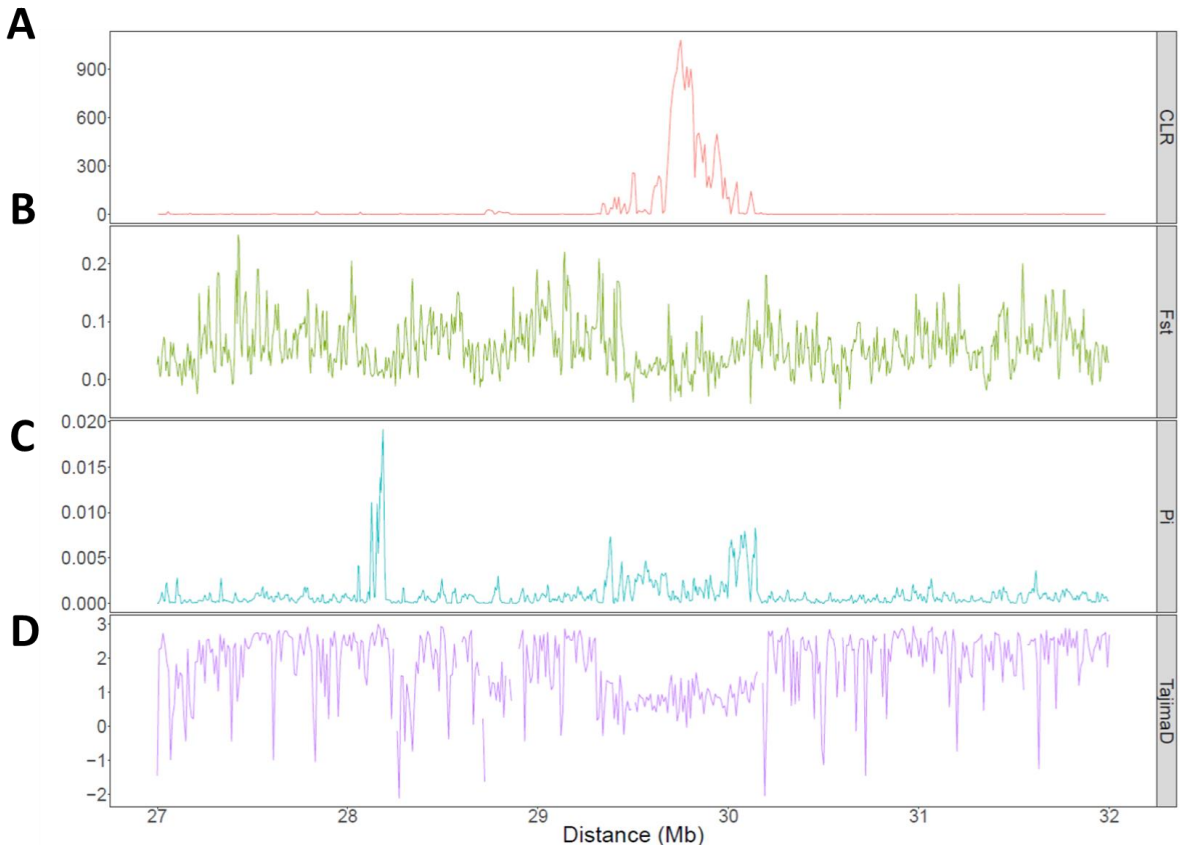


Figure 7.2.4.6 Summary statistics of putative selective sweep area on scaffold 1 (region 1 from group 2). Colours indicate the different scores. (**A**) Selective sweep likelihood quantified in CLR. (**B**) weighted population differentiation score in FST, between group 2 and all other genomic samples. (**C**)Nucleotide diversity (Pi). (**D**) Tajima's D statistic.

Region two that was found on scaffold 2 showed a single narrow CLR peak (figure 7.2.6.7A). However, the FST analysis showed a single narrow peak (figure 7.2.6.7B), which may indicate a strongly positive selected region. The nucleotide diversity line under the CLR peak was slightly more flattened (figure 7.2.6.7C) and Tajima's D was lower (figure 7.2.6.7D) compared to the surrounding areas. The three genes underlying this region (Gene ID: 0045230, 0045240, 0045250) were annotated as ncRNAs.

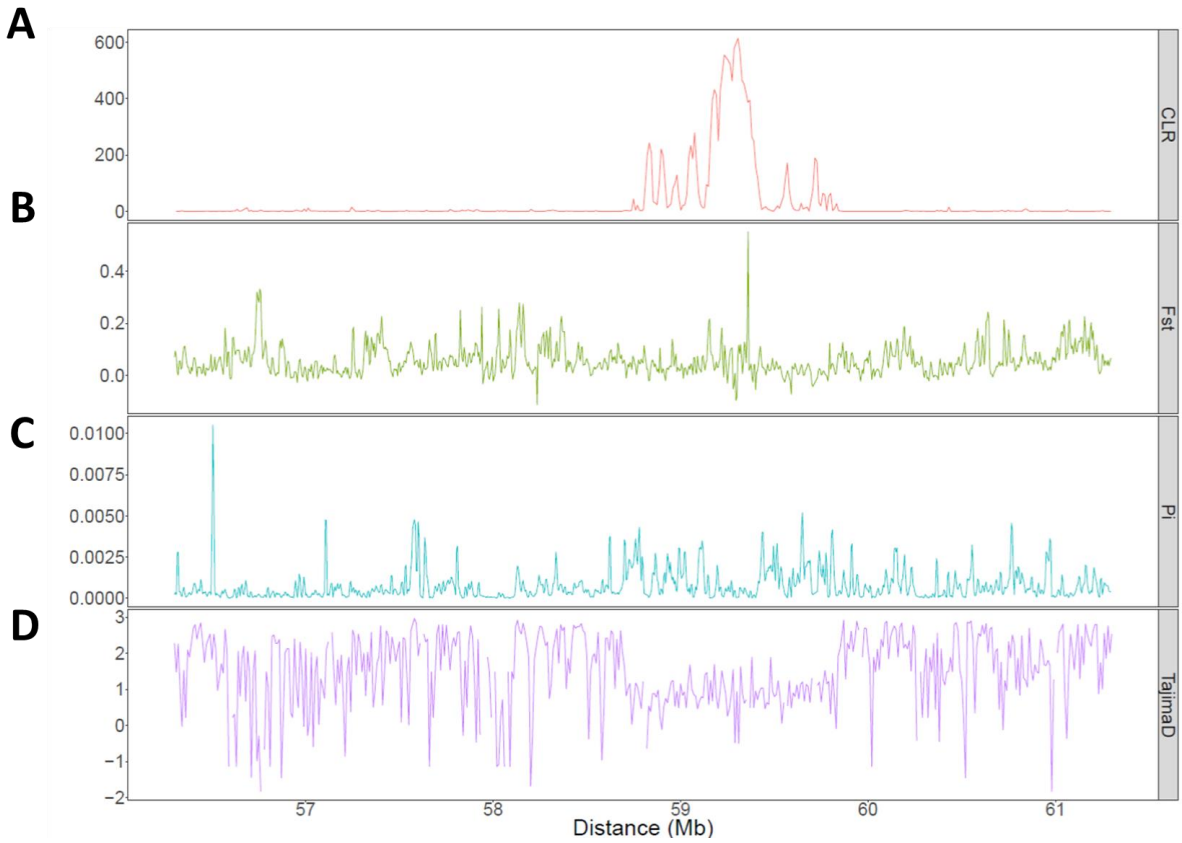


Figure 7.2.4.7 Summary statistics of putative selective sweep area on scaffold 1 (region 2 from group 2). Colours indicate the different scores. (**A**) Selective sweep likelihood quantified in CLR. (**B**) weighted population differentiation score in FST, between group 2 and all other genomic samples. (**C**)Nucleotide diversity (Pi). (**D**) Tajima's D statistic.

7.2.4.5 Selective sweep regions within group 3

In total ten putative selective sweep areas were found within the 6 scaffolds of group 3. Selective sweep regions five and eight that were found on scaffold 2 and scaffold 3, respectively. Region five was also found to be selective sweep regions in groups 2, 3 and 5. However, region 5 had no clear support in nucleotide diversity and Tajima's D analyses (figure 7.2.4.8A, C, D), though narrow dips in the FST were found (figure 7.2.4.8B). I found 37 genes within this region (Gene ID: 0128490, 0128500, ... up to 128850; table 7.2.4.1). One of these protein coding genes is depicted as an activity-regulated cytoskeleton-associated protein. These proteins are well known for their synaptic plasticity function in mammals (Nikolaienko et al., 2018) and shown to be a regulator of behaviour responses to starvation in *Drosophila* (Mattaliano et al., 2007).

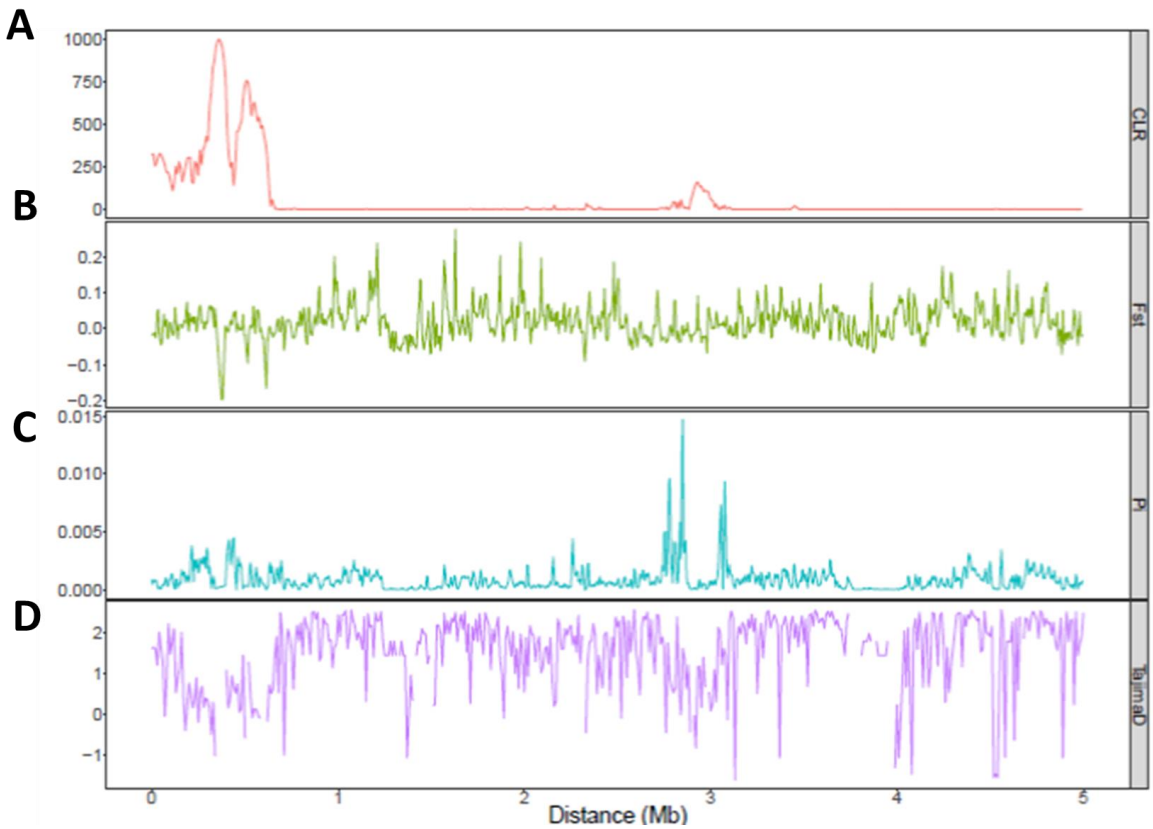


Figure 7.2.4.8 Summary statistics of putative selective sweep area on scaffold 2 (region 5 from group 3). Colours indicate the different scores. (**A**) Selective sweep likelihood quantified in CLR. (**B**) weighted population differentiation score in FST, between group 3 and all other genomic samples. (**C**)Nucleotide diversity (Pi). (**D**) Tajima's D statistic.

Table 7.2.4.1 Gene description of genes found within 5 Mb window near the selective sweep area on scaffold 2 (region 5 from group 3). Colours indicate the biotype of the gene. Blue = protein-coding gene and white = transposable element. Seven ncRNAs were removed from the table.

Gene	Biotype	Description	Gene ontology (GO) Term
128510	protein_coding_gene	CCHC-type domain-containing protein	GO:0003676, GO:0008270
128520	transposable_element_gene	Reverse transcriptase domain-containing protein	None
128540	protein_coding_gene	Activity-regulated cytoskeleton-associated protein	None
128550	transposable_element_gene	Retrovirus-related Pol polyprotein	GO:0003676, GO:0015074
128560	transposable_element_gene	Reverse transcriptase domain-containing protein	None
128570	transposable_element_gene	Reverse transcriptase domain-containing protein	None
128580	protein_coding_gene	MULE domain-containing protein	GO:0016020, GO:0016021
128590	protein_coding_gene	Dimer_Tnp_hAT domain-containing protein	GO:0046983
128600	predicted_gene	Unknown protein	None
128610	protein_coding_gene	U-box domain-containing protein 16	None
128620	transposable_element_gene	Integrase catalytic domain-containing protein	GO:0003676, GO:0015074

128630	transposable_element_gene	Reverse transcriptase domain-containing protein	None
128640	protein_coding_gene	DUF4806 domain-containing protein	None
128650	protein_coding_gene	DUF4806 domain-containing protein	None
128660	transposable_element_gene	Zinc finger BED domain-containing protein	GO:0046983
128670	protein_coding_gene	Zinc finger MYM-type protein 1	GO:0046983
128680	protein_coding_gene	DUF4806 domain-containing protein	None
128690	protein_coding_gene	DDE_Tnp_1_7 domain-containing protein	None
128700	transposable_element_gene	Unknown protein	None
128710	protein_coding_gene	DUF4806 domain-containing protein	None
128720	protein_coding_gene	HTH CENPB-type domain-containing protein	GO:0003677
128730	protein_coding_gene	DDE-1 domain-containing protein	GO:0003676
128740	protein_coding_gene	Integrase catalytic domain-containing protein	GO:0003676, GO:0015074
128750	transposable_element_gene	DNA_pol_B_2 domain-containing protein	GO:0000166, GO:0003677, GO:0003887, GO:0006260, GO:0071897
128760	protein_coding_gene	Integrase catalytic domain-containing protein	GO:0003676, GO:0015074
128770	protein_coding_gene	DUF1758 domain-containing protein	None
128780	protein_coding_gene	Unknown protein	None
128790	protein_coding_gene	RING-type domain-containing protein	None
128800	protein_coding_gene	Integrase_H2C2 domain-containing protein	None
128810	transposable_element_gene	CCHC-type domain-containing protein	GO:0003676, GO:0008270

The CLR of region eight on scaffold 3 (figure 7.2.4.9A) was supported by a narrow dip in FST (figure 7.2.4.9B), a narrow spike in the nucleotide diversity (figure 7.2.4.9C) and a slight fall in Tajima's D (figure 7.2.4.9D). I found three genes within this region (Gene ID: 0240760, 0240770, 0240780), of which one was depicted as a protein-coding gene (0240770) and the other were found to be ncRNAs. The protein coding gene was predicted as a PHD domain-containing protein (GO:0016020, GO:0016021, GO:0046872). The PHD zinc finger family is described as a epigenome reader that control gene expression via methylation (Sanchez and Zhou, 2011).

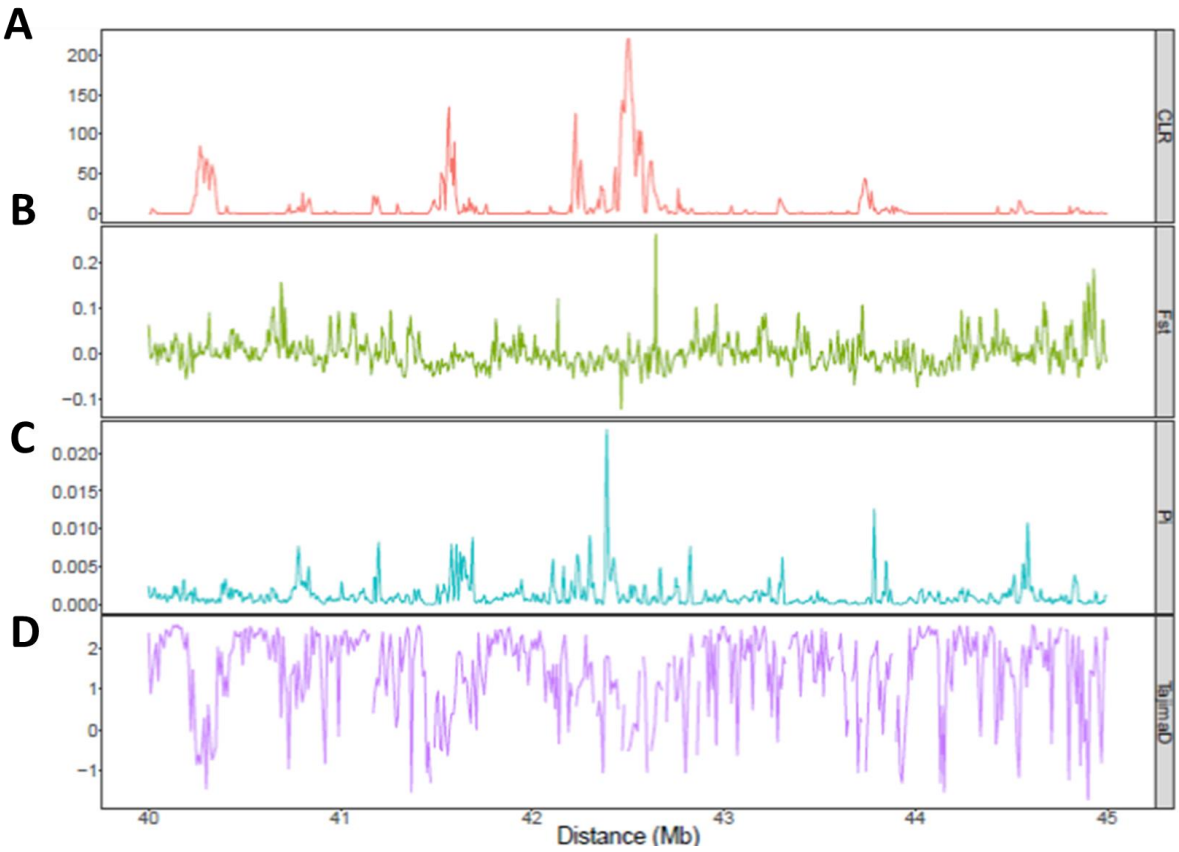


Figure 7.2.4.9 Summary statistics of putative selective sweep area on scaffold 3 (region 8 from group 3). Colours indicate the different scores. (A) Selective sweep likelihood quantified in CLR. (B) weighted population differentiation score in FST, between group 3 and all other genomic samples. (C)Nucleotide diversity (Pi). (D) Tajima's D statistic.

7.2.4.6 Selective sweep regions of group 4 (nicotianae subspecies cluster)

In total, nine putative selective sweep areas were found within the 6 scaffolds of group 4. I selected four selective sweep areas that were supported by FST, nucleotide diversity or Tajima's D analyses. These were selective sweep regions five, six, eight, and nine found on scaffolds 2, 2, 5 and 6, respectively. The CLR on region 5 on scaffold 2 (figure 7.2.4.10A) also had a single narrow FST peak (figure 7.2.4.10B) and a a fall in Tajima's D (figure 7.2.4.10D), though no difference in nucleotide diversity was observed (figure 7.2.4.10C). I found four genes within this region (Gene ID: 0093150, 0093160, 0093180, 0093190), including 2 protein-coding genes, one transposable element (table 7.2.4.2) and one ncRNA (0093190). Moreover, one gene was classified as a Arestin C domain-containing protein that could be interacting with G-proteins in cell signalling (Gurevich and Gurevich, 2018).

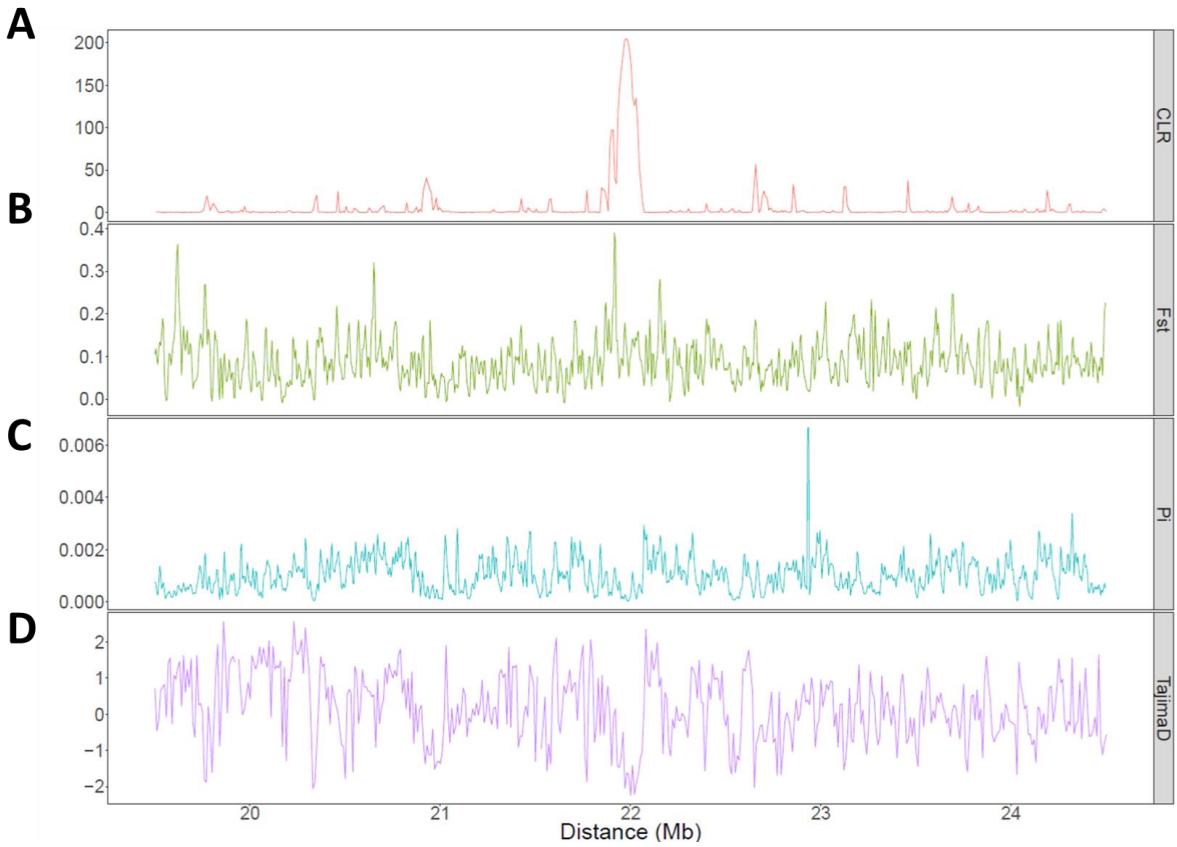


Figure 7.2.4.10 Summary statistics of putative selective sweep area on scaffold 2 (region 5 from group 4). Colours indicate the different scores. (**A**) Selective sweep likelihood quantified in CLR. (**B**) weighted population differentiation score in FST, between group 4 and all other genomic samples. (**C**)Nucleotide diversity (Pi). (**D**) Tajima's D statistic.

Table 7.2.4.2 Gene description of genes found within 5 Mb window near the selective sweep area on scaffold 2 (region 5 from group 4). Colours indicate the biotype of the gene. Blue = protein-coding gene and white = transposable element. The ncRNA single was removed from the table.

Gene	Biotype	Description	Gene-Ontology-Term
93150	protein_coding_gene	Nucleic-acid-binding protein	None
93160	transposable_element_gene	Reverse transcriptase domain-containing protein	None
93180	protein_coding_gene	Arrestin_C domain-containing protein	GO:0007165

The CLR of region six on scaffold 2 (figure 7.2.4.11A) overlapped with an increase in FST (figure 7.2.4.11B) and a fall in Tajima's D (figure 7.2.4.11D) and no obvious change in nucleotide diversity (figure 7.2.4.11C). I found 70 genes within this region (Gene ID: 0155220, 0155230, …, up to 0155910) of which two putative effectors (Gene ID: 0155220 and 0155260; table 7.2.4.3). Six more genes with unknown functions were found.

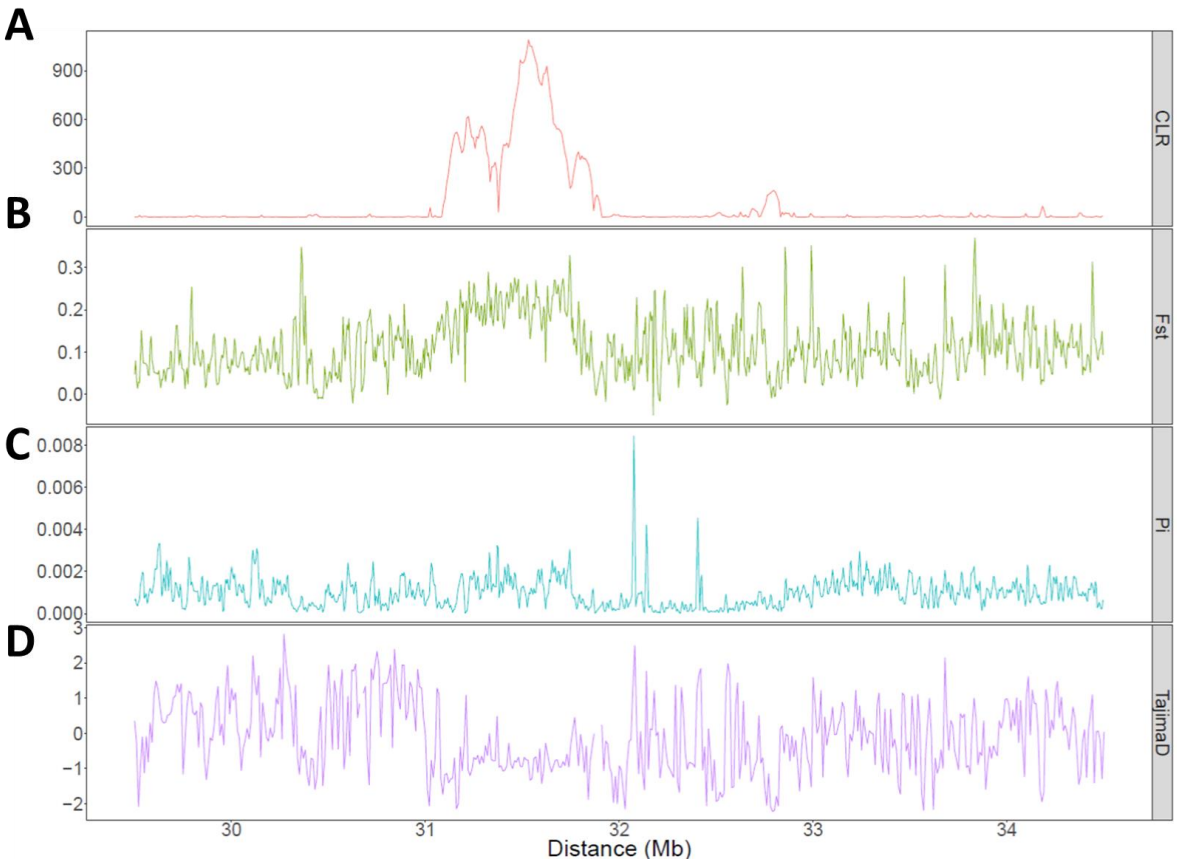


Figure 7.2.4.11 Summary statistics of putative selective sweep area on scaffold 2 (region 6 from group 4). Colours indicate the different scores. (**A**) Selective sweep likelihood quantified in CLR. (**B**) weighted population differentiation score in FST, between group 4 and all other genomic samples. (**C**)Nucleotide diversity (Pi). (**D**) Tajima's D statistic.

Table 7.2.4.3 Gene description of genes found within 5 Mb window near the selective sweep area on scaffold 2 (region 6 from group 4). Colours indicate the biotype of the gene. Blue = protein-coding gene and white = transposable element and orange = putative effector). The 21 ncRNAs were removed from the table.

Gene	Biotype	Description	Gene ontology term
155220	protein_coding_gene	Unknown protein	None
155240	transposable_element_gene	MULE domain-containing protein	None
155250	transposable_element_gene	THAP-type domain-containing protein	GO:0003676, GO:0004803, GO:0006313
155260	protein_coding_gene	Unknown protein	None
155270	protein_coding_gene	MULE domain-containing protein	None
155310	transposable_element_gene	DUF1758 domain-containing protein	GO:0004190, GO:0006508
155320	transposable_element_gene	Integrase catalytic domain-containing protein	GO:0003676, GO:0015074
155340	protein_coding_gene	THAP-type domain-containing protein	GO:0003676, GO:0003677, GO:0046983
155350	transposable_element_gene	Reverse transcriptase domain-containing protein	None

155360	protein_coding_gene	DDE-1 domain-containing protein	GO:0003676
155370	transposable_element_gene	Integrase catalytic domain-containing protein	GO:0003676, GO:0015074
155380	protein_coding_gene	NYD-SP28_assoc domain-containing protein	GO:0005858, GO:0070286
155390	transposable_element_gene	Reverse transcriptase domain-containing protein	None
155410	protein_coding_gene	Regulator of nonsense transcripts 1	None
155420	protein_coding_gene	FHA domain-containing protein	None
155430	protein_coding_gene	Unknown protein	None
155440	protein_coding_gene	Poly(A) RNA polymerase	None
155450	protein_coding_gene	Protein kinase domain-containing protein	GO:0000166, GO:0004672, GO:0004674, GO:0004689, GO:0005516, GO:0005524,
155460	protein_coding_gene	BHLH domain-containing protein	GO:0046983
155480	protein_coding_gene	Sterol regulatory element-binding protein 2	GO:0006355, GO:0016020, GO:0016021, GO:0046983
155490	protein_coding_gene	THAP-type domain-containing protein	GO:0003676, GO:0003677, GO:0004803, GO:0006313
155500	protein_coding_gene	THO complex subunit 5	None
155510	transposable_element_gene	DUF659 domain-containing protein	GO:0000122, GO:0001227, GO:0003690, GO:0005634, GO:0006357
155520	protein_coding_gene	Insulin-like growth factor 2 mRNA-binding protein 1	GO:0003676, GO:0003723
155530	protein_coding_gene	Insulin-like growth factor 2 mRNA-binding protein 3	GO:0003676, GO:0003723, GO:0016020, GO:0016021
155540	protein_coding_gene	ras-related protein Rab-8A	GO:0003924, GO:0005525
155560	protein_coding_gene	Usher syndrome type-1G protein like protein	GO:0007605, GO:0016020, GO:0016021, GO:0050953, GO:0050957
155570	predicted_gene	Unknown protein	None
155580	protein_coding_gene	Ras-related protein Rab-8A	GO:0003924, GO:0005525
155590	transposable_element_gene	HTH CENPB-type domain-containing protein	GO:0003676, GO:0003677
155600	protein_coding_gene	Tudor domain-containing protein 12	None
155610	protein_coding_gene	PRE_C2HC domain-containing protein	None
155620	protein_coding_gene	Unknown protein	None
155630	protein_coding_gene	MULE domain-containing protein	None
155640	protein_coding_gene	Abhydrolase_2 domain-containing protein	GO:0016787
155660	protein_coding_gene	Retrotrans_gag domain-containing protein	None
155670	protein_coding_gene	Unknown protein	None
155700	protein_coding_gene	Unknown protein	None
155710	protein_coding_gene	Enolase	GO:0000015, GO:0000287, GO:0004634, GO:0006096
155720	transposable_element_gene	Unknown protein	None
155740	transposable_element_gene	Integrase catalytic domain-containing protein	GO:0003676, GO:0015074
155750	protein_coding_gene	Retrotransposable element Tf2 protein type 1	GO:0003676, GO:0004190, GO:0006508, GO:0015074
155780	protein_coding_gene	UPF0670 protein	GO:0016020, GO:0016021
155790	protein_coding_gene	UPF0670 protein	GO:0016020, GO:0016021

155800	protein_coding_gene	Retrotrans_gag domain-containing protein	None
155810	protein_coding_gene	MFS domain-containing protein	None
155890	protein_coding_gene	DUF4808 domain-containing protein	GO:0016020, GO:0016021
155900	protein_coding_gene	Fibronectin type-III domain-containing protein	GO:0016020, GO:0016021
155910	transposable_element_gene	Zinc finger MYM-type protein 1	GO:0046983

The CLR of region eight on scaffold 5 (figure 7.2.4.12A) overlapped with multiple FST increases (figure 7.2.4.12B), reduced nucleotide diversity (figure 7.2.4.12C) and a fall in Tajima's D (figure 7.2.4.12D). This region may have rare alleles that have no shared ancestry with the other groups. The low nucleotide diversity could indicate a more recent selective sweep. I found 18 genes within this region (Gene ID: 0331290, 0331290, …, up to 0331460; table 7.2.4.4). The protein coding genes were predicted as transporters (Sec24C), transferases (ceramide choline phosphotransferase 1), heath shock proteins (J domain-containing protein) and a neurotransmitter (Glutamate receptor). Interestingly, glutamate receptors are often insecticide targets (Kane et al., 2000, Wang et al., 2016, Li et al., 2020a).

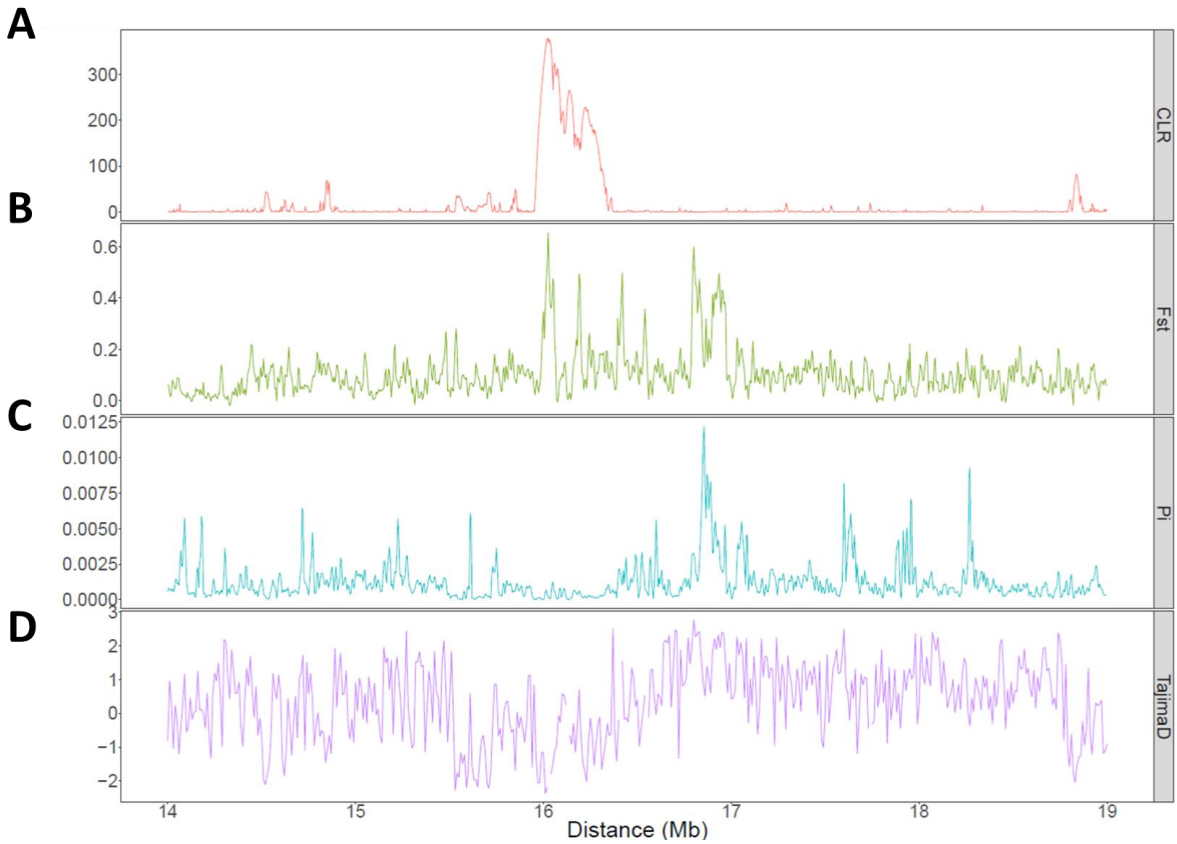


Figure 7.2.4.12 Summary statistics of putative selective sweep area on scaffold 5 (region 8 from group 4). Colours indicate the different scores. (**A**) Selective sweep likelihood

quantified in CLR. (**B**) weighted population differentiation score in FST, between group 4 and all other genomic samples. (**C**)Nucleotide diversity (Pi). (**D**) Tajima's D statistic.

Table 7.2.4.4 Gene description of genes found within 5 Mb window near the selective sweep area on scaffold 2 (region 6 from group 4). Colours indicate the biotype of the gene. Blue = protein-coding gene and white = transposable element and orange = putative effector) The six ncRNAs were removed from the table.

Gene	Biotype	Description	Gene ontology term
331290	protein_coding_gene	KDa protein in nof-fb transposable element	None
331310	protein_coding_gene	Glutamate receptor, ionotropic kainate 2	GO:0004970, GO:0005216, GO:0005886, GO:0006811, GO:0016020, GO:0016021,
331330	protein_coding_gene	Protein transport protein Sec24C	GO:0006886, GO:0006888, GO:0008270, GO:0015031, GO:0030127
331350	protein_coding_gene	Phosphatidylcholine:ceramide cholinephosphotransferase 1	GO:0016020, GO:0016021, GO:0016740
331360	protein_coding_gene	Phosphatidylcholine:ceramide cholinephosphotransferase 1	GO:0016020, GO:0016021, GO:0016740
331380	protein_coding_gene	J domain-containing protein	GO:0003676, GO:0016020, GO:0016021
331390	protein_coding_gene	PB1 domain-containing protein	GO:0000139, GO:0005737, GO:0006888, GO:0042802, GO:0048208
331400	protein_coding_gene	J domain-containing protein	None
331410	protein_coding_gene	Scavenger receptor class B member 1	GO:0016020, GO:0016021
331430	protein_coding_gene	Inositol polyphosphate 5-phosphatase OCRL-1	GO:0007165, GO:0016787, GO:0046856, GO:0052745
331440	protein_coding_gene	Cation-independent mannose-6-phosphate receptor	GO:0007041, GO:0038023, GO:0051219
331460	protein_coding_gene	Protein lap4	None

Th CLR region nine on scaffold 6 (figure 7.2.4.13A) is near a single spike in FST (figure 7.2.4.13B) and overlaps with reduced nucleotide diversity (figure 7.2.4.13C) and a fall in Tajima's D. I found 15 genes within this region (Gene ID: 0368920, 0368930, …, up to 0639060; table 7.2.4.5). One protein coding gene was predicted as G-protein coupled receptors. G-proteins are shown to regulate P450 gene expression in pesticide resistant of southern house mosquitoes (Li et al., 2015b).

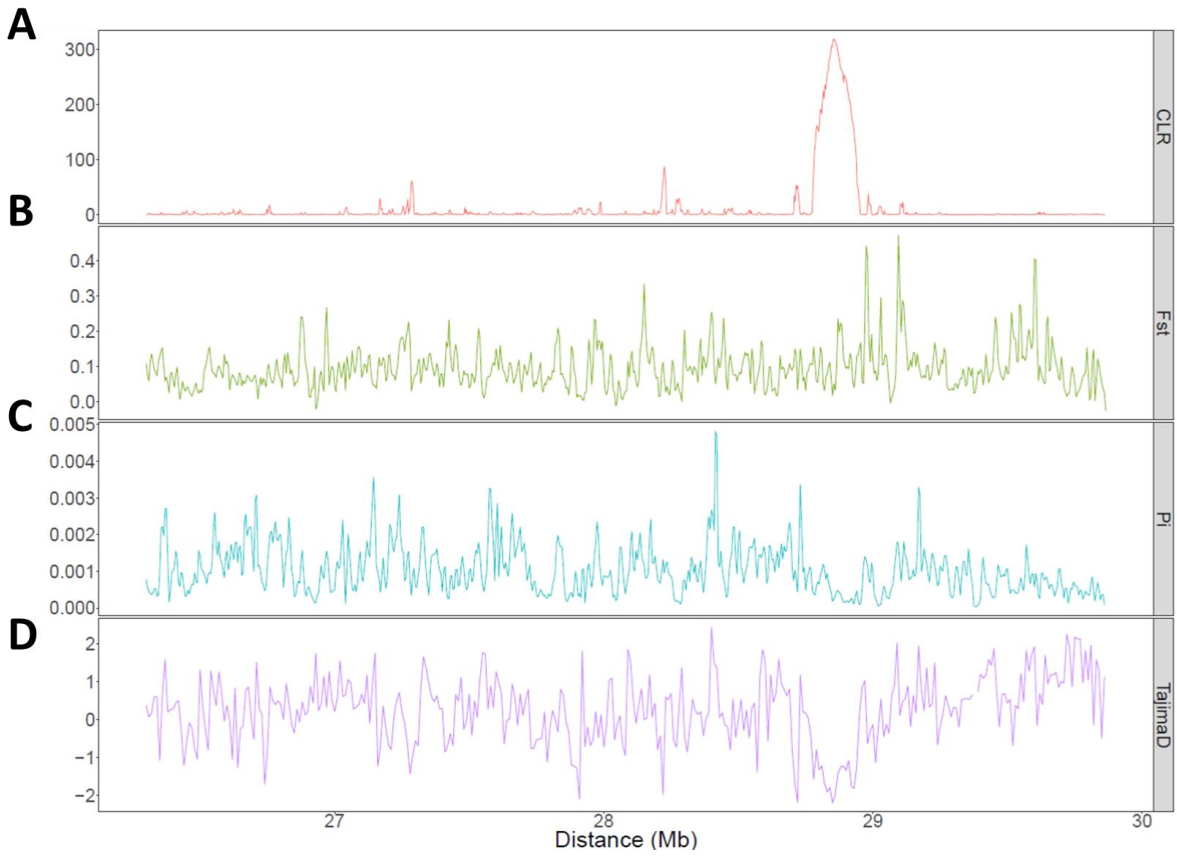


Figure 7.2.4.13 Summary statistics of putative selective sweep area on scaffold 6 (region 9 from group 4). Colours indicate the different scores. (**A**) Selective sweep likelihood quantified in CLR. (**B**) weighted population differentiation score in FST, between group 4 and all other genomic samples. (**C**)Nucleotide diversity (Pi). (**D**) Tajima's D statistic.

Table 7.2.4.5 Gene description of genes found within 5 Mb window near the selective sweep area on scaffold 2 (region 5 from group 4). Colours indicate the biotype of the gene. Blue = protein-coding gene and white = transposable element. The 4 ncRNAs were removed from the table.

Gene	Biotype	Description	Gene ontology term
368920	protein_coding_gene	Fibroblast growth factor receptor 4	GO:0000166, GO:0004672, GO:0004713, GO:0005524, GO:0006468, GO:0016020,
368930	protein_coding_gene	High affinity cationic amino acid transporter 1	GO:0016020, GO:0016021, GO:0022857, GO:0055085
368940	protein_coding_gene	Amino acid transporter	GO:0016020, GO:0016021, GO:0022857, GO:0055085
368970	protein_coding_gene	ACYPI53166 protein	GO:0016020, GO:0016021
368990	protein_coding_gene	G-protein coupled receptor	GO:0004888, GO:0004930, GO:0007166, GO:0007186, GO:0016020, GO:0016021
369000	protein_coding_gene	Beclin-1-like protein	GO:0006897, GO:0006914, GO:0032465
369010	protein_coding_gene	Nudix hydrolase domain-containing protein	GO:0016787, GO:0016818, GO:0046872
369020	protein_coding_gene	Pyridoxine-5'-phosphate oxidase	GO:0004733, GO:0008615, GO:0010181, GO:0016638, GO:0042823, GO:0048037, GO:0055114
369030	protein_coding_gene	Phosphoglycolate phosphatase	GO:0016311, GO:0016787, GO:0016791

369040	protein_coding_gene	Amino acid transporter	GO:0016020, GO:0016021, GO:0022857, GO:0055085
369060	protein_coding_gene	High affinity cationic amino acid transporter 1	GO:0016020, GO:0016021, GO:0022857, GO:0055085

7.2.4.7 Selective sweep regions of group 5

A total of five putative selective sweep regions were found within the 6 scaffolds of group 5. I selected two selective sweep regions that were supported by additional analyses. These were selective sweep regions two and three that were both found on scaffold 1.

The CLR region two on scaffold 1 (figure 7.2.4.14A) showed an increase in FST (figure 7.2.4.14B), increased and fluctuating nucleotide diversity (figure 7.2.4.14C) and a fall in Tajima's D. These data suggest a more ancient region within the selective sweep area with no shared ancestry between the other groups. I found 54 genes within this region (Gene ID: 0044580, 0044590, …, up to 0044970). All, were annotated as ncRNA, except for one gene that was annotated as a transposable element (GO:0003676, GO:0015074).

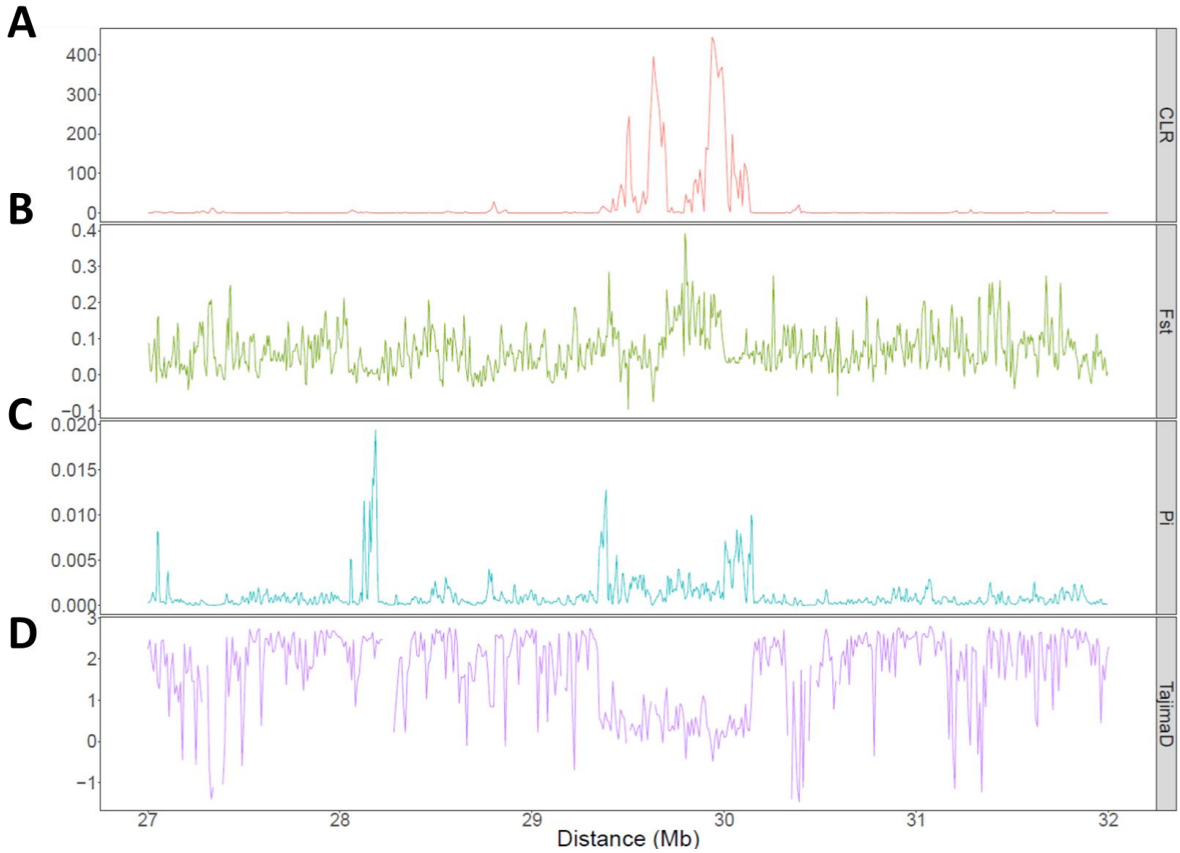


Figure 7.2.4.14 Summary statistics of putative selective sweep area on scaffold 1 (region 2 from group 5). Colours indicate the different scores. (**A**) Selective sweep likelihood quantified in CLR. (**B**) weighted population differentiation score in FST, between group 5 and all other genomic samples. (**C**)Nucleotide diversity (Pi). (**D**) Tajima's D statistic.

CLR region three on scaffold 1 (figure 7.2.4.15A) overlapped with an increase in FST, though other areas with higher FST were found (figure 7.2.4.15B). The region had fluctuating nucleotide diversity (figure 7.2.4.15C) and a fall in Tajima's D (figure 7.2.4.15D). I found 166 genes within this region (Gene ID: 0076720, 0076730, …, up to 0078370; table 7.2.4.6). In total 3 protein-coding genes were annotated, of which 2 had putative DNA-binding motifs (SAP domain-containing protein).

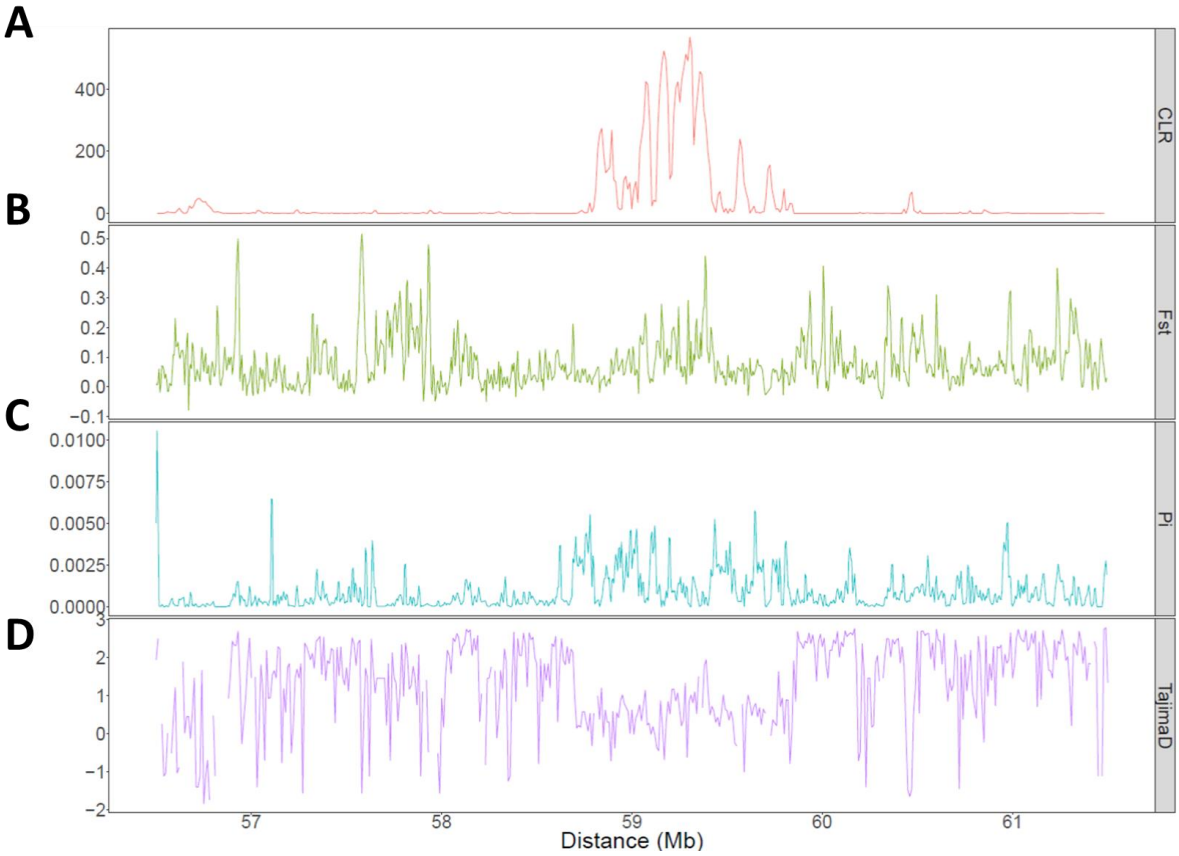


Figure 7.2.4.15 Summary statistics of putative selective sweep area on scaffold 1 (region 3 from group 5). Colours indicate the different scores. (**A**) Selective sweep likelihood quantified in CLR. (**B**) weighted population differentiation score in FST, between group 5 and all other genomic samples. (**C**)Nucleotide diversity (Pi). (**D**) Tajima's D statistic.

Table 7.2.4.6 Gene description of genes found within 5 Mb window near the selective sweep area on scaffold 2 (region 5 from group 4). Colours indicate the biotype of the gene. Blue = protein-coding gene and white = transposable element. 156 ncRNAs were removed from the table.

Gene	Biotype	Description	Gene ontology term
76760	protein_coding_gene	SAP domain-containing protein	None
76770	transposable_element_gene	Retrovirus-related Pol polyprotein	GO:0003676, GO:0015074
76780	protein_coding_gene	Retrotransposable element Tf2 protein type 2	None
76810	protein_coding_gene	SAP domain-containing protein	None
76820	transposable_element_gene	Retrovirus-related Pol polyprotein	GO:0003676, GO:0015074

77760	transposable_element_gene	CCHC-type domain-containing protein	GO:0003676, GO:0008270
77770	transposable_element_gene	DUF1758 domain-containing protein	None
77790	transposable_element_gene	SWIM-type domain-containing protein	GO:0008270
77800	transposable_element_gene	DUF1758 domain-containing protein	None
77840	transposable_element_gene	DUF1758 domain-containing protein	None

7.3 Discussion

In this chapter, I selected 101 of 131 *M. persicae* WGS for a population genomics study to understand the population dynamics of this aphid species. The selection of the samples was based on sequence quality, GC proportion and taxonomic annotation. For all 101 WGS, we had an average quality of over 37Q for the reads, an average GC proportion of around 27% and a high identity (≥99%) to the *M. persicae* genome sequence within our local database. Two samples were discarded for further analysis due to high heterozygosity and a high number of private alleles. Therefore, in total, 2.8 million SNPs of 99 samples were taken forward in this population study.

I tested four hypotheses on how this structure could be defined. **1.** Host-dependent, **2.** location-dependent, **3.** host/location dependent or **4.** no clear structure per host or location. Of these, the data show the most support for hypothesis 4, though one group was associated with a preference for *Nicotiana* sp. Moreover, I found that the samples may group into eight or ten groups that may reflect the population structure. Some groups were associated with the specific plant hosts, cabbage and canola and the countries Kenya and Australia. Denser sampling will be required to verify if this grouping is explained by plant host, country or both.

Five groups within the network were analysed for LD decay that showed some evidence of asexual reproduction of some groups. The current data shows evidence of between 8 and 10 *M. persicae* populations. Admixture and Bayesian clustering analyses showed evidence of recombination between most of these populations. The genetic differentiation between the selected groups was higher than observed in previous research projects that were based on microsatellite marker analyses (Zhao et al., 2015, Margaritopoulos et al., 2009, Wilson et al., 2004, Wilson et al., 2003, Blackman et al., 2007, Turcotte et al., 2013, Popkin et al., 2016, Monti et al., 2016, Sloane et al., 2001, Ramsey et al., 2007).

Based on a likelihood of a selective sweep of a > 200 CLR, in total 33 potential selective sweep regions were found in the genomes of *M. persicae*. There were 975 genes underlying these regions. Two of these were encoding putative effectors associated with one selective sweep region of group 4, which was interestingly the only group associated with *Nicotiana* as the host. FST, nucleotide diversity, and Tajima D analyses of the 33 selective sweep regions resulted in the short listing of eleven regions with the highest promise. Most of these selective sweep regions were group-specific, while one region was found in four of the five groups. Moreover, four putative selective sweep areas were found in group 4, which represents the *nicotianae* subgroup cluster. In one of these regions, 70 genes were found, and this included the two putative effector genes.

Some samples were other species or contaminated and needed to be discarded before conducting the population genomics analyses. In total, 30 samples were found that were not *M. persicae*, despite these sample being verified as *M. persicae* based on COI sequence analyses in chapter 6. The COI amplification procedure or the subsequent sequencing of COI may have been derived from contaminations, even though I included negative PCR controls in these PCRs. Nonetheless, because a lot of COI amplification was done and because PCR contaminations of fragments that are repeatedly amplified are common, COI PCR fragment contamination seems to be the most likely.

After quality control, I also discarded samples FRC09 and C25 because these showed high heterozygosity levels (>20%) with C25 at 60% and unique variance of over 90%. Nonetheless, the reads covered most of the *M. persicae* clone O genome, indicating that the sample was *M. persicae*. This sample may have been contaminated with another aphid species or may be derived from an individual that is a hybrid between M. persicae and another aphid. Hybridisation is well known in nature and is found in many organisms (Mallet, 2008). For instance, it is estimated that around 16% of the European butterfly species are derived from hybridisations with other species (Mallet, 2008, Mallet et al., 2011). Hybridisation is also known to occur in aphids (Dres and Mallet, 2002). For example, *R. padi* has hybridised with a closely related species (Delmotte et al., 2002, Delmotte et al., 2003). Hybridization within the species *M. persicae* is less investigated, but with its global distribution and its highly polyphagous behaviour, this aphid is likely to encounter closely related relatives to mate with. Further assessment if the samples C25 and FRC09 were derived from hybrids is needed.

With the *de novo* assembly BlobTools analyses, I found contigs that were most likely derived from bacteria that are potential (endo)symbionts of *M. persicae. M. persicae* has several (endo)symbionts, including *Hamiltonella defensa*, *Wolbachia*, *Regiella insecticola,* and *Serratia symbiotica* (Henry et al., 2015, Chen et al., 2019, Xu et al., 2020). It would be interesting to look further into the *M. persicae* genome sequence for the presence of bacterial (endo) symbionts. This is because the symbionts have important functions in insects. For example, a bacterial strain of *Ricketsia* increased the survival and fecundity of the whitefly *Bemisia tabaci* (Himler et al., 2011), *Spiroplasma* bacteria protect *Drosophila neostestacea* from nematode infections (Jaenike et al., 2010). For aphids, an U-type endosymbiont increases the fecundity of the pea aphid *A. pisum* upon host change (Tsuchida et al., 2004) and the endosymbiont *H. defensa* was found to reduce parasitism of *A. pisum* with the parasitic wasp *Aphidius ervi* (Oliver et al., 2003, Oliver et al., 2005).

Based on SNP analyses of WGS data, the lab-reared genotypes clone O, 4106a and FRC grouped into different clusters. In chapter 3, I found that clone O and 4106a have differences in their ability to colonize sugar beet. Moreover, clone O had similar performance rates on sugar beet as UK_SB and US1L did, and in this chapter, I found that the clones O, UK_SB and US1L cluster together. Therefore, results in chapter 3 and this chapter provide evidence that there are genetic differences that distinguish clones O, UK_SB and US1L from 4106a. These genetic differences include those that are associated with the ability of these aphids to colonize sugar beet.

Clones O, UK_SB and US1L all belong to group 3. These aphids were collected on separate occasions from the field. In group 3, there was a fourth aphid that was collected from potato in the UK in 2017. The four aphids in group 3 are extremely closely related and are likely from the same clonal lineage as is in agreement with the LD decay analysis, which showed an LD ratio of 0.25 that is higher than the ratios of the other groups. The four aphids in group 3 are likely to all genotype as clone O. Fenton et al. (2010) showed that a clonal population of a genotype O dominated within the UK between 2007 and 2009 and was described as a "super clone". Single asexual populations comprising 'super clones' can dominate due to having genetic characteristics that give high fitness gains, such as multiple insecticide resistances (Margaritopoulos et al., 2009). Beyond insecticide resistance, other environmental influences such as temperature or drought resistance can play a role. Among the 99 resequenced *M. persicae* genomes, 19 were derived from aphids collected in the UK from 2017-2019. Only four of these

19 clustered in group 3, and the 15 others distributed among several clusters in the network analysis. This indicates that the clone O genotype does not dominate in England. Rather, *M. persicae* populations in England consists of several genotypes.

Despite aphids having differences in their abilities to colonise sugar beet, there is no clear clustering of genomic SNPs and the ability to colonise sugar beet. This suggests that aphids colonise sugar beet even if they achieve higher reproductive rates on another crop species. Indeed, in chapter 3, I showed that Clones O, UK_SB and US1L do better on *B. rapa* than on sugar beet, and UK_SB was collected from sugar beet in England. In the 20 m^2 plots, most aphids were found on sugar beet. This may be due to a collection bias because the station managers that took care of the 20 m^2 plots were mostly interested in sugar beet. Nonetheless, it was also observed that the *B. rapa* plants on the 20 m^2 plots were often attacked by Cabbage white butterflies and flea beetles. Therefore, another possibility is that the aphids moved to sugar beet to escape from butterfly and flea beetle-infested *B. rapa* plants. Nonetheless, *M. persicae* with similar genotypes readily adjust to diverse plant species. This is in agreement with experiments in laboratory conditions that show that *M. persicae* clone O can be host-transferred to 9 plants species from 5 different families, including even a monocot, by changing gene expression in an apparent coordinated manner (Chen et al., 2020, Mathers et al., 2017).

None of the field samples grouped with group 4, which includes the aphids that are thought to be *M. persicae nicotianae* for which the genome sequence used in this analyses were taken from the Singh et al., (2020) study. These *M. persicae nicotianae* aphids have been collected from and reared on *Nicotiana tabacum* (tobacco) (Bass et al., 2014, Blackman et al., 2007, Eastop and Blackman, 2005). This subspecies has likely evolved to cope with chemical compounds, such as nicotine, from tobacco (Bass et al., 2014). Genomic differences between *M. persicae sensu stricto* and *nicotianae* include 1.5 million bases and the increased expression of an enzyme that detoxifies nicotine (Singh et al., 2020a). Within my analysis I found an LD decay within group 4, indicating that these individuals are derived from a single asexually reproducing population.

In the selective sweep analysis, I found 33 regions with a high likelihood of selective sweeps, either specific to a single group or found in multiple groups. I focussed on 11 of these regions, because these also showed features consistent with the presence of sweeps based on FST, nucleotide diversity and

Tajima D. Only one group (group 4) was associated with a specific plant host (*N. tabacum*) and one of the four promising selective sweeps identified in this group included two candidate effector genes. No candidate effector genes were found in any of the other sweeps. Whereas the functions of these genes as effectors have to be analysed, it is striking that these genes were found only in group 4.

Many of the genes that underlie the sweep areas are predicted to encode non-coding (nc) RNAs. In most genomes, the genes encoding ncRNAs have not been annotated, but because recent findings showed that *M. persicae* non-coding RNAs are transferred into the plant host upon feeding and have virulence functions (Chen et al., 2020), all ncRNAs genes were annotated in the *M. persicae* clone O genome (Singh et al., unpublished). The functions of ncRNAs are largely unknown, though they have been shown to be largely involved in the regulation of transcription and transcript splicing so far. Given that *M. persicae* undergoes massive transcriptome changes upon host change that includes ncRNA genes (Chen et al., 2020), further analyses of the ncRNA genes that underlie the sweeps appears warranted.

Several protein-coding genes that underlie the sweeps include those with functions in sensing and neural communication and that are used as insecticide targets and shown to be involved in the development of insecticide resistance (Ozoe, 2013, Gupta et al., 2019, Li and Liu, 2018). Therefore, some of the sweeps may have been the result of intense insecticide usage over the past decades. In agreement with this is that G-protein coupled receptor genes underlie the sweep regions of group 4. G-protein coupled receptor are known to function in P450 mediated permethrin resistance in mosquitoes (Li et al., 2015b). Group 4 consists of the *M. persicae nicotianae* population, and P450s are important in the colonisation of tobacco by mediating reduced sensitivity to the plant alkaloid nicotine (Bass et al., 2014).Taken together, the sweep analyses may have revealed some interesting genes to study further.

Chapter 8. General discussion

8.1 Summary of the research context

The green peach aphid *Myzus persicae* has a broad plant host range of over 400 species from over 40 families and is a significant insect pest of many crops. Furthermore, *M. persicae* transmits approximately 100 plant viruses and has developed resistance to more than 70 insecticides. So far, knowledge of crop resistance mechanisms to aphids is limited. More knowledge is needed about diversity within and between populations of aphid species as this information will enable the discovery of strategies that aphids may use to overcome plant defences and genetic resistance. Given that sugar beet production yields are challenged by insect-vectored viruses, particularly those transmitted by *M. persicae*, I investigated genome sequence variations of 99 aphids collected from *M. persicae* populations from sugar beet fields and other crops throughout the world.

In the years before neonicotinoid-like pesticides were introduced, aphids were causing a 20-30% yield loss of sugar beet in Europe. Sugar beet losses due to aphids and the viruses these aphids transmit were dramatically reduced to close to zero in the ± 15 years when neonicotinoid-like pesticides were regularly applied to sugar beet cultivation. During this time, the neonicotinoids were applied to sugar beet seed as a coating or sprayed in fields during the early stages of sugar beet growth when aphids, and particularly *M. persicae*, are known to be most abundant and to carry viruses (Smith and Hallsworth, 1990, Limburg et al., 1997, Schliephake et al., 2000, Kozłowska-Makulska et al., 2009). However, because of recent legislation introduced by the European Union to ban neonicotinoid pesticides, sugar beet crop losses due to infections by aphid-transmitted BYV and BMYV viruses have increased again and have been threatening the sugar beet industry of Europe (Hossain et al., 2020). Because of this, it has been extensively debated to reinstate the use of neonicotinoids as an emergency measure for certain crops, such as sugar beet, to make this crop economically valuable (European commission, 2021, Department for Environment Food and Rural affairs, 2021).

Because of the effectiveness of neonicotinoid-like pesticides for the control of aphids and aphid-transmitted viruses, investigations to increase sugar beet natural resistance to aphids and viruses were largely halted (Hossain et al., 2020). As such, there is no clear understanding of how far sugar beet can resist aphid and virus colonisation, including whether there is any level of resistance to these organisms in the existing sugar beet germplasm (Luterbacher et al., 2005). Moreover, genetic variations within and among

aphid, particularly *M. persicae* populations found on sugar beet, have not been investigated. Hence, it is currently unknown if there are aphid and virus resistance in available sugar beet germplasm and how far any identified resistance is stable, i.e. the capacity of aphids to overcome any resistance may be identified in sugar beet germplasm. To help to mitigate this situation, I started this iCASE studentship and collaboration with SESVanderHave to explore:

(i) If there is variation in resistance to *M. persicae* clones and genotypes and BMYV transmitted by this aphid species within SESVanderHave sugar beet germplasm and breeding lines;

(ii) The level of genome sequence variations among *M. persicae* populations found on sugar beet plots in countries across Europe.

8.2 Findings of this thesis

To address point (i) above, in **chapter 3**, I did firstly perform a host swap analysis to test if all four aphid clones could survive and reproduce on sugar beet. Secondly, I tested the various aphid clonal populations on multiple sugar beet lines with variable resistance levels against *M. persicae*. Thirdly, I assessed the virus transmission of BMYV on sugar beet lines with a single aphid clone (UK_SB). This showed that all aphid clones tested could survive and reproduce on sugar beet, but there is variable fecundity found on the multiple sugar beet lines dependent on the aphid clone. I showed that some sugar beet lines on which *M. persicae* had a lower fecundity. Moreover, I observed a similar pattern of reduced fecundity in the viral transmission of BMYV. Therefore, there is a level of resistance within the sugar beet germplasm to both *M. persicae* and BMYV. The work in this chapter also revealed that all four *M. persicae* clones, of which three clones were found very similar, are able to colonise and reproduce on sugar beet. However, the *M. persicae* clones have different fecundity rates per plant host, and these differences in fecundity were found for each *M. persicae* clone across the different sugar beet lines. These results suggest that there is genetic variation among *M. persicae* clones in the ability to colonise sugar beet lines. Given that *M. persicae* shows a variation in the ability to colonise sugar beet, it is possible that field populations of *M. persicae* may have adapted to colonise sugar beet. Perhaps similarly to *M. persicae nicotianae*, for which there is evidence of adaptation to colonise *Nicotiana* species, there may be a '*M. persicae betae*'

that has adapted to sugar beet. To investigate this further, it was necessary to genotype *M. persicae* clones and field-populations.

As a first effort to genotype *M. persicae*, I evaluated an existing genotyping method. In **chapter 4**, I determined whether genotypes could be distinguished from field-collected samples using microsatellite markers. This revealed that microsatellite markers were not specific enough to be used in large scale field studies. I showed that even with a more sensitive analysis and a higher resolution (down to 2 bp) of the amplicons' analyses, I was unable to distinguish between genotypes with high confidence. Altogether, the limited number of microsatellite markers used (MYZ9, M35, M40, M49, M63 and M86) were not ideal for genotyping *M. persicae* genotypes. Therefore, other genotyping methods had to be developed.

A number of genotyping methods are available for rapid screening, including Monsterplex (http://floodlightgenomics.com/), KASP markers, MYbaits (Li et al., 2013, Witek et al., 2016a, Witek et al., 2016b) and RAPDs (Criniti et al., 2006). Many of these require more information about variable sequences in genomes among populations and species. Moreover, with the advances of next-generation sequencing technologies, there is the possibility to increase the likelihood of finding genetic markers associated with a phenotype, such as the ability to colonise sugar beet, by sequencing and comparing entire genomes. To do this effectively, a high-quality genome assembly of *M. persicae* had to be generated first.

In **chapter 5**, I contributed optimisation of DNA extraction, library preparation and setting-up ONT MinION sequencing in the lab to obtain a chromosome-level assembly of *M. persicae* clone O. An assembly of this aphid was already available (Mathers et al., 2017), but it was highly fragmented, and segments were missing. The data collected in chapter 5 enabled the development of a protocol for the extraction of HMW DNA and for the sequencing of long-reads from (tiny) sap-feeding insects. The long-read sequence data generated in this chapter led to the generation of a chromosome-level *M. persicae* genome assembly. This put me in an advantageous position to conduct population genomics analyses on *M. persicae*.

In **chapter 6**, I worked with colleagues at SESVanderHave, Florimond Desprez, Syngenta and various research facilities around the world to sample *M. persicae* populations. With SESVanderHave, we set-up 20m^2 capture plots

with crop species in fields of Italy, France, Spain, Belgium, Germany, Netherlands, Austria, the UK and Russia. These plots include 5m^2 of each, sugar beet, Chinese cabbage, canola, and potato. Field station managers captured aphids on these crops and shipped these to me. To ship the single insects effectively, I designed the silica tube storage method and optimised DNA extractions from single dried aphids stored in these silica tubes. Because *M. persicae* can be confused with other species, I first genotyped the specimen. I received and found that 27% of the specimens were in fact, *M. persicae*. The study showed that sugar beet is being attacked by many aphid species, including *M. euphorbiae* and *A. fabae.* The finding of multiple other aphid species is important because all these aphids can transmit viruses. Nonetheless, *M. persicae* is the most economically relevant species because they are found early in the year, highly polyphagous and have high virus transmission rates (Smith and Hallsworth, 1990, Limburg et al., 1997, Schliephake et al., 2000, Kozłowska-Makulska et al., 2009, Hossain et al., 2020). Altogether, the work in chapter 6 gave the materials to start a whole-genome resequencing project of *M. persicae* populations collected worldwide.

In **chapter 7**, I resequenced the genomes of 119 *M. persicae* samples. Most of these came from sugar beet, but there were also specimens collected from canola, Chinese cabbage, potato, tobacco and peach. The SNP clustering showed that *M. persicae* collected from tobacco clustered separately. These genotypes were classified as *nicotianae* subspecies in Singh et al. (2020a). Therefore, this thesis's finding was in agreement with observations from others that there may be a *M. persicae nicotanae* subspecies. However, no such cluster was found to be associated with sugar beet. In fact, *M. persicae* found on sugar beet were distributed among all the network analysis clusters. This indicates that *M. persicae* on sugar beet can be diverse. Nonetheless, the genome scans revealed several sweeps within specific clusters. These sweeps appear to be associated with specific genes.

In conclusion, there is a level of resistance in sugar beet germplasm to some *M. persicae* clones. However, field populations of *M. persicae* found on sugar beet are geneticly highly diverse. Altogether, this suggests that any sugar beet resistance to a clone may be overcome rather quickly by other clones. For stable resistance of sugar beet, perhaps more work is needed to identify sugar beet susceptibility genes required for *M. persicae* colonisation.

8.3. The start for the search to obtain stable plant resistance to aphids

Several R-genes were identified for aphids and were mapped in crops. For example, RAG genes against the aphid *Aphis glycines* in soybean (Hill et al., 2012), Nr genes against the aphid *Nasonovia ribisnigri* in lettuce (Helden et al., 1993), *ER1* genes against *Dysaphis plantaginea* and *Eriosoma lanigerum,* and Dn genes against the aphid *Diuraphis noxia* in wheat (Liu et al., 2005). Furthermore, a few CC-NBS-LRRs were cloned, *Mi1-2* against *M.* euphorbiae (Nombela et al., 2003), VAT against *A. gossypii* (Boissot et al., 2016a) and AIN against *Acyrthosiphon kondoi* (Klingler et al., 2009). However, all known R-genes have already been broken by resistance-breaking aphid biotypes.

The occurrence of resistance-breaking biotypes had been recorded in aphids, for example, the aphid *Aphis glycines,* which has biotypes 2 and 3 that overcome Rag1 and Rag2 resistance in soybean (Mian et al., 2008, Hill et al., 2012); *Nasonovia ribisnigri* biotype Nr1 has overcome Nr resistance in lettuce (Thabuis et al., 2011); *Schizaphis graminum* biotypes have overcome resistances of wheat & sorghum (Curvetto and Webster, 1989, Kindler et al., 2001); *M. euphorbiae* biotypes can overcome *Mi1-2* resistance of tomato (Goggin et al., 2001); multiple *A. gossypii* biotypes overcome VAT resistance of melon (Boissot et al., 2016b); biotypes of the *Diuraphis noxia* overcome *Dn4* resistance of wheat (Basky et al., 2001, Collins et al., 2005, Peng et al., 2009); and *Dysaphis plantaginea* and *Eriosoma lanigerum* overcome *ER1*, *ER2* and *ER3* resistance in apple (Rath-Morris et al., 1998, Young et al., 1982). Hence, more investigation of the nature of genetic diversity within an aphid species is needed.

To obtain stable aphid resistance, it will be necessary to do more investigations on the aphids themselves. Such investigations may result in the identification of the aphid 'Achilles heel' via the following approaches: (1) characterization of conserved effector genes, e.g. those required for aphid feeding and the subsequent identification of (variation in) plant factors or processes targeted by these effectors; (2) targeting the aphid bacterial symbionts, which are required for aphid survival and enhancing their fitness. Aphid genes encoding for effectors and mediating aphid interactions with the symbionts could be targeted by specific plant factors and knocked down via RNA interference (RNAi).

M. persicae has a plethora of effector genes that they deploy for the colonisation of plant hosts (Bos et al., 2010). Some of these are found to be more conserved among aphid species and even among multiple species of the order Hemiptera. Nevertheless, currently, only a few conserved effector genes were characterised and tested for their contributions to aphid colonisation. Effector genes such as Mp10 and MpC002 were shown to be necessary for aphid colonisation of plants. Mp10 is found to suppress the *N. benthamiana* reactive oxygen species (ROS) burst that results in chlorosis and local cell death in response to flg22 and aphid elicitors (Bos et al., 2010, Rodriguez et al., 2014a) and the effector MpC002 was found to improve *M. persicae* colonisation (Pitino and Hogenhout, 2013, Bos et al., 2010). Mp1 is interacting with the plant hosts' Vacuolar Protein Sorting Associated Protein 52 (*VPS52*). By reducing the virulent factor VPS52 levels for improved fecundity was improved (Rodriguez et al., 2017). Yet, the number of aphid effectors for which functions are known is low. The genomic datasets generated in this thesis will facilitate the identification of effector genes that are conserved versus those that are variable.

8.4. Genotyping

The data that are generated in this thesis can be further used for developing new genetic markers to allow the opportunity to study the diversity of *M. persicae*. Genetic markers are useful for 1. Lowering the costs, 2. Enabling high-throughput processing of many individuals, and 3. reducing computational processing requirements. Genotyping technologies using Monsterplex, Mybaits or Kompetitive Allele-Specific PCR (KASP) markers could facilitate high-throughput genotyping of aphid populations.

Monsterplex Floodlight Genomics uses an optimised Hi-Plex approach to amplify targets in a single multiplex reaction. Monsterplex enables sequences of over 100 genetic targets from a single reaction used in 1000s of samples at the same time in high-throughput sequencing. This will enable accurate Illumina sequencing of variable regions but is limited to smaller fragments (http://floodlightgenomics.com/). The **MYbaits** technology allows sequence-specific region/genes of an organism via the next-generation sequencing techniques (Li et al., 2013, Witek et al., 2016a, Witek et al., 2016b). It uses biotinylated RNA baits complementary to the target sequence enabling the sequencing of specific DNA fragments from many samples. This will result in a greater number of samples that could be screened in similar amounts of

time and money as one whole-genome sequence (Faircloth et al., 2015, Nistelberger et al., 2016, Witek et al., 2016a, Witek et al., 2016b).

KASP is a uniplex SNP genotyping technology. It enables the screening of SNP KASP markers in up to 1536-well plates, and it is using accessible technologies used in many breeding companies. Therefore, developing KASP markers of *M. persicae* populations found worldwide could allow breeding companies to screen for populations found in locations that they grow their crop to deploy the most suitable resistance variety in that region. Nevertheless, KASP markers are limited to the segregation of no more than two nucleotides (homozygote or heterozygote), and therefore, many polymorphisms could be missed in screening with KASP markers.

Furthermore, the *M. persicae* genome sequence dataset will also enable the optimisation and development of additional **microsatellite markers**. Locations of microsatellites can be predicted from the genomic sequences, and primers can be designed accordingly. This would improve the selection and increase the development of multiple more microsatellite markers, such as previously described in Sloane et al. (2001) and Wilson et al. (2004).

8.5 Output of this thesis

I generated 3 open-access protocols

1. S. Mugford, **R.H.M. Wouters**, T.C. Mathers, S. Hogenhout, 2020. High quality DNA extraction from very small individual insects. **protocols.io**
2. **R.H.M. Wouters**, S. Mugford, S. Hogenhout. 2020. Ambient sample storage system of field-collected insect samples for genomics. **protocols.io**
3. **R.H.M. Wouters**, S. Mugford, R. Biello, D. Heavens, S. Hogenhout. 2020. Extraction of high molecular weight DNA from aphids and other sap-feeding insects for long-read sequencing. **protocols.io**

I was co-author on the following papers

1. S.-T. Cho, A. Zwolińska, W. Huang, **R.H.M. Wouters**, S.T. Mugford, S.A. Hogenhout, C.-H. Kuo. 2020. Complete Genome Sequence of "Candidatus Phytoplasma*" RP166, a Plant Pathogen Associated

with Rapeseed Phyllody Disease in Poland. **Microbiology Resource Announcements**. 9:e00760-20

2. T.C. Mathers, **R.H.M. Wouters**, S.T. Mugford, D. Swarbreck, C. Van Oosterhout and S.A. Hogenhout. 2020. Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. **MBE** msa246 and a preprint is found in **BioRXiv**

Bibliography

(MD), B. 1988. *National Center for Biotechnology Information (NCBI)* [Online]. National Library of Medicine (US), National Center for Biotechnology Information. Available: https://www.ncbi.nlm.nih.gov/ [Accessed October 2020].

ADAK, T., KUMAR, J., SHAKIL, N. A. & WALIA, S. 2012. Development of controlled release formulations of imidacloprid employing novel nano-ranged amphiphilic polymers. *J Environ Sci Health B,* 47**,** 217-25.

AGRANOVSKY, A. A., KOONIN, E. V., BOYKO, V. P., MAISS, E., FROTSCHL, R., LUNINA, N. A. & ATABEKOV, J. G. 1994. Beet yellows closterovirus: complete genome structure and identification of a leader papain-like thiol protease. *Virology,* 198**,** 311-24.

AHMAN, I., KIM, S. Y. & ZHU, L. H. 2019. Plant Genes Benefitting Aphids-Potential for Exploitation in Resistance Breeding. *Front Plant Sci,* 10**,** 1452.

AKBARI, A., VITTI, J. J., IRANMEHR, A., BAKHTIARI, M., SABETI, P. C., MIRARAB, S. & BAFNA, V. 2018. Identifying the favored mutation in a positive selective sweep. *Nat Methods,* 15**,** 279-282.

ALAVANJA, M. C. R. 2009. Pesticides Use and Exposure Extensive Worldwide. . *Reviews on environmental health. ,* 24**,** 303–309.

ALEXANDER, D. H. & LANGE, K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics,* 12**,** 246.

ALFORD, L., HUGHES, G. E., BLACKBURN, T. M. & BALE, J. S. 2012. Walking speed adaptation ability of Myzus persicae to different temperature conditions. *Bull Entomol Res,* 102**,** 303-13.

ALLENDORF, F. W., HOHENLOHE, P. A. & LUIKART, G. 2010. Genomics and the future of conservation genetics. *Nat Rev Genet,* 11**,** 697-709.

ANDRET-LINK, P. & FUCHS, M. 2005. Transmission specificity of plant viruses by vectors. *Journal of Plant Pathology,* 87**,** 153-165.

AREND, A. 2003. The possibility of Nasonovia ribisnigri resistance breaking biotype development due to plant host resistance: a literature study. *Eucarpia Leafy Vegetables***,** 75-81.

ASAI, D. J. & WILKES, D. E. 2004. The dynein heavy chain family. *J Eukaryot Microbiol,* 51**,** 23-9.

ATAMIAN, H. S., CHAUDHARY, R., CIN, V. D., BAO, E., GIRKE, T. & KALOSHIAN, I. 2013. In planta expression or delivery of potato aphid Macrosiphum euphorbiae effectors Me10 and Me23 enhances aphid fecundity. *Mol Plant Microbe Interact,* 26**,** 67-74.

AUSUBEL, F. M., KATAGIRI, F., MINDRINOS, M. & GLAZEBROOK, J. 1995. Use of Arabidopsis thaliana defense-related mutants to dissect the plant response to pathogens. *Proc Natl Acad Sci U S A,* 92**,** 4189-96.

BAGGS, E., DAGDAS, G. & KRASILEVA, K. V. 2017. NLR diversity, helpers and integrated domains: making sense of the NLR IDentity. *Curr Opin Plant Biol,* 38**,** 59-67.

BALE, J. S. 1987. Insect cold hardiness: Freezing and supercooling—An ecophysiological perspective. *Journal of Insect Physiology,* 33**,** 899-908.

BANDELT, H. J., FORSTER, P. & ROHL, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol,* 16**,** 37-48.

BARNES, M. R. & BREEN, G. 2010. genetic variation methods and protocols. 628.

BARTON, N. H. 2000. Genetic hitchhiking. *Philos Trans R Soc Lond B Biol Sci,* 355**,** 1553-62.

BASKY, Z., HOPPER, K. R., JORDAAN, J. & SAAYMAN, T. 2001. Biotypic differences in Russian wheat aphid (Diuraphis noxia) between South African and Hungarian agro-ecosystems. *Agriculture, Ecosystems & Environment,* 83**,** 121-128.

BASS, C., PUINEAN, A. M., ANDREWS, M., CUTLER, P., DANIELS, M., ELIAS, J., PAUL, V. L., CROSSTHWAITE, A. J., DENHOLM, I., FIELD, L. M., FOSTER, S. P., LIND, R., WILLIAMSON, M. S. & SLATER, R. 2011. Mutation of a nicotinic acetylcholine receptor beta subunit is associated with resistance to neonicotinoid insecticides in the aphid Myzus persicae. *BMC Neurosci,* 12**,** 51.

BASS, C., PUINEAN, A. M., ZIMMER, C. T., DENHOLM, I., FIELD, L. M., FOSTER, S. P., GUTBROD, O., NAUEN, R., SLATER, R. & WILLIAMSON, M. S. 2014. The evolution of insecticide resistance in the peach potato aphid, Myzus persicae. *Insect Biochem Mol Biol,* 51**,** 41-51.

BATES, S. L., ZHAO, J. Z., ROUSH, R. T. & SHELTON, A. M. 2005. Insect resistance management in GM crops: past, present and future. *Nat Biotechnol,* 23**,** 57-62.

BAXTER, S. W., ZHAO, J. Z., GAHAN, L. J., SHELTON, A. M., TABASHNIK, B. E. & HECKEL, D. G. 2005. Novel genetic basis of field-evolved resistance to Bt toxins in Plutella xylostella. *Insect Mol Biol,* 14**,** 327-34.

BBRO 2020. *https://bbro.co.uk/,* Accessed September 2020.

BEGUN, D. J. & AQUADRO, C. F. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. *Nature,* 356**,** 519-20.

BELEFANT-MILLER, H., PORTER, D. R., PIERCE, M. L. & MORT, A. J. 1994. An Early Indicator of Resistance in Barley to Russian Wheat Aphid. *Plant Physiol,* 105**,** 1289-1294.

BELKHADIR, Y., SUBRAMANIAM, R. & DANGL, J. L. 2004. Plant disease resistance protein signaling: NBS-LRR proteins and their partners. *Curr Opin Plant Biol,* 7**,** 391-9.

BELL, J. R., BLUMGART, D. & SHORTALL, C. R. 2020. Are insects declining and at what rate? An analysis of standardised, systematic catches of aphid and moth abundances across Great Britain. *Insect Conserv Divers,* 13**,** 115-126.

BERNAYS, E. & GRAHAM, M. 1988. On the Evolution of Host Specificity in Phytophagous Arthropods. *Ecology,* 69**,** 886-892.

BHARGAVA, A. & FUENTES, F. F. 2010. Mutational dynamics of microsatellites. *Mol Biotechnol,* 44**,** 250-66.

BIELLO, R., MATHERS, T. C., MUGFORD, S. T., LIU, Q., RODRIGUES, A. B. S., NETO, A. C. & HOGENHOUT, S. 2020. Draft genome assembly version 1 of the meadow spittlebug Philaenus spumarius (Linnaeus, 1758) (Hemiptera, Aphrophoridae). *Zenodo*.

BIELLO, R., SINGH, A., GODFREY, C. J., FERNANDEZ, F. F., MUGFORD, S. T., POWELL, G., HOGENHOUT, S. A. & MATHERS, T. C. 2021. A chromosome-level genome assembly of the woolly apple aphid, Eriosoma lanigerum Hausmann (Hemiptera: Aphididae). *Mol Ecol Resour,* 21**,** 316-326.

BLACKMAN, R. L. 1974. Life-cycle variation of Myzus persicae (Sulz.) (Hom., Aphididae) in different parts of the world, in relation to genotype and environment. *Bulletin of Entomological Research,* 63**,** 595.

BLACKMAN, R. L. 1987. Morphological Discrimination of a Tobacco-Feeding Form from Myzus-Persicae (Sulzer) (Hemiptera, Aphididae), and a Key to New World Myzus (Nectarosiphon) Species. *Bulletin of Entomological Research,* 77**,** 713-730.

BLACKMAN, R. L. & EASTOP, V. F. 2000a. *Aphids on the World's Crops.* Wiley: Chichester.

BLACKMAN, R. L. & EASTOP, V. F. 2000b. *Aphids on the world's crops : an identification and information guide,* Chichester, West Sussex, England; New York, Wiley.

BLACKMAN, R. L., MALARKY, G., MARGARITOPOULOS, J. T. & TSITSIPIS, J. A. 2007. Distribution of common genotypes of Myzus persicae (Hemiptera: Aphididae) in Greece, in relation to life cycle and host plant. *Bull Entomol Res,* 97**,** 253-63.

BOISSOT, N., SCHOENY, A. & VANLERBERGHE-MASUTTI, F. 2016a. Vat, an Amazing Gene Conferring Resistance to Aphids and Viruses They Carry: From Molecular Structure to Field Effects. *Front Plant Sci,* 7**,** 1420.

BOISSOT, N., THOMAS, S., CHOVELON, V. & LECOQ, H. 2016b. NBS-LRR-mediated resistance triggered by aphids: viruses do not adapt; aphids adapt via different mechanisms. *BMC Plant Biol,* 16**,** 25.

BOLLER, T. & FELIX, G. 2009. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu Rev Plant Biol,* 60**,** 379-406.

BONAVENTURE, G. 2011. The Nicotiana attenuata LECTIN RECEPTOR KINASE 1 is involved in the perception of insect feeding. *Plant Signal Behav,* 6**,** 2060-3.

BONMATIN, J. M., GIORIO, C., GIROLAMI, V., GOULSON, D., KREUTZWEISER, D. P., KRUPKE, C., LIESS, M., LONG, E., MARZARO, M., MITCHELL, E. A., NOOME, D. A., SIMON-DELSO, N. & TAPPARO, A. 2015. Environmental fate and exposure; neonicotinoids and fipronil. *Environ Sci Pollut Res Int,* 22**,** 35-67.

BOS, J. I., PRINCE, D., PITINO, M., MAFFEI, M. E., WIN, J. & HOGENHOUT, S. A. 2010. A functional genomics approach identifies candidate effectors from the aphid species Myzus persicae (green peach aphid). *PLoS Genet,* 6**,** e1001216.

BOUWMEESTER, K. & GOVERS, F. 2009. Arabidopsis L-type lectin receptor kinases: phylogeny, classification, and expression profiles. *J Exp Bot,* 60**,** 4383-96.

BRADIC, M. & CARLTON, J. M. 2018. Does the common sexually transmitted parasite Trichomonas vaginalis have sex? *PLoS Pathog,* 14**,** e1006831.

BRUINS, M. 2020. Sugar Rush: What it Takes to Develop a New Sugar Beet Variety.

CASTALANELLI, M. A., SEVERTSON, D. L., BRUMLEY, C. J., SZITO, A., FOOTTIT, R. G., GRIMM, M., MUNYARD, K. & GROTH, D. M. 2010. A rapid non-destructive DNA extraction method for insects and other arthropods. *Journal of Asia-Pacific Entomology,* 13**,** 243-248.

CASTRO, S., ROMEIRAS, M. M., CASTRO, M., DUARTE, M. C. & LOUREIRO, J. 2013. Hidden diversity in wild Beta taxa from Portugal: insights from genome size and ploidy level estimations using flow cytometry. *Plant Sci,* 207**,** 72-8.

CHANDRE, F., DARRIER, F., MANGA, L., AKOGBETO, M., FAYE, O., MOUCHET, J. & GUILLET, P. 1999. Status of pyrethroid resistance in Anopheles gambiae sensu lato / F. Chandre ... [et al.]. *Bulletin of the World Health Organization : the International Journal of Public Health 1999 ; 77(3) : 230-234*.

CHARLESWORTH, B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet,* 10**,** 195-205.

CHEN, H., RANGASAMY, M., TAN, S. Y., WANG, H. & SIEGFRIED, B. D. 2010. Evaluation of five methods for total DNA extraction from western corn rootworm beetles. *PLoS One,* 5**,** e11963.

CHEN, R., SU, X., CHEN, J., JIANG, L. & QIAO, G. X. 2019. Wolbachia Infection in Two Species: Novel Views on the Colonization Ability of Wolbachia in Aphids. *Environ Entomol,* 48**,** 1388-1393.

CHEN, X., SHANG, J., CHEN, D., LEI, C., ZOU, Y., ZHAI, W., LIU, G., XU, J., LING, Z., CAO, G., MA, B., WANG, Y., ZHAO, X., LI, S. & ZHU, L. 2006. A B-lectin receptor kinase gene conferring rice blast resistance. *Plant J,* 46**,** 794-804.

CHEN, Y., SINGH, A., KAITHAKOTTIL, G. G., MATHERS, T. C., GRAVINO, M., MUGFORD, S. T., VAN OOSTERHOUT, C., SWARBRECK, D. & HOGENHOUT, S. A. 2020. An aphid RNA transcript migrates systemically within plants and is a virulence factor. *Proc Natl Acad Sci U S A,* 117**,** 12763-12771.

CHENG, X., WU, Y., GUO, J., DU, B., CHEN, R., ZHU, L. & HE, G. 2013. A rice lectin receptor-like kinase that is involved in innate immune responses also contributes to seed germination. *Plant J,* 76**,** 687-98.

CHEWAPREECHA, C., HOLDEN, M. T., VEHKALA, M., VALIMAKI, N., YANG, Z., HARRIS, S. R., MATHER, A. E., TUANYOK, A., DE SMET, B., LE HELLO, S., BIZET, C., MAYO, M., WUTHIEKANUN, V., LIMMATHUROTSAKUL, D., PHETSOUVANH, R., SPRATT, B. G., CORANDER, J., KEIM, P., DOUGAN, G., DANCE, D. A., CURRIE, B. J., PARKHILL, J. & PEACOCK, S. J. 2017. Global and regional dissemination and evolution of Burkholderia pseudomallei. *Nat Microbiol,* 2**,** 16263.

CHO, S. T., ZWOLINSKA, A., HUANG, W., WOUTERS, R. H. M., MUGFORD, S. T., HOGENHOUT, S. A. & KUO, C. H. 2020. Complete Genome Sequence of "Candidatus Phytoplasma asteris" RP166, a Plant Pathogen Associated with Rapeseed Phyllody Disease in Poland. *Microbiol Resour Announc,* 9.

COCKBAIN, A. J., GIBBS, A. J. & HEATHCOTE, G. D. 1963. Some factors affecting the transmission of sugar-beet mosaic and pea mosaic viruses by Aphis fubue and Myzus persicae. *Annals of Applied Biology,* 52**,** 133-143.

COLEMAN, A. D., MUGFORD, S. T. & HOGENHOUT, S. A. 2016. Silencing of Aphid Genes by dsRNA Feeding from Plants. *In:* CZOSNEK, H. & GHANIM, M. (eds.) *Management of Insect Pests to Agriculture: Lessons Learned from Deciphering their Genome, Transcriptome and Proteome.* Cham: Springer International Publishing.

COLLINS, M. B., HALEY, S. D., PEAIRS, F. B. & RUDOLPH, J. B. 2005. Biotype 2 Russian Wheat Aphid Resistance among Wheat Germplasm Accessions. *Crop Science,* 45**,** 1877-1880.

COOK, D. E., MESARICH, C. H. & THOMMA, B. P. 2015. Understanding plant immunity as a surveillance system to detect invasion. *Annu Rev Phytopathol,* 53**,** 541-63.

COOKE, D. A. & SCOTT, J. 2012. *The sugar beet crop*, Springer Science & Business Media.

COOPER, W. R., JIA, L. & GOGGIN, L. 2005. Effects of jasmonate-induced defenses on root-knot nematode infection of resistant and susceptible tomato cultivars. *J Chem Ecol,* 31**,** 1953-67.

CRINITI, A., MAZZONI, E., PECCHIONI, N., RAU, D., CASSANELLI, S., BIZZARO, D. & MANICARDI, G. 2006. Genetic variability among different Italian populations of the aphidMyzus persicae. *Caryologia,* 59**,** 326-333.

CUI, N., LU, H., WANG, T., ZHANG, W., KANG, L. & CUI, F. 2019. Armet, an aphid effector protein, induces pathogen resistance in plants by promoting the accumulation of salicylic acid. *Philos Trans R Soc Lond B Biol Sci,* 374**,** 20180314.

CURVETTO, R. O. & WEBSTER, J. A. 1989. Resistance mechanisms of PI 240675 Rey to biotype F greenbug. *Southwestern Entomologist,* 23**,** 97-103.

CUTTER, A. D., BAIRD, S. E. & CHARLESWORTH, D. 2006. High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of Caenorhabditis remanei. *Genetics,* 174**,** 901-13.

D'ALESSANDRO, M. & TURLINGS, T. C. 2006. Advances and challenges in the identification of volatiles that mediate interactions among plants and arthropods. *Analyst,* 131**,** 24-32.

DAVID, S., RUSNIOK, C., MENTASTI, M., GOMEZ-VALERO, L., HARRIS, S. R., LECHAT, P., LEES, J., GINEVRA, C., GLASER, P., MA, L., BOUCHIER, C., UNDERWOOD, A., JARRAUD, S., HARRISON, T. G., PARKHILL, J. & BUCHRIESER, C. 2016. Multiple major disease-associated clones of Legionella pneumophila have emerged recently and independently. *Genome Res,* 26**,** 1555-1564.

DE LA PAZ CELORIO-MANCERA, M., WHEAT, C. W., VOGEL, H., SODERLIND, L., JANZ, N. & NYLIN, S. 2013. Mechanisms of macroevolution: polyphagous plasticity in butterfly larvae revealed by RNA-Seq. *Mol Ecol,* 22**,** 4884-95.

DE SMET, I., VASSILEVA, V., DE RYBEL, B., LEVESQUE, M. P., GRUNEWALD, W., VAN DAMME, D., VAN NOORDEN, G., NAUDTS, M., VAN ISTERDAEL, G., DE CLERCQ, R., WANG, J. Y., MEULI, N., VANNESTE, S., FRIML, J., HILSON, P., JURGENS, G., INGRAM, G. C., INZE, D., BENFEY, P. N. & BEECKMAN, T. 2008. Receptor-like kinase ACR4 restricts formative cell divisions in the Arabidopsis root. *Science,* 322**,** 594-7.

DE SMET, I., VOSS, U., JURGENS, G. & BEECKMAN, T. 2009. Receptor-like kinases shape the plant. *Nat Cell Biol,* 11**,** 1166-73.

DEDRYVER, C. A., LE RALEC, A. & FABRE, F. 2010. The conflicting relationships between aphids and men: a review of aphid damage and control strategies. *C R Biol,* 333**,** 539-53.

DELMOTTE, F., LETERME, N., GAUTHIER, J. P., RISPE, C. & SIMON, J. C. 2002. Genetic architecture of sexual and asexual populations of the aphid Rhopalosiphum padi based on allozyme and microsatellite markers. *Mol Ecol,* 11**,** 711-23.

DELMOTTE, F., SABATER-MUNOZ, B., PRUNIER-LETERME, N., LATORRE, A., SUNNUCKS, P., RISPE, C. & SIMON, J. C. 2003. Phylogenetic evidence for hybrid origins of asexual lineages in an aphid species. *Evolution,* 57**,** 1291-303.

DEPARTMENT FOR ENVIRONMENT FOOD AND RURAL AFFAIRS. 2021. *Statement on the decision to issue – with strict conditions – emergency authorisation to use a product containing a neonicotinoid to treat sugar beet seed in 2021* [Online]. [Accessed February 2021].

DEVONSHIRE, A. L., DENHOLM, I., PICKETT, J. A., DEVONSHIRE, A. L., FIELD, L. M., FOSTER, S. P., MOORES, G. D., WILLIAMSON, M. S. & BLACKMAN, R. L. 1998. The evolution of insecticide resistance in the peach–potato aphid, Myzus persicae. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences,* 353**,** 1677-1684.

DEWAR, A., HAYLOCK, L. & ECCLESTONE, P. 1996. Strategies for controlling virus yellows in sugar beet. *British Sugar Beet Review* 64**,** 44-48.

DIXON, A. 1998. *Aphid Ecology: An Optimization Approach. ,* London, Springer.

DODDS, P. N. & RATHJEN, J. P. 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet,* 11**,** 539-48.

DOGIMONT, C., BENDAHMANE, A., CHOVELON, V. & BOISSOT, N. 2010. Host plant resistance to aphids in cultivated crops: genetic and molecular bases, and interactions with aphid populations. *C R Biol,* 333**,** 566-73.

DOUGLAS, A. E. 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. *Annu Rev Entomol,* 43**,** 17-37.

DRES, M. & MALLET, J. 2002. Host races in plant-feeding insects and their importance in sympatric speciation. *Philos Trans R Soc Lond B Biol Sci,* 357**,** 471-92.

DU, B., CHEN, R., GUO, J. & HE, G. 2020. Current understanding of the genomic, genetic, and molecular control of insect resistance in rice. *Molecular Breeding,* 40.

DU, B., ZHANG, W., LIU, B., HU, J., WEI, Z., SHI, Z., HE, R., ZHU, L., CHEN, R., HAN, B. & HE, G. 2009. Identification and characterization of Bph14, a gene conferring resistance to brown planthopper in rice. *Proc Natl Acad Sci U S A,* 106**,** 22163-8.

EASTOP, V. F. & BLACKMAN, R. L. 2005. Some new synonyms in Aphididae (Hemiptera : Sternorrhyncha). *Zootaxa,* 1089**,** 1-36.

ECKARDT, N. A. 2002. Plant Disease Susceptibility Genes? *The Plant Cell Online,* 14**,** 1983-1986.

ELZINGA, D. A., DE VOS, M. & JANDER, G. 2014. Suppression of plant defenses by a Myzus persicae (green peach aphid) salivary effector protein. *Mol Plant Microbe Interact,* 27**,** 747-56.

ERIKSSON, A. & MANICA, A. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc Natl Acad Sci U S A,* 109**,** 13956-60.

ESCUDERO-MARTINEZ, C., RODRIGUEZ, P. A., LIU, S., SANTOS, P. A., STEPHENS, J. & BOS, J. I. B. 2020. An aphid effector promotes barley susceptibility through suppression of defence gene expression. *J Exp Bot,* 71**,** 2796-2807.

EUROPEAN COMMISSION. 2021. *Pesticides emergency authorisations* [Online]. Available: https://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/ppp/pppeas/screen/home [Accessed January 2021].

FAIRCLOTH, B. C., BRANSTETTER, M. G., WHITE, N. D. & BRADY, S. G. 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Mol Ecol Resour,* 15**,** 489-501.

FENTON, B., MALLOCH, G., WOODFORD, J. A. T., FOSTER, S. P., ANSTEAD, J., DENHOLM, I., KING, L. & PICKUP, J. 2005. The attack of the clones: tracking the movement of insecticide-resistant peach–potato aphids Myzus persicae (Hemiptera: Aphididae). *Bulletin of Entomological Research,* 95.

FENTON, B., MARGARITOPOULOS, J. T., MALLOCH, G. L. & FOSTER, S. P. 2010. Micro-evolutionary change in relation to insecticide resistance in the peach potato aphid,Myzus persicae. *Ecological Entomology,* 35**,** 131-146.

FERRARI, J., GODFRAY, H. C. J., FAULCONBRIDGE, A. S., PRIOR, K. & VIA, S. 2006. Population Differentiation and Genetic Variation in Host Choice among Pea Aphids from Eight Host Plant Genera. *Evolution,* 60**,** 1574-1584.

FERRARI, J., VIA, S. & GODFRAY, H. C. 2008. Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. *Evolution,* 62**,** 2508-24.

FIELD, L. M., JAVED, N., STRIBLEY, M. F. & DEVONSHIRE, A. L. 1994. The peach-potato aphid Myzus persicae and the tobacco aphid Myzus nicotianae have the same esterase-based mechanisms of insecticide resistance. *Insect Molecular Biology,* 3**,** 143-148.

FOLMER, O., BLACK, M., HOEH, W., LUTZ, R. & VRIJENHOEK, R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol,* 3**,** 294-299.

FOOTTIT, R. G., MAW, H. E., CD, V. D. & HEBERT, P. D. 2008. Species identification of aphids (Insecta: Hemiptera: Aphididae) through DNA barcodes. *Mol Ecol Resour,* 8**,** 1189-201.

FORISTER, M. L., NOVOTNY, V., PANORSKA, A. K., BAJE, L., BASSET, Y., BUTTERILL, P. T., CIZEK, L., COLEY, P. D., DEM, F., DINIZ, I. R., DROZD, P., FOX, M., GLASSMIRE, A. E., HAZEN, R., HRCEK, J., JAHNER, J. P., KAMAN, O., KOZUBOWSKI, T. J., KURSAR, T. A., LEWIS, O. T., LILL, J., MARQUIS, R. J., MILLER, S. E., MORAIS, H. C., MURAKAMI, M., NICKEL, H., PARDIKES, N. A., RICKLEFS, R. E., SINGER, M. S., SMILANICH, A. M., STIREMAN, J. O., VILLAMARIN-CORTEZ, S., VODKA, S., VOLF, M., WAGNER, D. L., WALLA, T., WEIBLEN, G. D. & DYER, L. A. 2015. The global distribution of diet breadth in insect herbivores. *Proc Natl Acad Sci U S A,* 112**,** 442-7.

FRADIN, E. F., ZHANG, Z., JUAREZ AYALA, J. C., CASTROVERDE, C. D., NAZAR, R. N., ROBB, J., LIU, C. M. & THOMMA, B. P. 2009. Genetic dissection of Verticillium wilt resistance mediated by tomato Ve1. *Plant Physiol,* 150**,** 320-32.

FREEMAN, B. C. & BEATTIE, G. A. 2008. An Overview of Plant Defenses against Pathogens and Herbivores. *The Plant Health Instructor*.

GABRIELS, S. H., VOSSEN, J. H., EKENGREN, S. K., VAN OOIJEN, G., ABD-EL-HALIEM, A. M., VAN DEN BERG, G. C., RAINEY, D. Y., MARTIN, G. B., TAKKEN, F. L., DE WIT, P. J. & JOOSTEN, M. H. 2007. An NB-LRR protein required for HR signalling mediated by both extra- and intracellular resistance proteins. *Plant J,* 50**,** 14-28.

GALLAGHER, S. 2001. Quantitation of nucleic acids with absorption spectroscopy. *Curr Protoc Protein Sci,* Appendix 4**,** Appendix 4K.

GARKE, C., YTOURNEL, F., BED'HOM, B., GUT, I., LATHROP, M., WEIGEND, S. & SIMIANER, H. 2012. Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Anim Genet,* 43**,** 419-28.

GAUFFRE, B. & D'ACIER, A. C. 2006. New polymorphic microsatellite loci, cross-species amplification and PCR multiplexing in the black aphid, Aphis fabae Scopoli. *Molecular Ecology Notes,* 6**,** 440-442.

GAUTAM, S., GADHAVE, K. R., BUCK, J. W., DUTTA, B., COOLONG, T., ADKINS, S. & SRINIVASAN, R. 2020. Virus-virus interactions in a plant host and in a hemipteran vector: Implications for vector fitness and virus epidemics. *Virus Res,* 286**,** 198069.

GEIJTENBEEK, T. B. & GRINGHUIS, S. I. 2009. Signalling through C-type lectin receptors: shaping immune responses. *Nat Rev Immunol,* 9**,** 465-79.

GEISLER, M., NADEAU, J. & SACK, F. D. 2000. Oriented asymmetric divisions that generate the stomatal spacing pattern in arabidopsis are disrupted by the too many mouths mutation. *Plant Cell,* 12**,** 2075-86.

GEMAYEL, R., CHO, J., BOEYNAEMS, S. & VERSTREPEN, K. J. 2012. Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. *Genes (Basel),* 3**,** 461-80.

GERGERICH, R. C. & DOLJA, V. V. 2006. Introduction to Plant Viruses, the Invisible Foe. *The Plant Health Instructor*.

GILARDONI, P. A., HETTENHAUSEN, C., BALDWIN, I. T. & BONAVENTURE, G. 2011. Nicotiana attenuata LECTIN RECEPTOR KINASE1 suppresses the insect-mediated inhibition of induced defense responses during Manduca sexta herbivory. *Plant Cell,* 23**,** 3512-32.

GILBERT, M. T., MOORE, W., MELCHIOR, L. & WOROBEY, M. 2007. DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS One,* 2**,** e272.

GILLBARD, E. 2020. 4-point plan to control virus disease in sugar beet. *Farmers weekly*.

GLAWISCHNIG, E. 2007. Camalexin. *Phytochemistry,* 68**,** 401-6.

GLOVER, K. A., HANSEN, M. M., LIEN, S., ALS, T. D., HOYHEIM, B. & SKAALA, O. 2010. A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. *BMC Genet,* 11**,** 2.

GOGGIN, F. L., WILLIAMSON, V. M. & ULLMAN, D. E. 2001. Variability in the Response ofMacrosiphum euphorbiaeandMyzus persicae(Hemiptera: Aphididae) to the Tomato Resistance GeneMi. *Environmental Entomology,* 30**,** 101-106.

GOULSON, D. & KLEIJN, D. 2013. REVIEW: An overview of the environmental risks posed by neonicotinoid insecticides. *Journal of Applied Ecology,* 50**,** 977-987.

GRABENER, S., OLDELAND, J., SHORTALL, C. R. & HARRINGTON, R. 2020. Changes in phenology and abundance of suction-trapped Diptera from a farmland site in the
UK
over four decades. *Ecological Entomology,* 45**,** 1215-1219.

GRBIC, M., VAN LEEUWEN, T., CLARK, R. M., ROMBAUTS, S., ROUZE, P., GRBIC, V., OSBORNE, E. J., DERMAUW, W., NGOC, P. C., ORTEGO, F., HERNANDEZ-CRESPO, P., DIAZ, I., MARTINEZ, M., NAVAJAS, M., SUCENA, E., MAGALHAES, S., NAGY, L., PACE, R. M., DJURANOVIC, S., SMAGGHE, G., IGA, M., CHRISTIAENS, O., VEENSTRA, J. A., EWER, J., VILLALOBOS, R. M., HUTTER, J. L., HUDSON, S. D., VELEZ, M., YI, S. V., ZENG, J., PIRES-DASILVA, A., ROCH, F., CAZAUX, M., NAVARRO, M., ZHUROV, V., ACEVEDO, G., BJELICA, A., FAWCETT, J. A., BONNET, E., MARTENS, C., BAELE, G., WISSLER, L., SANCHEZ-RODRIGUEZ, A., TIRRY, L., BLAIS, C., DEMEESTERE, K., HENZ, S. R., GREGORY, T. R., MATHIEU, J., VERDON, L., FARINELLI, L., SCHMUTZ, J., LINDQUIST, E., FEYEREISEN, R. & VAN DE PEER, Y. 2011. The genome of Tetranychus urticae reveals herbivorous pest adaptations. *Nature,* 479**,** 487-92.

GROVER, A. & SHARMA, P. C. 2011. Is spatial occurrence of microsatellites in the genome a determinant of their function and dynamics contributing to genome evolution? *Current Science,* 100**,** 859-869.

GRUBE, A., DONALDSON, D., KIELY, T. & WU, L. 2011. Pesticides Industry Sales and Usage; 2006 and 2007 Market Estimates. *US Environmental Protection Agency.* Washington (DC).

GUILLEMAUD, T., MIEUZET, L. & SIMON, J. C. 2003. Spatial and temporal genetic variability in French populations of the peach-potato aphid, Myzus persicae. *Heredity (Edinb),* 91**,** 143-52.

GUILLEY, H., RICHARDS, K. E. & JONARD, G. 1995. Nucleotide sequence of beet mild yellowing virus RNA. *Archives of Virology,* 140**,** 1109-1118.

GUILLOT, G. & FOLL, M. 2009. Correcting for ascertainment bias in the inference of population structure. *Bioinformatics,* 25**,** 552-4.

GUO, J., XU, C., WU, D., ZHAO, Y., QIU, Y., WANG, X., OUYANG, Y., CAI, B., LIU, X., JING, S., SHANGGUAN, X., WANG, H., MA, Y., HU, L., WU, Y., SHI, S., WANG, W., ZHU, L., XU, X., CHEN,

R., FENG, Y., DU, B. & HE, G. 2018. Bph6 encodes an exocyst-localized protein and confers broad resistance to planthoppers in rice. *Nat Genet,* 50**,** 297-306.

GUPTA, R. C., MILLER MUKHERJEE, I. R., MALIK, J. K., DOSS, R. B., DETTBARN, W.-D. & MILATOVIC, D. 2019. Chapter 26 - Insecticides. *In:* GUPTA, R. C. (ed.) *Biomarkers in Toxicology (Second Edition).* Academic Press.

GUREVICH, V. V. & GUREVICH, E. V. 2018. Arrestin-mediated signaling: Is there a controversy? *World J Biol Chem,* 9**,** 25-35.

GUY, E., BOULAIN, H., AIGU, Y., LE PENNEC, C., CHAWKI, K., MORLIERE, S., SCHADEL, K., KUNERT, G., SIMON, J. C. & SUGIO, A. 2016. Optimization of Agroinfiltration in Pisum sativum Provides a New Tool for Studying the Salivary Protein Functions in the Pea Aphid Complex. *Front Plant Sci,* 7**,** 1171.

HAASL, R. J. & PAYSEUR, B. A. 2011. Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity (Edinb),* 106**,** 158-71.

HALPIN, C. 2005. Gene stacking in transgenic plants--the challenge for 21st century plant biotechnology. *Plant Biotechnol J,* 3**,** 141-55.

HARMEL, N., LETOCART, E., CHERQUI, A., GIORDANENGO, P., MAZZUCCHELLI, G., GUILLONNEAU, F., DE PAUW, E., HAUBRUGE, E. & FRANCIS, F. 2008. Identification of aphid salivary proteins: a proteomic investigation of Myzus persicae. *Insect Mol Biol,* 17**,** 165-74.

HARRIS, S. R., FEIL, E. J., HOLDEN, M. T., QUAIL, M. A., NICKERSON, E. K., CHANTRATITA, N., GARDETE, S., TAVARES, A., DAY, N., LINDSAY, J. A., EDGEWORTH, J. D., DE LENCASTRE, H., PARKHILL, J., PEACOCK, S. J. & BENTLEY, S. D. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. *Science,* 327**,** 469-74.

HAUER, M., HANSEN, A. L., MANDERYCK, B., OLSSON, Å., RAAIJMAKERS, E., HANSE, B., STOCKFISCH, N. & MÄRLÄNDER, B. 2017. Neonicotinoids in sugar beet cultivation in Central and Northern Europe: Efficacy and environmental impact of neonicotinoid seed treatments and alternative measures. *Crop Protection,* 93**,** 132-142.

HAUSER, S., STEVENS, M., BEUVE, M. & LEMAIRE, O. 2002. Biological properties and molecular characterization

of beet chlorosis virus. *Arch Virol,* 147**,** 745–762.

HEBERT, P. D., RATNASINGHAM, S. & DEWAARD, J. R. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc Biol Sci,* 270 Suppl 1**,** S96-9.

HEIMBACH, F., RUSS, A., SCHIMMER, M. & BORN, K. 2016. Large-scale monitoring of effects of clothianidin dressed oilseed rape seeds on pollinating insects in Northern Germany: implementation of the monitoring project and its representativeness. *Ecotoxicology,* 25**,** 1630-1647.

HELDEN, M., TJALLINGII, W. F. & DIELEMAN, F. L. 1993. The resistance of lettuce (Lactuca sativa L.) to Nasonovia ribisnigri: bionomics of N. ribisnigri on near isogenic lettuce lines. *Entomologia Experimentalis et Applicata,* 66**,** 53-58.

HENRY, L. M., MAIDEN, M. C., FERRARI, J. & GODFRAY, H. C. 2015. Insect life history and the evolution of bacterial mutualism. *Ecol Lett,* 18**,** 516-25.

HERVE, C., SERRES, J., DABOS, P., CANUT, H., BARRE, A., ROUGE, P. & LESCURE, B. 1999. Characterization of the Arabidopsis lecRK-a genes: members of a superfamily encoding putative receptors with an extracellular domain homologous to legume lectins. *Plant Mol Biol,* 39**,** 671-82.

HILL, C. B., CHIRUMAMILLA, A. & HARTMAN, G. L. 2012. Resistance and virulence in the soybean-Aphis glycines interaction. *Euphytica,* 186**,** 635-646.

HILL, W. G. & ROBERTSON, A. 1968. Linkage disequilibrium in finite populations. *Theor Appl Genet,* 38**,** 226-31.

HIMLER, A. G., ADACHI-HAGIMORI, T., BERGEN, J. E., KOZUCH, A., KELLY, S. E., TABASHNIK, B. E., CHIEL, E., DUCKWORTH, V. E., DENNEHY, T. J., ZCHORI-FEIN, E. & HUNTER, M. S. 2011. Rapid Spread of a Bacterial Symbiont in an Invasive Whitefly Is Driven by Fitness Benefits and Female Bias. *Science,* 332**,** 254-256.

HOGENHOUT, S. A., AMMAR EL, D., WHITFIELD, A. E. & REDINBAUGH, M. G. 2008. Insect vector interactions with persistently transmitted viruses. *Annu Rev Phytopathol,* 46**,** 327-59.

HOGENHOUT, S. A. & BOS, J. I. 2011. Effector proteins that modulate plant--insect interactions. *Curr Opin Plant Biol,* 14**,** 422-8.

HOGENHOUT, S. A., VAN DER HOORN, R. A., TERAUCHI, R. & KAMOUN, S. 2009. Emerging concepts in effector biology of plant-associated organisms. *Mol Plant Microbe Interact,* 22**,** 115-22.

HOHN, T. 2007. Plant virus transmission from the insect point of view. *Proc Natl Acad Sci U S A,* 104**,** 17905-6.

HONG, F., HAN, H. L., PU, P., WEI, D., WANG, J. & LIU, Y. 2019. Effects of Five Host Plant Species on the Life History and Population Growth Parameters of Myzus persicae (Hemiptera: Aphididae). *J Insect Sci,* 19.

HOPMANS, E. S., NATSOULIS, G., BELL, J. M., GRIMES, S. M., SIEH, W. & JI, H. P. 2014. A programmable method for massively parallel targeted sequencing. *Nucleic Acids Res,* 42**,** e88.

HOSSAIN, R., MENZEL, W., LACHMANN, C. & VARRELMANN, M. 2020. New insights into virus yellows distribution in Europe and effects of beet yellows virus, beet mild yellowing virus, and beet chlorosis virus on sugar beet yield following field inoculation. *Plant Pathology*.

HULL, R. 2014a. Chapter 2 - Plant Viruses and Their Classification. *Plant Virology (Fifth Edition).* Boston: Academic Press.

HULL, R. 2014b. Chapter 14 - Ecology, Epidemiology, and Control of Plant Viruses. *Plant Virology (Fifth Edition).* Boston: Academic Press.

HULLÉ, M., CHAUBET, B., TURPEAU, E. & SIMON, J.-C. 2020. Encyclop'Aphid: a website on aphids and their natural enemies. *Entomologia Generalis,* 40**,** 97-101.

HUMPHREY, T. V., BONETTA, D. T. & GORING, D. R. 2007. Sentinels at the wall: cell wall receptors and sensors. *New Phytol,* 176**,** 7-21.

HUSON, D. H. & BRYANT, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol,* 23**,** 254-67.

ICTV. 2019. *International Committee on Taxonomy of Viruses* [Online]. Available: https://talk.ictvonline.org/ [Accessed 13/09 2020].

IMMS, A. D. 1964. *Outlines of Entomology,* Dunfermline, United Kingdom, Methuen.

INTERNATIONAL APHID GENOMICS, C. 2010. Genome sequence of the pea aphid Acyrthosiphon pisum. *PLoS Biol,* 8**,** e1000313.

IP, C. L. C., LOOSE, M., TYSON, J. R., DE CESARE, M., BROWN, B. L., JAIN, M., LEGGETT, R. M., ECCLES, D. A., ZALUNIN, V., URBAN, J. M., PIAZZA, P., BOWDEN, R. J., PATEN, B., MWAIGWISYA, S., BATTY, E. M., SIMPSON, J. T., SNUTCH, T. P., BIRNEY, E., BUCK, D., GOODWIN, S., JANSEN, H. J., O'GRADY, J., OLSEN, H. E., MIN, I. O. N. A. & REFERENCE, C. 2015. MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Res,* 4**,** 1075.

JACOB, F., VERNALDI, S. & MAEKAWA, T. 2013. Evolution and Conservation of Plant NLR Functions. *Front Immunol,* 4**,** 297.

JACTEL, H., VERHEGGEN, F., THIERY, D., ESCOBAR-GUTIERREZ, A. J., GACHET, E., DESNEUX, N. & NEONICOTINOIDS WORKING, G. 2019. Alternatives to neonicotinoids. *Environ Int,* 129**,** 423-429.

JAENIKE, J., UNCKLESS, R., COCKBURN, S. N., BOELIO, L. M. & PERLMAN, S. J. 2010. Adaptation via symbiosis: recent spread of a Drosophila defensive symbiont. *Science,* 329**,** 212-5.

JAOUANNET, M., RODRIGUEZ, P. A., THORPE, P., LENOIR, C. J., MACLEOD, R., ESCUDERO-MARTINEZ, C. & BOS, J. I. 2014. Plant immunity in plant-aphid interactions. *Front Plant Sci,* 5**,** 663.

JENA, K. K. & KIM, S.-M. 2010. Current Status of Brown Planthopper (BPH) Resistance and Genetics. *Rice,* 3**,** 161-171.

JESCHKE, P., NAUEN, R., SCHINDLER, M. & ELBERT, A. 2011. Overview of the status and global strategy for neonicotinoids. *J Agric Food Chem,* 59**,** 2897-908.

JI, H., KIM, S. R., KIM, Y. H., SUH, J. P., PARK, H. M., SREENIVASULU, N., MISRA, G., KIM, S. M., HECHANOVA, S. L., KIM, H., LEE, G. S., YOON, U. H., KIM, T. H., LIM, H., SUH, S. C., YANG, J., AN, G. & JENA, K. K. 2016. Map-based Cloning and Characterization of the BPH18 Gene from Wild Rice Conferring Resistance to Brown Planthopper (BPH) Insect Pest. *Sci Rep,* 6**,** 34376.

JI, R., YU, H., FU, Q., CHEN, H., YE, W., LI, S. & LOU, Y. 2013. Comparative transcriptome analysis of salivary glands of two populations of rice brown planthopper, Nilaparvata lugens, that differ in virulence. *PLoS One,* 8**,** e79612.

JONES, J. D. & DANGL, J. L. 2006. The plant immune system. *Nature,* 444**,** 323-9.

JONES, R. A. C. & NAIDU, R. A. 2019. Global Dimensions of Plant Virus Diseases: Current Status and Future Perspectives. *Annu Rev Virol,* 6**,** 387-409.

JOUET, A. 2016. *The adaptive evolution of the plant pathogen Albugo candida.* PhD thesis, University of East Anglia.

JOUSSELIN, E., CRUAUD, A., GENSON, G., CHEVENET, F., FOOTTIT, R. G. & COEUR D'ACIER, A. 2013. Is ecological speciation a major trend in aphids? Insights from a molecular phylogeny of the conifer-feeding genus Cinara. *Front Zool,* 10**,** 56.

KALOSHIAN, I., KINSEY, M. G., WILLIAMSON, V. M. & ULLMAN, D. E. 2000. Mi-Mediated Resistance Against the Potato AphidMacrosiphum euphorbiae(Hemiptera: Aphididae) Limits Sieve Element Ingestion. *Environmental Entomology,* 29**,** 690-695.

KALOSHIAN, I. & WALLING, L. L. 2016. Hemipteran and dipteran pests: Effectors and plant host immune regulators. *J Integr Plant Biol,* 58**,** 350-61.

KANE, N. S., HIRSCHBERG, B., QIAN, S., HUNT, D., THOMAS, B., BROCHU, R., LUDMERER, S. W., ZHENG, Y., SMITH, M., ARENA, J. P., COHEN, C. J., SCHMATZ, D., WARMKE, J. & CULLY, D. F. 2000. Drug-resistant Drosophila indicate glutamate-gated chloride channels are targets for the antiparasitics nodulisporic acid and ivermectin. *Proc Natl Acad Sci U S A,* 97**,** 13949-54.

KANTURSKI, M., LEE, Y., CHOI, J. & LEE, S. 2018. DNA barcoding and a precise morphological comparison revealed a cryptic species in the Nippolachnus piri complex (Hemiptera: Aphididae: Lachninae). *Sci Rep,* 8**,** 8998.

KARASEV, A. V. 2000. Genetic Diversity and Evolution of Closteroviruses. *Annu Rev Phytopathol,* 38**,** 293-324.

KARASEV, A. V., NIKOLAEVA, O. V., MUSHEGIAN, A. R., LEE, R. F. & DAWSON, W. O. 1996. Organization of the 3'-terminal half of beet yellow stunt virus genome and implications for the evolution of closteroviruses. *Virology,* 221**,** 199-207.

KATI, A. N., MANDRIOLI, M., SKOURAS, P. J., MALLOCH, G. L., VOUDOURIS, C. C., VENTURELLI, M., MANICARDI, G. C., TSITSIPIS, J. A., FENTON, B. & MARGARITOPOULOS, J. T. 2014. Recent changes in the distribution of carboxylesterase genes and associated chromosomal rearrangements in Greek populations of the tobacco aphidMyzus persicae nicotianae. *Biological Journal of the Linnean Society,* 113**,** 455-470.

KENNEDY, G. G. & KISHABA, A. N. 1977. Response of Alate Melon Aphids to Resistant and Susceptible Muskmelon Lines. *Journal of Economic Entomology,* 70**,** 407-410.

KETTLES, G. J., DRUREY, C., SCHOONBEEK, H. J., MAULE, A. J. & HOGENHOUT, S. A. 2013. Resistance of Arabidopsis thaliana to the green peach aphid, Myzus persicae, involves camalexin and is regulated by microRNAs. *New Phytol,* 198**,** 1178-90.

KETTLES, G. J. & KALOSHIAN, I. 2016. The Potato Aphid Salivary Effector Me47 Is a Glutathione-S-Transferase Involved in Modifying Plant Responses to Aphid Infestation. *Front Plant Sci,* 7**,** 1142.

KHUSH, G. S., KARIM, A. N. M. R. & ANGELES, E. R. 1985. Genetics of resistance of rice cultivar ARC10550 to Bangladesh brown pianthopper teletype. *Journal of Genetics,* 64**,** 121-125.

KIM, Y. & STEPHAN, W. 2002. Detecting a Local Signature of Genetic Hitchhiking Along a Recombining Chromosome. *Genetics,* 160**,** 765-777.

KINDLER, S. D., HARVEY, T. L., WILDE, G. E., SHUFRAN, R. A., BROOKS, H. L. & SLODERBECK, P. E. 2001. Occurrence of greenbug biotype K in the field. *Journal of Agricultural and Urban Entomology,* 18**,** 23-24.

KING, J. R. & PORTER, S. D. 2004. Recommendations on the use of alcohols for preservation of ant specimens (Hymenoptera, Formicidae). *Insectes Sociaux,* 51**,** 197-202.

KINGAN, S. B., HEATON, H., CUDINI, J., LAMBERT, C. C., BAYBAYAN, P., GALVIN, B. D., DURBIN, R., KORLACH, J. & LAWNICZAK, M. K. N. 2019. A High-Quality De novo Genome Assembly from a Single Mosquito Using PacBio Sequencing. *Genes (Basel),* 10.

KLAGES, U. 2016. Review of field and monitoring studies investigating the role of nitro-substituted neonicotinoid insecticides in the reported losses of honey bee colonies (Apis mellifera). *Focus on the health of honey bee and bumble bee colonies as well as wild bees:.* Monheim am Rhein, Germany.

KLINGLER, J., CREASY, R., GAO, L., NAIR, R. M., CALIX, A. S., JACOB, H. S., EDWARDS, O. R. & SINGH, K. B. 2005. Aphid resistance in Medicago truncatula involves antixenosis and phloem-specific, inducible antibiosis, and maps to a single locus flanked by NBS-LRR resistance gene analogs. *Plant Physiol,* 137**,** 1445-55.

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. *J Exp Bot,* 60**,** 4115-27.

KLOTH, K. J., BUSSCHER-LANGE, J., WIEGERS, G. L., KRUIJER, W., BUIJS, G., MEYER, R. C., ALBRECTSEN, B. R., BOUWMEESTER, H. J., DICKE, M. & JONGSMA, M. A. 2017. SIEVE ELEMENT-LINING CHAPERONE1 Restricts Aphid Feeding on Arabidopsis during Heat Stress. *Plant Cell,* 29**,** 2450-2464.

KOHN, M. H., YORK, E. C., KAMRADT, D. A., HAUGHT, G., SAUVAJOT, R. M. & WAYNE, R. K. 1999. Estimating population size by genotyping faeces. *Proc Biol Sci,* 266**,** 657-63.

KOLLIAS, S., POORTVLIET, M., SMOLINA, I. & HOARAU, G. 2015. Low cost sequencing of mitogenomes from museum samples using baits capture and Ion Torrent. *Conservation Genetics Resources,* 7**,** 345-348.

KOZŁOWSKA-MAKULSKA, A., BEUVE, M., SYLLER, J., SZYNDEL, M. S., LEMAIRE, O., BOUZOUBAA, S. & HERRBACH, E. 2009. Aphid transmissibility of different European beet polerovirus isolates. *European Journal of Plant Pathology,* 125**,** 337-341.

KROL, E., MENTZEL, T., CHINCHILLA, D., BOLLER, T., FELIX, G., KEMMERLING, B., POSTEL, S., ARENTS, M., JEWORUTZKI, E., AL-RASHEID, K. A., BECKER, D. & HEDRICH, R. 2010. Perception of the Arabidopsis danger signal peptide 1 involves the pattern recognition receptor AtPEPR1 and its close homologue AtPEPR2. *J Biol Chem,* 285**,** 13471-9.

KUSABA, M., DWYER, K., HENDERSHOT, J., VREBALOV, J., NASRALLAH, J. B. & NASRALLAH, M. E. 2001. Self-Incompatibility in the Genus Arabidopsis: Characterization of the S Locus in the Outcrossing A. lyrata and Its Autogamous Relative A. thaliana. *The Plant Cell,* 13**,** 627-643.

LAETSCH, D. R., KOUTSOVOULOS, G., BOOTH, T., STAJICH, J. & KUMAR, S. 2017. DRL/blobtools: BlobTools v1.0.1. *Zenodo*.

LAMBERT, P. & PASCAL, T. 2011. Mapping Rm2 gene conferring resistance to the green peach aphid (Myzus persicae Sulzer) in the peach cultivar "Rubira®". *Tree Genetics & Genomes,* 7**,** 1057-1068.

LANGE, W. 1999. Taxonomy and cultonomy of beet (Beta vulgaris L.). *Botanical Journal of the Linnean Society,* 130**,** 81-96.

LANGE, W. H. 1987. Insect Pests of Sugar Beet. *Annual Review of Entomology,* 32**,** 341-360.

LARSON, E. L., TINGHITELLA, R. M. & TAYLOR, S. A. 2019. Insect Hybridization and Climate Change. *Frontiers in Ecology and Evolution,* 7.

LAWSON, M. J. & ZHANG, L. 2006. Distinct patterns of SSR distribution in the Arabidopsis thaliana and rice genomes. *Genome Biol,* 7**,** R14.

LEE, H.-R., LEE, S., PARK, S., VAN KLEEFF, P. J. M., SCHUURINK, R. C. & RYU, C.-M. 2018. Transient Expression of Whitefly Effectors in Nicotiana benthamiana Leaves Activates Systemic Immunity Against the Leaf Pathogen Pseudomonas syringae and Soil-Borne Pathogen Ralstonia solanacearum. *Frontiers in Ecology and Evolution,* 6.

LEE, W., KIM, H., LIM, J., CHOI, H. R., KIM, Y., KIM, Y. S., JI, J. Y., FOOTTIT, R. G. & LEE, S. 2011. Barcoding aphids (Hemiptera: Aphididae) of the Korean Peninsula: updating the global data set. *Mol Ecol Resour,* 11**,** 32-7.

LEFEUVRE, P., MARTIN, D. P., ELENA, S. F., SHEPHERD, D. N., ROUMAGNAC, P. & VARSANI, A. 2019. Evolution and ecology of plant viruses. *Nat Rev Microbiol,* 17**,** 632-644.

LETUNIC, I. & BORK, P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res,* 47**,** W256-W259.

LEWONTIN, R. C. 1964. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics,* 49**,** 49-67.

LI, B. & KIMMEL, M. 2013. Factors influencing ascertainment bias of microsatellite allele sizes: impact on estimates of mutation rates. *Genetics,* 195**,** 563-72.

LI, C., HOFREITER, M., STRAUBE, N., CORRIGAN, S. & NAYLOR, G. J. 2013. Capturing protein-coding genes across highly divergent species. *Biotechniques,* 54**,** 321-6.

LI, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics,* 27**,** 2987-93.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, J., CAO, J., NIU, J., LIU, X. & ZHANG, Q. 2015a. Identification of the Population Structure of Myzus persicae (Hemiptera: Aphididae) on Peach Trees in China Using Microsatellites. *J Insect Sci,* 15.

LI, J., HUANG, H., ZHU, M., HUANG, S., ZHANG, W., DINESH-KUMAR, S. P. & TAO, X. 2019. A Plant Immune Receptor Adopts a Two-Step Recognition Mechanism to Enhance Viral Effector Perception. *Mol Plant,* 12**,** 248-262.

LI, T., CAO, C., YANG, T., ZHANG, L., HE, L., XI, Z., BIAN, G. & LIU, N. 2015b. A G-protein-coupled receptor regulation pathway in cytochrome P450-mediated permethrin-resistance in mosquitoes, Culex quinquefasciatus. *Sci Rep,* 5**,** 17772.

LI, T. & LIU, N. 2018. The Function of G-Protein-Coupled Receptor-Regulatory Cascade in Southern House Mosquitoes (Diptera: Culicidae). *J Med Entomol,* 55**,** 862-870.

LI, W., ZHU, Y.-C., LI, F. & HE, Y. 2020a. In silico discovery of genes encoding insecticide targets and detoxifying enzymes in Brevicoryne brassicae and Lipaphis erysimi. *Journal of Asia-Pacific Entomology,* 23**,** 159-166.

LI, Y., MO, Y., LI, Z., YANG, M., TANG, L., CHENG, L. & QIU, Y. 2020b. Characterization and application of a gall midge resistance gene (Gm6) from Oryza sativa 'Kangwenqingzhan'. *Theor Appl Genet,* 133**,** 579-591.

LIMBURG, D. D., MAUK, P. A. & GODFREY, L. D. 1997. Characteristics of Beet Yellows Closterovirus Transmission to Sugar Beets by Aphis fabae. *Phytopathology,* 87**,** 766-71.

LISCHER, H. E. & EXCOFFIER, L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics,* 28**,** 298-9.

LIU, P. F., AVRAMOVA, L. V. & PARK, C. 2009. Revisiting absorbance at 230nm as a protein unfolding probe. *Anal Biochem,* 389**,** 165-70.

LIU, S., XIE, J., CHENG, J., LI, B., CHEN, T., FU, Y., LI, G., WANG, M., JIN, H., WAN, H. & JIANG, D. 2016. Fungal DNA virus infects a mycophagous insect and utilizes it as a transmission vector. *Proc Natl Acad Sci U S A,* 113**,** 12803-12808.

LIU, X. M., SMITH, C. M., FRIEBE, B. R. & GILL, B. S. 2005. Molecular Mapping and Allelic Relationships of Russian Wheat Aphid–Resistance Genes. *Crop Science,* 45**,** 2273-2280.

LIU, Y., WU, H., CHEN, H., LIU, Y., HE, J., KANG, H., SUN, Z., PAN, G., WANG, Q., HU, J., ZHOU, F., ZHOU, K., ZHENG, X., REN, Y., CHEN, L., WANG, Y., ZHAO, Z., LIN, Q., WU, F., ZHANG, X., GUO, X., CHENG, X., JIANG, L., WU, C., WANG, H. & WAN, J. 2015. A gene cluster encoding lectin receptor kinases confers broad-spectrum and durable insect resistance in rice. *Nat Biotechnol,* 33**,** 301-5.

LLEWELLYN, K. S., LOXDALE, H. D., HARRINGTON, R., CLARK, S. J. & SUNNUCKS, P. 2004. Evidence for gene flow and local clonal selection in field populations of the grain aphid (Sitobion avenae) in Britain revealed using microsatellites. *Heredity (Edinb),* 93**,** 143-53.

LOUIS, J., MONDAL, H. A. & SHAH, J. 2012. Green peach aphid infestation induces Arabidopsis PHYTOALEXIN-DEFICIENT4 expression at site of insect feeding. *Plant Signal Behav,* 7**,** 1431-3.

LOUIS, J. & SHAH, J. 2015. Plant defence against aphids: the PAD4 signalling nexus. *J Exp Bot,* 66**,** 449-54.

LU, H., GIORDANO, F. & NING, Z. 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics,* 14**,** 265-279.

LUTERBACHER, M. C., ASHER, M. J. C., DEAMBROGIO, E., BIANCARDI, E., STEVENATO, P. & FRESE, L. 2005. Sources of resistance to diseases of sugar beet in related Beta germplasm: I. Foliar diseases. *Euphytica,* 139**,** 105-121.

MADDISON, D. 2013. *How to collect beetles for DNA studies* [Online]. Available: https://subulatepalpomere.com/2013/04/12/how-to-collect-beetles-for-dna-studies/ [Accessed January 2017].

MAIENFISCH, P., ANGST, M., BRANDL, F., FISCHER, W., HOFER, D., KAYSER, H., KOBEL, W., RINDLISBACHER, A., SENN, R., STEINEMANN, A. & WIDMER, H. 2001. Chemistry and biology of thiamethoxam: a second generation neonicotinoid. *Pest Manag Sci,* 57**,** 906-13.

MALLET, J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philos Trans R Soc Lond B Biol Sci,* 363**,** 2971-86.

MALLET, J., WYNNE, I. R. & THOMAS, C. D. 2011. Hybridisation and climate change: brown argus butterflies in Britain (Polyommatus subgenus Aricia). *Insect Conservation and Diversity,* 4**,** 192-199.

MANDRIOLI, M. & MANICARDI, G. C. 2012. Unlocking holocentric chromosomes: new perspectives from comparative and functional genomics? *Curr Genomics,* 13**,** 343-9.

MARGARITOPOULOS, J. T., KASPROWICZ, L., MALLOCH, G. L. & FENTON, B. 2009. Tracking the global dispersal of a cosmopolitan insect pest, the peach potato aphid. *BMC Ecol,* 9**,** 13.

MARGARITOPOULOS, J. T., KATI, A. N., VOUDOURIS, C. C., SKOURAS, P. J. & TSITSIPIS, J. A. 2021. Long-term studies on the evolution of resistance of Myzus persicae (Hemiptera: Aphididae) to insecticides in Greece. *Bull Entomol Res,* 111**,** 1-16.

MARGARITOPOULOS, J. T., MALARKY, G., TSITSIPIS, J. A. & BLACKMAN, R. L. 2007. Microsatellite DNA and behavioural studies provide evidence of host-mediated speciation in Myzus persicae (Hemiptera: Aphididae). *Biological Journal of the Linnean Society,* 91**,** 687-702.

MARTIN, F. N. & TOOLEY, P. W. 2017. Phylogenetic relationships amongPhytophthoraspecies inferred from sequence analysis of mitochondrially encoded cytochrome oxidase I and II genes. *Mycologia,* 95**,** 269-284.

MARZACHÌ, C., ALMA, A., D'AQUILIO, M., MINUTO, G. & BOCCARDO, G. 1999. Detection and identification of phytoplasmas infecting cultivated and wild plants in Liguria (Italian Riviera). *Journal of Plant Pathology,* 81**,** 127-136.

MATHERS, T. C., CHEN, Y., KAITHAKOTTIL, G., LEGEAI, F., MUGFORD, S. T., BAA-PUYOULET, P., BRETAUDEAU, A., CLAVIJO, B., COLELLA, S., COLLIN, O., DALMAY, T., DERRIEN, T., FENG, H., GABALDON, T., JORDAN, A., JULCA, I., KETTLES, G. J., KOWITWANICH, K., LAVENIER, D., LENZI, P., LOPEZ-GOMOLLON, S., LOSKA, D., MAPLESON, D., MAUMUS, F., MOXON, S., PRICE, D. R., SUGIO, A., VAN MUNSTER, M., UZEST, M., WAITE, D., JANDER, G., TAGU, D., WILSON, A. C., VAN OOSTERHOUT, C., SWARBRECK, D. & HOGENHOUT, S. A. 2017. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biol,* 18**,** 27.

MATHERS, T. C., MUGFORD, S. T., HOGENHOUT, S. A. & TRIPATHI, L. 2020a. Genome Sequence of the Banana Aphid, Pentalonia nigronervosa Coquerel (Hemiptera: Aphididae) and Its Symbionts. *G3 (Bethesda),* 10**,** 4315-4321.

MATHERS, T. C., WOUTERS, R. H. M., MUGFORD, S. T., SWARBRECK, D., VAN OOSTERHOUT, C. & HOGENHOUT, S. A. 2020b. Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Mol Biol Evol*.

MATSUMOTO, Y. & HATTORI, M. 2018. The green rice leafhopper, Nephotettix cincticeps (Hemiptera: Cicadellidae), salivary protein NcSP75 is a key effector for successful phloem ingestion. *PLoS One,* 13**,** e0202492.

MATTALIANO, M. D., MONTANA, E. S., PARISKY, K. M., LITTLETON, J. T. & GRIFFITH, L. C. 2007. The Drosophila ARC homolog regulates behavioral responses to starvation. *Mol Cell Neurosci,* 36**,** 211-21.

MAYJONADE, B., GOUZY, J., DONNADIEU, C., POUILLY, N., MARANDE, W., CALLOT, C., LANGLADE, N. & MUNOS, S. 2016. Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *Biotechniques,* 61**,** 203-205.

MCLEAN, G. D., GARRETT, R. G. & RUESINK, W. G. 1986. *Plant virus epidemics: monitoring, modelling and predicting outbreaks*, Academic Press.

MCMULLAN, M., GARDINER, A., BAILEY, K., KEMEN, E., WARD, B. J., CEVIK, V., ROBERT-SEILANIANTZ, A., SCHULTZ-LARSEN, T., BALMUTH, A., HOLUB, E., VAN OOSTERHOUT, C. & JONES, J. D. 2015. Evidence for suppression of immunity as a driver for genomic introgressions and host range expansion in races of Albugo candida, a generalist parasite. *Elife,* 4.

MCVEAN, G. 2007. The structure of linkage disequilibrium around a selective sweep. *Genetics,* 175**,** 1395-406.

MENDY, B., WANG'OMBE, M. W., RADAKOVIC, Z. S., HOLBEIN, J., ILYAS, M., CHOPRA, D., HOLTON, N., ZIPFEL, C., GRUNDLER, F. M. & SIDDIQUE, S. 2017. Arabidopsis leucine-rich repeat receptor-like kinase NILR1 is required for induction of innate immunity to parasitic nematodes. *PLoS Pathog,* 13**,** e1006284.

MIAN, M. A. R., HAMMOND, R. B. & ST. MARTIN, S. K. 2008. New Plant Introductions with Resistance to the Soybean Aphid. *Crop Science,* 48.

MINGEOT, D., HAUTIER, L. & JANSEN, J. P. 2021. Structuration of multilocus genotypes associated with insecticide resistance of the peach potato aphid, Myzus persicae (Sulzer), in potato fields in southern Belgium. *Pest Manag Sci,* 77**,** 482-491.

MONTI, V., LOXDALE, H. D., CESARI, M., FRATTINI, M., PANINI, M., MAZZONI, E. & MANICARDI, G. C. 2016. Ecological genetics of Italian peach-potato aphid (Myzus persicae) populations in relation to geography, dispersal and insecticide resistance as studied using microsatellite and resistance markers. *Agricultural and Forest Entomology,* 18**,** 376-389.

MOREAU, C. S., WRAY, B. D., CZEKANSKI-MOIR, J. E. & RUBIN, B. E. R. 2013. DNA preservation: a test of commonly used preservatives for insects. *Invertebrate Systematics,* 27**,** 81-86.

MUGFORD, S., WOUTERS, R., MATHERS, T. C. & HOGENHOUT, S. 2020. High quality DNA extraction from very small individual insects. *Protocols.io*.

MUGFORD, S. T., BARCLAY, E., DRUREY, C., FINDLAY, K. C. & HOGENHOUT, S. A. 2016. An Immuno-Suppressive Aphid Saliva Protein Is Delivered into the Cytosol of Plant Mesophyll Cells During Feeding. *Mol Plant Microbe Interact,* 29**,** 854-861.

MUNDT, C. C. 2014. Durable resistance: a key to sustainable management of pathogens and pests. *Infect Genet Evol,* 27**,** 446-55.

MUTREJA, A., KIM, D. W., THOMSON, N. R., CONNOR, T. R., LEE, J. H., KARIUKI, S., CROUCHER, N. J., CHOI, S. Y., HARRIS, S. R., LEBENS, M., NIYOGI, S. K., KIM, E. J., RAMAMURTHY, T., CHUN, J., WOOD, J. L., CLEMENS, J. D., CZERKINSKY, C., NAIR, G. B., HOLMGREN, J., PARKHILL, J. & DOUGAN, G. 2011. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature,* 477**,** 462-5.

MUTTI, N. S., LOUIS, J., PAPPAN, L. K., PAPPAN, K., BEGUM, K., CHEN, M. S., PARK, Y., DITTMER, N., MARSHALL, J., REESE, J. C. & REECK, G. R. 2008. A protein from the salivary glands of the pea aphid, Acyrthosiphon pisum, is essential in feeding on a host plant. *Proc Natl Acad Sci U S A,* 105**,** 9965-9.

MUTTI, N. S., PARK, Y., REESE, J. C. & REECK, G. R. 2005. RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, Acyrthosiphon pisum. *J Insect Sci,* 6.

NAKABACHI, A. & ISHIKAWA, H. 1999. Provision of riboflavin to the host aphid, Acyrthosiphon pisum, by endosymbiotic bacteria, Buchnera. *Journal of Insect Physiology,* 45**,** 1-6.

NANDETY, R. S., CAPLAN, J. L., CAVANAUGH, K., PERROUD, B., WROBLEWSKI, T., MICHELMORE, R. W. & MEYERS, B. C. 2013. The role of TIR-NBS and TIR-X proteins in plant basal defense responses. *Plant Physiol,* 162**,** 1459-72.

NAUEN, R., JESCHKE, P. & COPPING, L. 2008. In Focus: neonicotinoid insecticides. *Pest Manag Sci,* 64**,** 1081.

NG, J. C. & FALK, B. W. 2006. Virus-vector interactions mediating nonpersistent and semipersistent transmission of plant viruses. *Annu Rev Phytopathol,* 44**,** 183-212.

NIELSEN, R. 2005. Molecular signatures of natural selection. *Annu Rev Genet,* 39**,** 197-218.

NIELSEN, R., WILLIAMSON, S., KIM, Y., HUBISZ, M. J., CLARK, A. G. & BUSTAMANTE, C. 2005. Genomic scans for selective sweeps using SNP data. *Genome Res,* 15**,** 1566-75.

NIKOLAIENKO, O., PATIL, S., ERIKSEN, M. S. & BRAMHAM, C. R. 2018. Arc protein: a flexible hub for synaptic plasticity and cognition. *Semin Cell Dev Biol,* 77**,** 33-42.

NIKOLAKAKIS, N. N., MARGARITOPOULOS, J. T. & TSITSIPIS, J. A. 2003. Performance of Myzus persicae (Hemiptera: Aphididae) clones on different host-plants and their host preference. *Bull Entomol Res,* 93**,** 235-42.

NISTELBERGER, H. M., SMITH, O., WALES, N., STAR, B. & BOESSENKOOL, S. 2016. The efficacy of high-throughput sequencing and target enrichment on charred archaeobotanical remains. *Sci Rep,* 6**,** 37347.

NOLEPPA, S. 2017. Banning neonicotinoids in the European Union: An ex-post assessment of economic and environmental costs. HFFA Research GmbH.

NOMBELA, G., WILLIAMSON, V. M. & MUNIZ, M. 2003. The root-knot nematode resistance gene Mi-1.2 of tomato is responsible for resistance against the whitefly Bemisia tabaci. *Mol Plant Microbe Interact,* 16**,** 645-9.

OERKE, E. C. 2005. Crop losses to pests. *The Journal of Agricultural Science,* 144**,** 31.

OLIVER, K. M., MORAN, N. A. & HUNTER, M. S. 2005. Variation in Resistance to Parasitism in Aphids Is Due to Symbionts Not Host Genotype. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 12795-12800.

OLIVER, K. M., RUSSELL, J. A., MORAN, N. A. & HUNTER, M. S. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences,* 100**,** 1803-1807.

OZEROV, M., VASEMAGI, A., WENNEVIK, V., DIAZ-FERNANDEZ, R., KENT, M., GILBEY, J., PRUSOV, S., NIEMELA, E. & VAHA, J. P. 2013. Finding markers that make a difference: DNA pooling and

SNP-arrays identify population informative markers for genetic stock identification. *PLoS One,* 8**,** e82434.
OZOE, Y. 2013. Chapter Four - γ-Aminobutyrate- and Glutamate-gated Chloride Channels as Targets of Insecticides. *In:* COHEN, E. (ed.) *Advances in Insect Physiology.* Academic Press.
PANELLA, L. & LEWELLEN, R. T. 2006. Broadening the genetic base of sugar beet: introgression from wild relatives. *Euphytica,* 154**,** 383-400.
PAVAN, S., JACOBSEN, E., VISSER, R. G. & BAI, Y. 2010. Loss of susceptibility as a novel breeding strategy for durable and broad-spectrum resistance. *Mol Breed,* 25**,** 1-12.
PAVLIDIS, P. & ALACHIOTIS, N. 2017. A survey of methods and tools to detect recent and strong positive selection. *J Biol Res (Thessalon),* 24**,** 7.
PAVLIDIS, P., ZIVKOVIC, D., STAMATAKIS, A. & ALACHIOTIS, N. 2013. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol,* 30**,** 2224-34.
PECCOUD, J., OLLIVIER, A., PLANTEGENEST, M. & SIMON, J. C. 2009. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc Natl Acad Sci U S A,* 106**,** 7495-500.
PELISSIE, B., CROSSLEY, M. S., COHEN, Z. P. & SCHOVILLE, S. D. 2018. Rapid evolution in insect pests: the importance of space and time in population genomics studies. *Curr Opin Insect Sci,* 26**,** 8-16.
PENG, J. H., BAI, Y., HALEY, S. D. & LAPITAN, N. L. 2009. Microsatellite-based molecular diversity of bread wheat germplasm and association mapping of wheat resistance to the Russian wheat aphid. *Genetica,* 135**,** 95-122.
PETERS, B., GAO, Z. & ZUMKIER, U. 2016. Large-scale monitoring of effects of clothianidin-dressed oilseed rape seeds on pollinating insects in Northern Germany: effects on red mason bees (Osmia bicornis). *Ecotoxicology,* 25**,** 1679-1690.
PETERSEN, T. N., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods,* 8**,** 785-6.
PETIT, R. J., DEGUILLOUX, M. F., CHAT, J., GRIVET, D., GARNIER-GERE, P. & VENDRAMIN, G. G. 2005. Standardizing for microsatellite length in comparisons of genetic diversity. *Mol Ecol,* 14**,** 885-90.
PIASECKA, A., JEDRZEJCZAK-REY, N. & BEDNAREK, P. 2015. Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. *New Phytol,* 206**,** 948-64.
PITINO, M., COLEMAN, A. D., MAFFEI, M. E., RIDOUT, C. J. & HOGENHOUT, S. A. 2011. Silencing of aphid genes by dsRNA feeding from plants. *PLoS One,* 6**,** e25709.
PITINO, M. & HOGENHOUT, S. A. 2013. Aphid protein effectors promote aphid colonization in a plant species-specific manner. *Mol Plant Microbe Interact,* 26**,** 130-9.
POPKIN, M., PIFFARETTI, J., CLAMENS, A.-L., QIAO, G.-X., CHEN, J., VITALIS, R., VANLERBERGHE-MASUTTI, F., GUPTA, R. K., LAMAARI, M., LANGELLA, O., COEUR D'ACIER, A. & JOUSSELIN, E. 2016. Large-scale phylogeographic study of the cosmopolitan aphid pestBrachycaudus helichrysireveals host plant associated lineages that evolved in allopatry. *Biological Journal of the Linnean Society*.
POPLIN, R., RUANO-RUBIO, V., DEPRISTO, M. A., FENNELL, T. J., CARNEIRO, M. O., VAN DER AUWERA, G. A., KLING, D. E., GAUTHIER, L. D., LEVY-MOONSHINE, A., ROAZEN, D., SHAKIR, K., THIBAULT, J., CHANDRAN, S., WHELAN, C., LEK, M., GABRIEL, S., DALY, M. J., NEALE, B., MACARTHUR, D. G. & BANKS, E. 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*.
POPP, J. & HANTOS, K. 2013. The impact of crop protection on agricultural production. *Studies in Agricultural Economics,* 113**,** 47-66.
PORCHER, M. H. 2005. *Sorting BETA names* [Online]. Available: http://www.plantnames.unimelb.edu.au/Sorting/Beta.html#vulgaris-vulgaris-crassa [Accessed 12th-February 2017].

PRINCE, D. C., DRUREY, C., ZIPFEL, C. & HOGENHOUT, S. A. 2014. The leucine-rich repeat receptor-like kinase BRASSINOSTEROID INSENSITIVE1-ASSOCIATED KINASE1 and the cytochrome P450 PHYTOALEXIN DEFICIENT3 contribute to innate immunity to aphids in Arabidopsis. *Plant Physiol,* 164**,** 2207-19.

PRITCHARD, J. K., STEPHENS, M. & DONNELLY, P. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics,* 155**,** 945-959.

PUINEAN, A. M., FOSTER, S. P., OLIPHANT, L., DENHOLM, I., FIELD, L. M., MILLAR, N. S., WILLIAMSON, M. S. & BASS, C. 2010. Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid Myzus persicae. *PLoS Genet,* 6**,** e1000999.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I., DALY, M. J. & SHAM, P. C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet,* 81**,** 559-75.

PUTMAN, A. I. & CARBONE, I. 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecol Evol,* 4**,** 4399-428.

QI, A., DEWAR, A. M. & HARRINGTON, R. 2004. Decision making in controlling virus yellows of sugar beet in the UK. *Pest Manag Sci,* 60**,** 727-32.

QIN, Y. J., BUAHOM, N., KROSCH, M. N., DU, Y., WU, Y., MALACRIDA, A. R., DENG, Y. L., LIU, J. Q., JIANG, X. L. & LI, Z. H. 2016. Genetic diversity and population structure in Bactrocera correcta (Diptera: Tephritidae) inferred from mtDNA cox1 and microsatellite markers. *Sci Rep,* 6**,** 38476.

RAMSEY, J. S., WILSON, A. C., DE VOS, M., SUN, Q., TAMBORINDEGUY, C., WINFIELD, A., MALLOCH, G., SMITH, D. M., FENTON, B., GRAY, S. M. & JANDER, G. 2007. Genomic resources for Myzus persicae: EST sequencing, SNP identification, and microarray design. *BMC Genomics,* 8**,** 423.

RANSON, H., JENSEN, B., VULULE, J. M., WANG, X., HEMINGWAY, J. & COLLINS, F. H. 2000. Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan Anopheles gambiae associated with resistance to DDT and pyrethroids. *Insect Mol Biol,* 9**,** 491-7.

RANSON, H., N'GUESSAN, R., LINES, J., MOIROUX, N., NKUNI, Z. & CORBEL, V. 2011. Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? *Trends Parasitol,* 27**,** 91-8.

RAO, W., ZHENG, X., LIU, B., GUO, Q., GUO, J., WU, Y., SHANGGUAN, X., WANG, H., WU, D., WANG, Z., HU, L., XU, C., JIANG, W., HUANG, J., SHI, S. & HE, G. 2019. Secretome Analysis and In Planta Expression of Salivary Proteins Identify Candidate Effectors from the Brown Planthopper Nilaparvata lugens. *Mol Plant Microbe Interact,* 32**,** 227-239.

RATH-MORRIS, E., CROWTHER, S. & GUESSOUM, M. 1998. Resistance-breaking biotypes of rosy apple aphid *Dysaphis plantaginea* on the resistant cultivar Florina. *Handout IOBC meeting Dundee*.

RATTAN, R. S. 2010. Mechanism of action of insecticidal secondary metabolites of plant origin. *Crop Protection,* 29**,** 913-920.

REN, J., GAO, F., WU, X., LU, X., ZENG, L., LV, J., SU, X., LUO, H. & REN, G. 2016. Bph32, a novel gene encoding an unknown SCR domain-containing protein, confers resistance against the brown planthopper in rice. *Sci Rep,* 6**,** 37645.

REYMOND, P. 2013. Perception, signaling and molecular basis of oviposition-mediated plant responses. *Planta,* 238**,** 247-58.

REYMOND, P. & CALANDRA, T. 2015. Plant Immune Responses: Aphids Strike Back. *Curr Biol,* 25**,** R604-6.

RHOADS, A. & AU, K. F. 2015. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics,* 13**,** 278-89.

ROBIDEAU, G. P., DE COCK, A. W., COFFEY, M. D., VOGLMAYR, H., BROUWER, H., BALA, K., CHITTY, D. W., DESAULNIERS, N., EGGERTSON, Q. A., GACHON, C. M., HU, C. H., KUPPER, F. C.,

RINTOUL, T. L., SARHAN, E., VERSTAPPEN, E. C., ZHANG, Y., BONANTS, P. J., RISTAINO, J. B. & LEVESQUE, C. A. 2011. DNA barcoding of oomycetes with cytochrome c oxidase subunit I and internal transcribed spacer. *Mol Ecol Resour,* 11**,** 1002-11.

ROCKMAN, M. V. & KRUGLYAK, L. 2009. Recombinational landscape and population genomics of Caenorhabditis elegans. *PLoS Genet,* 5**,** e1000419.

RODRIGUEZ, P. A., ESCUDERO-MARTINEZ, C. & BOS, J. I. 2017. An Aphid Effector Targets Trafficking Protein VPS52 in a Host-Specific Manner to Promote Virulence. *Plant Physiol,* 173**,** 1892-1903.

RODRIGUEZ, P. A., HOGENHOUT, S. A. & BOS, J. I. B. 2014a. Leaf-Disc Assay Based on Transient Over-Expression in Nicotiana benthamiana to Allow Functional Screening of Candidate Effectors from Aphids. *In:* BIRCH, P., JONES, J. T. & BOS, J. I. B. (eds.) *Plant-Pathogen Interactions: Methods and Protocols.* Totowa, NJ: Humana Press.

RODRIGUEZ, P. A., STAM, R., WARBROEK, T. & BOS, J. I. 2014b. Mp10 and Mp42 from the aphid species Myzus persicae trigger plant defenses in Nicotiana benthamiana through different activities. *Mol Plant Microbe Interact,* 27**,** 30-9.

ROLKE, D., FUCHS, S., GRUNEWALD, B., GAO, Z. & BLENAU, W. 2016a. Large-scale monitoring of effects of clothianidin-dressed oilseed rape seeds on pollinating insects in Northern Germany: effects on honey bees (Apis mellifera). *Ecotoxicology,* 25**,** 1648-1665.

ROLKE, D., PERSIGEHL, M., PETERS, B., STERK, G. & BLENAU, W. 2016b. Large-scale monitoring of effects of clothianidin-dressed oilseed rape seeds on pollinating insects in northern Germany: residues of clothianidin in pollen, nectar and honey. *Ecotoxicology,* 25**,** 1691-1701.

SANCHEZ, J. A., LA-SPINA, M., GUIRAO, P. & CANOVAS, F. 2013. Inferring the population structure of Myzus persicae in diverse agroecosystems using microsatellite markers. *Bull Entomol Res,* 103**,** 473-84.

SANCHEZ, R. & ZHOU, M. M. 2011. The PHD finger: a versatile epigenome reader. *Trends Biochem Sci,* 36**,** 364-72.

SANDHU, S. & KANG, M. S. 2017. Advances in Breeding for Resistance to Insects. *In:* ARORA, R. & SANDHU, S. (eds.) *Breeding Insect Resistant Crops for Sustainable Agriculture.* Singapore: Springer Singapore.

SASTRY, K. S. A. & ZITTER, T. 2013. *Plant Virus and Viroid Diseases in the Tropics*.

SAUGE, M.-H., LACROZE, J.-P., POESSEL, J.-L., PASCAL, T. & KERVELLA, J. 2002. Induced resistance by Myzus persicae in the peach cultivar 'Rubira'. *Entomologia Experimentalis et Applicata,* 102**,** 29-37.

SAUGE, M. H., LAMBERT, P. & PASCAL, T. 2012. Co-localisation of host plant resistance QTLs affecting the performance and feeding behaviour of the aphid Myzus persicae in the peach tree. *Heredity (Edinb),* 108**,** 292-301.

SCHALAMUN, M., KAINER, D., BEAVAN, E., NAGAR, R., ECCLES, D., RATHJEN, J., LANFEAR, R. & SCHWESSINGER, B. 2018. A comprehensive toolkit to enable MinION long-read sequencing in any laboratory.

SCHLIEPHAKE, E., GRAICHEN, K. & RABENSTEIN, F. 2000. Investigations on the vector transmission of the Beet mild yellowing virus (BMYV) and the Turnip yellows virus (TuYV). *Zeitschrift Fur Pflanzenkrankheiten Und Pflanzenschutz-Journal of Plant Diseases and Protection,* 107**,** 81-87.

SCHMUCK, R. & LEWIS, G. 2016. Review of field and monitoring studies investigating the role of nitro-substituted neonicotinoid insecticides in the reported losses of honey bee colonies (Apis mellifera). *Ecotoxicology,* 25**,** 1617-1629.

SCHOONHOVEN, L. M., VAN LOON, J. J. A. & DICKE, M. 2005. *Insect-Plant Biology*, Oxford University Press.

SCHUMAN, M. C. & BALDWIN, I. T. 2016. The Layers of Plant Responses to Insect Herbivores. *Annu Rev Entomol,* 61**,** 373-94.

SCHWESSINGER, B. 2017. High quality DNA from Fungi for long read sequencing e.g. PacBio, Nanopore MinION. *Protocols.io*.

SCHWESSINGER, B. & RATHJEN, J. P. 2017. Extraction of High Molecular Weight DNA from Fungal Rust Spores for Long Read Sequencing. *In:* PERIYANNAN, S. (ed.) *Wheat Rust Diseases: Methods and Protocols.* New York, NY: Springer New York.

SEGONZAC, C. & ZIPFEL, C. 2011. Activation of plant pattern-recognition receptors by bacteria. *Curr Opin Microbiol,* 14**,** 54-61.

SHANGGUAN, X., ZHANG, J., LIU, B., ZHAO, Y., WANG, H., WANG, Z., GUO, J., RAO, W., JING, S., GUAN, W., MA, Y., WU, Y., HU, L., CHEN, R., DU, B., ZHU, L., YU, D. & HE, G. 2018. A Mucin-Like Protein of Planthopper Is Required for Feeding and Induces Immunity Response in Plants. *Plant Physiol,* 176**,** 552-565.

SHEN, W., LE, S., LI, Y. & HU, F. 2016. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One,* 11**,** e0163962.

SHEPPARD, L. W., BELL, J. R., HARRINGTON, R. & REUMAN, D. C. 2015. Changes in large-scale climate alter spatial synchrony of aphid pests. *Nature Climate Change,* 6**,** 610-613.

SHINGLETON, A. W., SISK, G. C. & STERN, D. L. 2003. Diapause in the pea aphid (Acyrthosiphon pisum) is a slowing but not a cessation of development. *BMC Dev Biol,* 3**,** 7.

SHIU, S. H. & BLEECKER, A. B. 2001. Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci U S A,* 98**,** 10763-8.

SHIU, S. H., KARLOWSKI, W. M., PAN, R., TZENG, Y. H., MAYER, K. F. & LI, W. H. 2004. Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. *Plant Cell,* 16**,** 1220-34.

SILVA, A. X., BACIGALUPE, L. D., LUNA-RUDLOFF, M. & FIGUEROA, C. C. 2012. Insecticide resistance mechanisms in the green peach aphid Myzus persicae (Hemiptera: Aphididae) II: Costs and benefits. *PLoS One,* 7**,** e36810.

SIMON-DELSO, N., AMARAL-ROGERS, V., BELZUNCES, L. P., BONMATIN, J. M., CHAGNON, M., DOWNS, C., FURLAN, L., GIBBONS, D. W., GIORIO, C., GIROLAMI, V., GOULSON, D., KREUTZWEISER, D. P., KRUPKE, C. H., LIESS, M., LONG, E., MCFIELD, M., MINEAU, P., MITCHELL, E. A., MORRISSEY, C. A., NOOME, D. A., PISA, L., SETTELE, J., STARK, J. D., TAPPARO, A., VAN DYCK, H., VAN PRAAGH, J., VAN DER SLUIJS, J. P., WHITEHORN, P. R. & WIEMERS, M. 2015. Systemic insecticides (neonicotinoids and fipronil): trends, uses, mode of action and metabolites. *Environ Sci Pollut Res Int,* 22**,** 5-34.

SIMON, J. C., BAUMANN, S., SUNNUCKS, P., HEBERT, P. D., PIERRE, J. S., LE GALLIC, J. F. & DEDRYVER, C. A. 1999. Reproductive mode and population genetic structure of the cereal aphid Sitobion avenae studied using phenotypic and microsatellite markers. *Mol Ecol,* 8**,** 531-45.

SINGH, K. S., TROCZKA, B. J., DUARTE, A., BALABANIDOU, V., TRISSI, N., CARABAJAL PALADINO, L. Z., NGUYEN, P., ZIMMER, C. T., PAPAPOSTOLOU, K. M., RANDALL, E., LUEKE, B., MAREC, F., MAZZONI, E., WILLIAMSON, M. S., HAYWARD, A., NAUEN, R., VONTAS, J. & BASS, C. 2020a. The genetic architecture of a host shift: An adaptive walk protected an aphid and its endosymbiont from plant chemical defenses. *Sci Adv,* 6**,** eaba1070.

SINGH, N. K., CHANCLUD, E. & CROLL, D. 2020b. Population-level deep sequencing reveals the interplay of clonal and sexual reproduction in the fungal wheat pathogen Zymoseptoria tritici. *bioRxiv***,** 2020.07.07.191510.

SINGH, P., KUO, Y. C., MISHRA, S., TSAI, C. H., CHIEN, C. C., CHEN, C. W., DESCLOS-THEVENIAU, M., CHU, P. W., SCHULZE, B., CHINCHILLA, D., BOLLER, T. & ZIMMERLI, L. 2012. The lectin receptor kinase-VI.2 is required for priming and positively regulates Arabidopsis pattern-triggered immunity. *Plant Cell,* 24**,** 1256-70.

SLATKIN, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics,* 139**,** 457-462.

SLOANE, M. A., SUNNUCKS, P., WILSON, A. C. & HALES, D. F. 2001. Microsatellite isolation, linkage group identification and determination of recombination frequency in the peach-potato aphid, Myzus persicae (Sulzer) (Hemiptera: Aphididae). *Genet Res,* 77**,** 251-60.

SMITH, H. G. & HALLSWORTH, P. B. 1990. The effects of yellowing viruses on yield of sugar beet in field trials, 1985 and 1987. *Annals of Applied Biology,* 116**,** 503-511.

SMITH, J. M. & HAIGH, J. 2009. The hitch-hiking effect of a favourable gene. *Genetical Research,* 23**,** 23-35.

SNOEREN, T. A., KAPPERS, I. F., BROEKGAARDEN, C., MUMM, R., DICKE, M. & BOUWMEESTER, H. J. 2010. Natural variation in herbivore-induced volatiles in Arabidopsis thaliana. *J Exp Bot,* 61**,** 3041-56.

SODERLUND, D. M. 2012. Molecular mechanisms of pyrethroid insecticide neurotoxicity: recent advances. *Arch Toxicol,* 86**,** 165-81.

SPENCE, J. M., BLACKMAN, R. L., TESTA, J. M. & READY, P. D. 1998. A 169-base pair tandem repeat DNA marker for subtelomeric heterochromatin and chromosomal rearrangements in aphids of the Myzus persicae group. *Chromosome Res,* 6**,** 167-75.

SPENCER, C. C., NEIGEL, J. E. & LEBERG, P. L. 2000. Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Molecular Ecology,* 9**,** 1517-1528.

STEINBRENNER, A. D., MUNOZ-AMATRIAIN, M., CHAPARRO, A. F., AGUILAR-VENEGAS, J. M., LO, S., OKUDA, S., GLAUSER, G., DONGIOVANNI, J., SHI, D., HALL, M., CRUBAUGH, D., HOLTON, N., ZIPFEL, C., ABAGYAN, R., TURLINGS, T. C. J., CLOSE, T. J., HUFFAKER, A. & SCHMELZ, E. A. 2020. A receptor-like protein mediates plant immune responses to herbivore-associated molecular patterns. *Proc Natl Acad Sci U S A,* 117**,** 31510-31518.

STERK, G., PETERS, B., GAO, Z. & ZUMKIER, U. 2016. Large-scale monitoring of effects of clothianidin-dressed OSR seeds on pollinating insects in Northern Germany: effects on large earth bumble bees (Bombus terrestris). *Ecotoxicology,* 25**,** 1666-1678.

STEVANATO, P., CHIODI, C., BROCCANELLO, C., CONCHERI, G., BIANCARDI, E., PAVLI, O. & SKARACIS, G. 2019. Sustainability of the Sugar Beet Crop. *Sugar Tech,* 21**,** 703-716.

STEVENS, M., HALLSWORTH, P. B. & SMITH, H. G. 2004. The effects of Beet mild yellowing virus and Beet chlorosis virus on the yield of UK field-grown sugar beet in 1997,1999 and 2000. *Annals of Applied Biology,* 144**,** 113-119.

SU, Q., PENG, Z., TONG, H., XIE, W., WANG, S., WU, Q., ZHANG, J., LI, C. & ZHANG, Y. 2019. A salivary ferritin in the whitefly suppresses plant defenses and facilitates host exploitation. *J Exp Bot,* 70**,** 3343-3355.

SUAREZ, A. V. & TSUTSUI, N. D. 2004. The Value of Museum Collections for Research and Society. *BioScience,* 54.

SUMMERS, E. G., NEWTON, A. S., KIRK, M. & TEMPLE, S. R. 1990. Transmission of Beet Yellows and Beet Mosaic Viruses by Noncolonizing Aphid Vectors. *Journal of Economic Entomology,* 83**,** 2448-2451.

SUN, M., VOORRIPS, R. E., STEENHUIS-BROERS, G., VAN'T WESTENDE, W. & VOSMAN, B. 2018. Reduced phloem uptake of Myzus persicae on an aphid resistant pepper accession. *BMC Plant Biol,* 18**,** 138.

SUN, M., VOORRIPS, R. E., VAN KAAUWEN, M., VISSER, R. G. F. & VOSMAN, B. 2020a. The ability to manipulate ROS metabolism in pepper may affect aphid virulence. *Hortic Res,* 7**,** 6.

SUN, M., VOORRIPS, R. E. & VOSMAN, B. 2020b. Aphid populations showing differential levels of virulence on Capsicum accessions. *Insect Sci,* 27**,** 336-348.

SUNNUCKS, P. & HALES, D. F. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus Sitobion (Hemiptera: Aphididae). *Mol Biol Evol,* 13**,** 510-24.

SUVENDHU, D. S., DIVYA, D., DURGA RANI, C. V., DAYAKAR REDDY, T., VISALAKSHMI, V., CHERALA, C., IBOHAL SINGH, K. & BENTUR, J. S. 2014. Characterization of gall midge resistant rice genotypes using resistance gene specific markers. *Journal of Experimental Biology and Agricultural Sciences,* 2**,** 439-446.

SYLVESTER, E. S. 1980. Circulative and Propagative Virus Transmission by Aphids. *Annual Review of Entomology,* 25**,** 257-286.

TABASHNIK, B. E., GASSMANN, A. J., CROWDER, D. W. & CARRIERE, Y. 2008. Insect resistance to Bt crops: evidence versus theory. *Nat Biotechnol,* 26**,** 199-202.

TAKKEN, F. L., ALBRECHT, M. & TAMELING, W. I. 2006. Resistance proteins: molecular switches of plant defence. *Curr Opin Plant Biol,* 9**,** 383-90.

TAKKEN, F. L. & GOVERSE, A. 2012. How to build a pathogen detector: structural basis of NB-LRR function. *Curr Opin Plant Biol,* 15**,** 375-84.

TAMELING, W. I., VOSSEN, J. H., ALBRECHT, M., LENGAUER, T., BERDEN, J. A., HARING, M. A., CORNELISSEN, B. J. & TAKKEN, F. L. 2006. Mutations in the NB-ARC domain of I-2 that impair ATP hydrolysis cause autoactivation. *Plant Physiol,* 140**,** 1233-45.

TAMURA, Y., HATTORI, M., YOSHIOKA, H., YOSHIOKA, M., TAKAHASHI, A., WU, J., SENTOKU, N. & YASUI, H. 2014. Map-based cloning and characterization of a brown planthopper resistance gene BPH26 from Oryza sativa L. ssp. indica cultivar ADR52. *Sci Rep,* 4**,** 5872.

TEAM, R. C. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

THABUIS, A. P. P., TEEKENS, C. K. & VAN HERWIJNEN, Z. O. 2011. *Lettuce that is resistant to the lettuce aphid Nasonovia ribisnigri biotype 1*. US and EU patent application.

THANGARAJ, S. R., MCCULLOCH, G. A., SUBBARAYALU, M., SUBRAMANIAM, C. & WALTER, G. H. 2016. Development of microsatellite markers and a preliminary assessment of population structuring in the rice weevil, Sitophilus oryzae (L.). *Journal of Stored Products Research,* 66**,** 12-17.

THOMPSON, H. M. 2012. Interaction between pesticides and other factors in effects on bees. Sand Hutton, York Food and Environment Research Agency.

THOMSEN, P. F., ELIAS, S., GILBERT, M. T., HAILE, J., MUNCH, K., KUZMINA, S., FROESE, D. G., SHER, A., HOLDAWAY, R. N. & WILLERSLEV, E. 2009. Non-destructive sampling of ancient insect DNA. *PLoS One,* 4**,** e5048.

TJALLINGII, W. F. 1978. Electronic Recording of Penetration Behavior by Aphids. *Entomologia Experimentalis Et Applicata,* 24**,** 721-730.

TJALLINGII, W. F. 1985. Membrane potentials as an indication for plant cell penetration by aphid stylets. *Entomologia Experimentalis et Applicata,* 38**,** 187-193.

TOBE, S. S., KITCHENER, A. C. & LINACRE, A. M. 2010. Reconstructing mammalian phylogenies: a detailed comparison of the cytochrome B and cytochrome oxidase subunit I mitochondrial genes. *PLoS One,* 5**,** e14156.

TOCKO-MARABENA, B. K., SILLA, S., SIMIAND, C., ZINGA, I., LEGG, J., REYNAUD, B. & DELATTE, H. 2017. Genetic diversity of Bemisia tabaci species colonizing cassava in Central African Republic characterized by analysis of cytochrome c oxidase subunit I. *PLoS One,* 12**,** e0182749.

TOR, M., LOTZE, M. T. & HOLTON, N. 2009. Receptor-mediated signalling in plants: molecular patterns and programmes. *J Exp Bot,* 60**,** 3645-54.

TSUCHIDA, T., KOGA, R. & FUKATSU, T. 2004. Host plant specialization governed by facultative symbiont. *Science,* 303**,** 1989.

TURCOTTE, M. M., REZNICK, D. N. & DANIEL HARE, J. 2013. Experimental test of an eco-evolutionary dynamic feedback loop between evolution and population density in the green peach aphid. *Am Nat,* 181 Suppl 1**,** S46-57.

VAID, N., MACOVEI, A. & TUTEJA, N. 2013. Knights in action: lectin receptor-like kinases in plant development and stress responses. *Mol Plant,* 6**,** 1405-18.

VALENCIA-MONTOYA, W. A., ELFEKIH, S., NORTH, H. L., MEIER, J. I., WARREN, I. A., TAY, W. T., GORDON, K. H. J., SPECHT, A., PAULA-MORAES, S. V., RANE, R., WALSH, T. K. & JIGGINS, C. D. 2020. Adaptive Introgression across Semipermeable Species Boundaries between Local Helicoverpa zea and Invasive Helicoverpa armigera Moths. *Mol Biol Evol,* 37**,** 2568-2583.

VALI, U., EINARSSON, A., WAITS, L. & ELLEGREN, H. 2008. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Mol Ecol,* 17**,** 3808-17.

VAN DAMME, E. J., NAKAMURA-TSURUTA, S., SMITH, D. F., ONGENAERT, M., WINTER, H. C., ROUGE, P., GOLDSTEIN, I. J., MO, H., KOMINAMI, J., CULERRIER, R., BARRE, A., HIRABAYASHI, J. & PEUMANS, W. J. 2007. Phylogenetic and specificity studies of two-domain GNA-related lectins: generation of multispecificity through domain duplication and divergent evolution. *Biochem J,* 404**,** 51-61.

VAN DER BIEZEN, E. A. & JONES, J. D. G. 1998. The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Current Biology,* 8**,** R226-R228.

VAN EMDEN, H. F. & HARRINGTON, R. 2017. aphids as crop pests. *CABI*.

VEIDT, I., LOT, H., LEISER, M., SCHEIDECKER, D., GUILLEY, H., RICHARDS, K. & JONARD, G. 1988. Nucleotide sequence of beet western yellows virus RNA. *Nucleic Acids Research,* 16**,** 9917-9932.

VERDUGO, J. A., SAUGE, M.-H., LACROZE, J.-P., FRANCIS, F. & RAMIREZ, C. C. 2015. Drought-stress and plant resistance affect herbivore performance and proteome: the case of the green peach aphidMyzus persicae(Hemiptera: Aphididae). *Physiological Entomology,* 40**,** 265-276.

VIA, S. 1999. Reproductive Isolation between Sympatric Races of Pea Aphids. I. Gene Flow Restriction and Habitat Choice. *Evolution,* 53**,** 1446-1457.

VIA, S. & HAWTHORNE, D. J. 2002. The genetic architecture of ecological specialization: correlated gene effects on host use and habitat choice in pea aphids. *Am Nat,* 159 Suppl 3**,** S76-88.

VIEIRA, M. L., SANTINI, L., DINIZ, A. L. & MUNHOZ CDE, F. 2016. Microsatellite markers: what they mean and why they are so useful. *Genet Mol Biol,* 39**,** 312-28.

VORBURGER, C. 2006. Temporal dynamics of genotypic diversity reveal strong clonal selection in the aphid Myzus persicae. *J Evol Biol,* 19**,** 97-107.

VORBURGER, C., LANCASTER, M. & SUNNUCKS, P. 2003. Environmentally related patterns of reproductive modes in the aphid Myzus persicae and the predominance of two 'superclones' in Victoria, Australia. *Molecular Ecology,* 12**,** 3493-3504.

VOUDOURIS, C. C., WILLIAMSON, M. S., SKOURAS, P. J., KATI, A. N., SAHINOGLOU, A. J. & MARGARITOPOULOS, J. T. 2017. Evolution of imidacloprid resistance in Myzus persicae in Greece and susceptibility data for spirotetramat. *Pest Manag Sci,* 73**,** 1804-1812.

WAITS, L., TABERLET, P., SWENSON, J. E., SANDEGREN, F. & FRANZEN, R. 2000. Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (Ursus arctos). *Molecular Ecology,* 9**,** 421-431.

WALLING, L. L. 2008. Avoiding effective defenses: strategies employed by phloem-feeding insects. *Plant Physiol,* 146**,** 859-66.

WANG, G., DONG, Y., LIU, X., YAO, G., YU, X. & YANG, M. 2018. The Current Status and Development of Insect-Resistant Genetically Engineered Poplar in China. *Front Plant Sci,* 9**,** 1408.

WANG, N., ZHAO, P., MA, Y., YAO, X., SUN, Y., HUANG, X., JIN, J., ZHANG, Y., ZHU, C., FANG, R. & YE, J. 2019. A whitefly effector Bsp9 targets host immunity regulator WRKY33 to promote performance. *Philos Trans R Soc Lond B Biol Sci,* 374**,** 20180313.

WANG, W., DAI, H., ZHANG, Y., CHANDRASEKAR, R., LUO, L., HIROMASA, Y., SHENG, C., PENG, G., CHEN, S., TOMICH, J. M., REESE, J., EDWARDS, O., KANG, L., REECK, G. & CUI, F. 2015a. Armet is an effector protein mediating aphid-plant interactions. *FASEB J,* 29**,** 2032-45.

WANG, X., WANG, R., YANG, Y., WU, S., O'REILLY, A. O. & WU, Y. 2016. A point mutation in the glutamate-gated chloride channel of Plutella xylostella is associated with resistance to abamectin. *Insect Mol Biol,* 25**,** 116-25.

WANG, Y., CAO, L., ZHANG, Y., CAO, C., LIU, F., HUANG, F., QIU, Y., LI, R. & LOU, X. 2015b. Map-based cloning and characterization of BPH29, a B3 domain-containing recessive gene conferring brown planthopper resistance in rice. *J Exp Bot,* 66**,** 6035-45.

WATERHOUSE, D. F. 1993. The major arthropod pests and weeds of agriculture in Southeast Asia. *ACIAR Canberra Australia*.

WATSON, M. A. 1940. Studies on the transmission of sugar-beet yellows virus by the aphis, Myzus persicae
(Sulz.). *Proceedings of the Royal Society of London. Series B - Biological Sciences,* 128**,** 535-552.

WEBER, G. 1985. Genetic-Variability in Host Plant Adaptation of the Green Peach Aphid, Myzus-Persicae. *Entomologia Experimentalis Et Applicata,* 38**,** 49-56.

WEIR, B. S. & COCKERHAM, C. C. 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution,* 38.

WENGER, J. A., CASSONE, B. J., LEGEAI, F., JOHNSTON, J. S., BANSAL, R., YATES, A. D., COATES, B. S., PAVINATO, V. A. & MICHEL, A. 2017. Whole genome sequence of the soybean aphid, Aphis glycines. *Insect Biochem Mol Biol*.

WIECZOREK, K. & CHŁOND, D. 2019. The first detection of the alien species: green-peach aphid Myzus (Nectarosiphon) persicae (Insecta, Hemiptera, Aphididae) in the Svalbard archipelago. *Polar Biology,* 42**,** 1947-1951.

WILL, T., TJALLINGII, W. F., THONNESSEN, A. & VAN BEL, A. J. 2007. Molecular sabotage of plant defense by aphid saliva. *Proc Natl Acad Sci U S A,* 104**,** 10536-41.

WILSON, A. C. C., MASSONNET, B., SIMON, J. C., PRUNIER-LETERME, N., DOLATTI, L., LLEWELLYN, K. S., FIGUEROA, C. C., RAMIREZ, C. C., BLACKMAN, R. L., ESTOUP, A. & SUNNUCKS, P. 2004. Cross-species amplification of microsatellite loci in aphids: assessment and application. *Molecular Ecology Notes,* 4**,** 104-109.

WILSON, A. C. C., SUNNUCKS, P. & HALES, D. F. 2003. Heritable genetic variation and potential for adaptive evolution in asexual aphids (Aphidoidea). *Biological Journal of the Linnean Society,* 79**,** 115-135.

WITEK, K., JUPE, F., WITEK, A. I., BAKER, D., CLARK, M. D. & JONES, J. D. 2016a. Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. *Nat Biotechnol,* 34**,** 656-60.

WITEK, K., JUPE, F., WITEK, A. I., BAKER, D., CLARK, M. D. & JONES, J. D. G. 2016b. SMRT RenSeq protocol. *Protocol Exchange*.

WOUTERS, R., MUGFORD, S., BIELLO, R., HEAVENS, D. & HOGENHOUT, S. 2020a. Extraction of high molecular weight DNA from aphids and other sap-feeding insects for long-read sequencing. *protocols.io*.

WOUTERS, R. H., MUGFORD, S. T. & HOGENHOUT, S. 2020b. Ambient sample storage system of field-collected insect samples for genomics. *Protocols.io*.

WU-SMART, J. & SPIVAK, M. 2016. Sub-lethal effects of dietary neonicotinoid insecticide exposure on honey bee queen fecundity and colony development. *Sci Rep,* 6**,** 32108.

WU, C. H., ABD-EL-HALIEM, A., BOZKURT, T. O., BELHAJ, K., TERAUCHI, R., VOSSEN, J. H. & KAMOUN, S. 2017. NLR network mediates immunity to diverse plant pathogens. *Proc Natl Acad Sci U S A,* 114**,** 8113-8118.

WU, C. H., BELHAJ, K., BOZKURT, T. O., BIRK, M. S. & KAMOUN, S. 2016. Helper NLR proteins NRC2a/b and NRC3 but not NRC1 are required for Pto-mediated cell death and resistance in Nicotiana benthamiana. *New Phytol,* 209**,** 1344-52.

WU, J. & BALDWIN, I. T. 2010. New insights into plant responses to the attack from insect herbivores. *Annu Rev Genet,* 44**,** 1-24.

WU, S., SHAN, L. & HE, P. 2014. Microbial signature-triggered plant defense responses and early signaling mechanisms. *Plant Sci,* 228**,** 118-26.

XU, H. X., QIAN, L. X., WANG, X. W., SHAO, R. X., HONG, Y., LIU, S. S. & WANG, X. W. 2019. A salivary effector enables whitefly to feed on host plants by eliciting salicylic acid-signaling pathway. *Proc Natl Acad Sci U S A,* 116**,** 490-495.

XU, S., JIANG, L., QIAO, G. & CHEN, J. 2020. Diversity of bacterial symbionts associated with Myzus persicae (Sulzer) (Hemiptera: Aphididae: Aphidinae) revealed by 16S rRNA Illumina sequencing. *Microb Ecol*.

YANG, C. H., GUO, J. Y., CHU, D., DING, T. B., WEI, K. K., CHENG, D. F. & WAN, F. H. 2017. Secretory laccase 1 in Bemisia tabaci MED is involved in whitefly-plant interaction. *Sci Rep,* 7**,** 3623.

YANG, M., CHENG, L., YAN, L., SHU, W., WANG, X. & QIU, Y. 2019. Mapping and characterization of a quantitative trait locus resistance to the brown planthopper in the rice variety IR64. *Hereditas,* 156**,** 22.

YE, W., YU, H., JIAN, Y., ZENG, J., JI, R., CHEN, H. & LOU, Y. 2017. A salivary EF-hand calcium-binding protein of the brown planthopper Nilaparvata lugens functions as an effector for defense responses in rice. *Sci Rep,* 7**,** 40498.

YOUNG, E., ROCK, G. C. & ZEIGER, D. C. 1982. Infestation of some Malus cultivars by the North Carolina woolly apple aphid biotype. *HortSience* 17**,** 787-788.

YU, X., WANG, G., HUANG, S., MA, Y. & XIA, L. 2014. Engineering plants for aphid resistance: current status and future perspectives. *Theor Appl Genet,* 127**,** 2065-83.

ZHANG, C. L., XU, D. C., JIANG, X. C., ZHOU, Y., CUI, J., ZHANG, C. X., CHEN, D. F., FOWLER, M. R., ELLIOTT, M. C., SCOTT, N. W., DEWAR, A. M. & SLATER, A. 2008. Genetic approaches to sustainable pest management in sugar beet (Beta vulgaris). *Annals of Applied Biology,* 152**,** 143-156.

ZHANG, L., WANG, F., QIAO, L., DIETRICH, C. H., MATSUMURA, M. & QIN, D. 2019. Population structure and genetic differentiation of tea green leafhopper, Empoasca (Matsumurasca) onukii, in China based on microsatellite markers. *Sci Rep,* 9**,** 1202.

ZHANG, W., ZHANG, X., LI, K., WANG, C., CAI, L., ZHUANG, W., XIANG, M. & LIU, X. 2018. Introgression and gene family contraction drive the evolution of lifestyle and host shifts of hypocrealean fungi. *Mycology,* 9**,** 176-188.

ZHAO, C., YANG, X. M., TANG, S. H., XU, P. J., BIAN, W. J., WANG, X. F., WANG, X. W. & REN, G. W. 2015. Population genetic structure of Myzus persicae nicotianae (Hemiptera: Aphididae) in China by microsatellite analysis. *Genet Mol Res,* 14**,** 17159-69.

ZHAO, Y., HUANG, J., WANG, Z., JING, S., WANG, Y., OUYANG, Y., CAI, B., XIN, X. F., LIU, X., ZHANG, C., PAN, Y., MA, R., LI, Q., JIANG, W., ZENG, Y., SHANGGUAN, X., WANG, H., DU, B., ZHU, L., XU, X., FENG, Y. Q., HE, S. Y., CHEN, R., ZHANG, Q. & HE, G. 2016. Allelic diversity in an NLR gene BPH9 enables rice to combat planthopper variation. *Proc Natl Acad Sci U S A,* 113**,** 12850-12855.

ZIEGLER, J. O., WÄLTHER, M., LINZER, T. R., SEGELBACHER, G., STAUSS, M., ROOS, C., LOESCHCKE, V. & TOMIUK, J. 2009. Frequent non-reciprocal exchange in microsatellite-containing-DNA-regions of vertebrates. *Journal of Zoological Systematics and Evolutionary Research,* 47**,** 15-20.

ZIPFEL, C. 2008. Pattern-recognition receptors in plant innate immunity. *Curr Opin Immunol,* 20**,** 10-6.

Appendices

Table S1 Number of samples collected per host, country and during what month.

	2017				Total 2017	2018						Total 2018	2019					Total 2019
	5	6	7	8		1	5	6	7	8	10		4	5	6	7	8	
Belgium		**31**		**45**	**76**									**1**	**2**			**3**
Apple		9			9													
bramble		1			1													
fodder beet		2			2													
oat		5			5													
Potato		1			1													
Roses				10	10													
Sugar beet		3		35	38									1	2			3
Thistle		2			2													
Wheat		2			2													
winter ramanas		6			6													
France							**19**	**37**				**56**	**9**	**55**	**71**	**21**		**156**
Barley														2				2
Hazel tree								4				4						
Pittosporum								3				3						
Roses								2				2		1 6	2 2	 1		39
Sugar beet							1 9	2 8				47	8	3 2	4 2	1 4		96
Wheat													1					1
Maize														5	7	6		18
Germany								**2**				**2**		**3**	**1**			**4**
Cabbage								2				2						
Sugar beet														3	1			4
Hungary								**3**	**10**			**13**						
peach									8			8						
Sugar beet								3	2			5						
Italy							**1**	**4**	**29**	**31**	**5**	**70**	**2**	**24**	**21**	**4**	**2**	**53**
acacia tree																	1	1
alfa-alfa										4		4					1	1
Apple									1	2		3						
beet							1	2	5			8						
cucumber									1			1			1			1
melon									1			1						
peach									6	13		19			2			2
pomegranate										2		2						
Potato									2			2		1	2			3

prunus																2			2
Roses						1	1			2			5	3					8
													1	1					
Sugar beet									5	5			5	3	1				29
tomato						1	1			2									
weed							6	2		8									
wild grass											2		3						5
wisteria							2	2		4					1				1
Maize							3	6		9									
						1	**2**	**1**					**2**						
Russia				**7**		**7**	**1**	**5**		**60**			**2**	**4**		**7**			**33**
Achillea							5			5									
Amaranthus														3		2			5
Ambrosia							5	5		10									
Carduus													6						6
Cichorium							6			6			6						6
Euphorbia																5			5
Helianthus annuus													4	1					5
Onopordum						5				5									
Roses				7						7									
Rumex						5				5									
Sonchus						7				7									
Sugar beet													6						6
Tanacetum							5	1 0		15									
						3	**1**						**5**	**2**					
Spain						**6**	**7**			**53**			**4**	**7**					**81**
Almond						9				9			5						5
													1						
artichoke													0						10
Beans						7	6			13									
canola						6	6			12									
Chinese cabbage						2				2									
Cotton							2			2									
													1						
peach													0						10
pomegranate													1 4						14
Potato						1				1									
						1							1	2					
Sugar beet						1	3			14			5	7					42
The	**3**	**5**				**3**	**2**						**1**	**5**					
Netherlands	**5**	**4**	**89**			**6**	**4**			**60**			**0**	**1**	**1**				**62**
canola														1					1
Chinese cabbage						1				1				1					1
Hazelnut						1				1									
milk thistle														8					8
						1													
Mustard						5				15									

Potato								1				1				6			6
								1	2										
Roses								1	4			35							
		3	5											1	2				
Sugar beet		5	4		89			7				7		0	8	1		39
tomato															7			7
United																**1**		
Kingdom	**14**				**14**	**3**		**6**	**9**	**2**		**20**		**1**	**8**	**2**		**21**
Cabbage						3						3						
Mustard								2	2			4						
Potato	2				2			4	6	1		11			4	9		13
Sugar beet	12				12				1	1		2		1	4	3		8
								1	**1**					**1**	**1**			
		6	**5**	**4**			**2**	**4**	**1**	**4**				**7**	**8**	**3**		
Total	**14**	**6**	**4**	**5**	**179**	**3**	**7**	**1**	**0**	**8**	**5**	**334**	**11**	**0**	**5**	**8**	**9**	**413**

Table S2 Species identification of samples collected. Specified per host, country and timepoint of collection.

	2017			2017 Total	2018					2018 Total	2019					2019 Total	Total
Months ->	**5**	**6**	**10**		**1**	**5**	**6**	**7**	**8**	**9**		**4**	**5**	**6**	**7**	**8**		
Australia			**7**	**7**														**7**
canola			**5**	**5**														**5**
Aphis fabae			1	1														1
Myzus persicae			2	2														2
Parasitoid wasp			2	2														2
Field peas			**1**	**1**														**1**
Aphis fabae			1	1														1
Sugar beet			**1**	**1**														**1**
Aulacorthum solani			1	1														1
Belgium		**15**		**15**		**9**	**1**				**10**				**1**		**1**	**26**
Apple		**3**		**3**		**1**					**1**							**4**
Aphis pomi		1		1														1
Myzus persicae		1		1		1					1							2
Parasitoid wasp		1		1														1
Chinese cabbage		**1**		**1**											**1**		**1**	**2**
Brevicoryne brassicae		1		1														1
Myzus persicae															1		1	1
fodder beet		**3**		**3**														**3**
Macrosiphum euphorbiae		1		1														1
Myzus persicae		1		1														1
Parasitoid wasp		1		1														1
potato		**3**		**3**		**1**	**1**				**2**							**5**
Macrosiphum euphorbiae		1		1		1					1							2
Myzus persicae		1		1														1
Parasitoid wasp		1		1			1				1							2
Sugar beet		**5**		**5**		**7**					**7**							**12**
Aphis fabae		1		1														1
Hyperomyzus lactucae		1		1														1
Macrosiphum euphorbiae		2		2		4					4							6
Myzus persicae		1		1		3					3							4
China	**14**	**2**		**36**														**36**
Cheesewood	**3**	**1**		**4**														**4**
Aphis spiraecola	2			2														2
Myzus persicae	1	1		2														2
Peach	**2**	**6**		**8**														**8**
Dysaphis plantaginea		2		2														2

Hyalopterus pruni	2	2	4									4
Myzus persicae		1	1									1
Parasitoid wasp		1	1									1
Rose		**2**	**2**									**2**
Myzus persicae		1	1									1
Parasitoid wasp		1	1									1
tea	**5**	**5**	**10**									**10**
Myzus persicae	2	2	4									4
Myzus varians	1	1	2									2
Parasitoid wasp	1	1	2									2
Sitobion avenae	1	1	2									2
unkown	**3**	**4**	**7**									**7**
Aphis fabae	1		1									1
Aphis sp.		1	1									1
Aphis spiraecola	1	1	2									2
Parasitoid wasp	1	1	2									2
Sitobion avenae		1	1									1
Wheat	**1**	**4**	**5**									**5**
Myzus persicae	1	3	4									4
Sitobion avenae		1	1									1
France				**12**	**5**	**17**	**8**	**18**	**16**	**7**	**49**	**66**
Maize										**1**	**1**	**1**
Parasitoid wasp										1	1	1
Pittosporum					**2**	**2**						**2**
Aphis fabae					2	2						2
Rose				**7**		**7**		**6**	**8**		**14**	**21**
Aphis fabae								1			1	1
Aulacorthum solani				1		1						1
Macrosiphum rosae				2		2			1		1	3
Macrosiphum sp.								2			2	2
Metopolophium dirhodum				2		2						2
Myzus persicae				2		2						2
Parasitoid wasp								1	3		4	4
Wahlgreniella nervata								2	4		6	6
Sugar beet				**5**	**3**	**8**	**7**	**12**	**8**	**6**	**33**	**41**
Aphis fabae				1	2	3	1	10	6	3	20	23
Aphis kurosawai				1		1						1
Aulacorthum solani				1	1	2						2
Macrosiphum rosae							3				3	3
Myzus persicae				1		1			2		2	3
Parasitoid wasp				1		1		1			1	2

Rhopalosiphum cerasifoliae								1				1	1
Wahlgreniella nervata							3			3		6	6
Wheat							**1**					**1**	**1**
Sitobion avenae							1					1	1
Germany		**2**				**2**							**2**
Sugar beet		**2**				**2**							**2**
Macrosiphum sp.		1				1							1
Parasitoid wasp		1				1							1
Hungary		**10**	**11**			**22**							**22**
Peach		**8**	**12**			**20**							**20**
Hyalopterus pruni		6	1			17							17
Myzus persicae		2				2							2
Parasitoid wasp			1			1							1
Sugar beet		**2**				**2**							**2**
Aphis fabae		2				2							2
Israel		**2**				**2**							**2**
Chinese cabbage		**2**				**2**							**2**
Myzus persicae		2				2							2
Italy		**2**	**10**	**11**	**2**	**25**	**1**	**18**	**11**	**1**	**1**	**32**	**57**
alfa-alfa				**1**		**1**					**1**	**1**	**2**
Myzus varians				1		1							1
Parasitoid wasp											1	1	1
cucumber									**1**			**1**	**1**
Myzus persicae									1			1	1
Peach			**7**	**7**		**14**							**14**
Aphis craccivora				1		1							1
Aphis fabae mordvilkoi			1			1							1
Aphis kurosawai				1		1							1
Aphis lhasaensis				1		1							1
Hyalopterus pruni			5	2		7							7
Parasitoid wasp			1			1							1
Uroleucon sp.				2		2							2
Rose			**1**			**1**		**4**	**2**			**6**	**7**
Aphis fabae								2	1			3	3
Hyperomyzus lactucae								1				1	1
Macrosiphum rosae								1				1	1
Parasitoid wasp			1			1			1			1	2
Sugar beet		**2**	**2**	**3**	**2**	**9**		**12**	**8**	**1**		**21**	**30**
Amphorophora rubi								1				1	1

Aphis fabae			1	1		1	3			4	7			11	14
Aphis fabae mordvilkoi			1	1	1		3								3
Aphis spiraecola												1		1	1
Aulacorthum solani					1		1								1
Hyperomyzus lactucae										2				2	2
Macrosiphum rosae										3				3	3
Myzus persicae						1	1								1
Parasitoid wasp					1		1			2	1			3	4
Wild grass								**1**		**2**				**3**	**3**
Aphis fabae								1		2				3	3
Kenya		**5**					**5**								**5**
kale		**5**					**5**								**5**
Aphis craccivora		1					1								1
Macrosiphum rosae		1					1								1
Myzus persicae		3					3								3
Netherlands			**1 7**	**1 3**			**30**			**1 6**	**2 3**	**3**		**42**	**72**
Chinese cabbage										**1**				**1**	**1**
Myzus persicae										1				1	1
Milk thistle										**2**				**2**	**2**
Macrosiphum euphorbiae										1				1	1
Parasitoid wasp										1				1	1
mustard			**7**				**7**								**7**
Aphis craccivora			1				1								1
Myzus persicae			5				5								5
Parasitoid wasp			1				1								1
potato			**4**				**4**								**4**
Parasitoid wasp			4				4								4
Rose				**1 0**			**10**								**10**
Macrosiphum rosae				3			3								3
Macrosiphum rosae				4			4								4
Myzus persicae				3			3								3
Sugar beet			**6**	**1**			**7**			**1 0**	**2 3**	**3**		**37**	**44**
Macrosiphum euphorbiae										1				1	1
Myzus persicae			6	1			7			1 0	2 3	3		36	43
Tomato										**2**				**2**	**2**
Macrosiphum euphorbiae										1				1	1
Nasonovia ribis-nigri										1				1	1
weed				**2**			**2**								**2**

Hyalopterus pruni				2		2					2
Russia		**3**	**10**	**10**	**1**	**24**		**22**	**4**	**8**	**34**	**58**
Achillea			**1**	**3**		**4**						**4**
Aphis kurosawai			1	1		2						2
Hyperomyzus lactucae				2		2						2
Amaranthus									**3**	**2**	**5**	**5**
Aphis craccivora									2	2	4	4
Aphis glycines									1		1	1
Ambrosia				**2**	**1**	**3**						**3**
Aphis fabae					1	1						1
Aphis kurosawai				2		2						2
Carduus								**6**			**6**	**6**
Aphis fabae								5			5	5
Parasitoid wasp								1			1	1
Cichorium				**1**		**1**		**6**			**6**	**7**
Aphis craccivora								1			1	1
Aphis glycines								3			3	3
Aphis intybi								1			1	1
Parasitoid wasp				1		1		1			1	2
Euphorbia										**6**	**6**	**6**
Aphis craccivora										2	2	2
Aphis fabae										1	1	1
Aphis glycines										1	1	1
Parasitoid wasp										2	2	2
Helianthus annuus								**3**	**1**		**4**	**4**
Aphis fabae								1	1		2	2
Parasitoid wasp								2			2	2
Onopordum			**1**			**1**						**1**
Lipaphis pseudobrassicae			1			1						1
Rose		**2**	**2**			**4**						**4**
Aphis fabae		2				2						2
Hyalopterus pruni			2			2						2
Rumex		**1**				**1**		**1**			**1**	**2**
Aphis kurosawai		1				1						1
Myzus persicae								1			1	1
Sonchus			**4**			**4**						**4**
Hyperomyzus lactucae			4			4						4
Sugar beet								**6**			**6**	**6**
Aphis fabae								5			5	5
Parasitoid wasp								1			1	1
Tanacetum				**4**		**4**						**4**
Metopeurum fuscoviride				4		4						4

thisle				**2**		**2**					**2**
Myzus persicae				2		2					2
South Africa	**2**	**2**									**2**
canola	**1**	**1**									**1**
Parasitoid wasp	1	1									1
Rose	**1**	**1**									**1**
Macrosiphum rosae	1	1									1
Spain			**3**	**2**	**9**	**34**	**3**	**4**	**9**	**43**	**77**
Almond			**1**	**3**	**1**	**5**	**1**			**1**	**6**
Brachycaudus prunicola							1			1	1
Hyalopterus amygdali				3		3					3
Hyalopterus amygdali			1		1	2					2
Almonds							**1**			**1**	**1**
Brachycaudus sp.							1			1	1
Artichoke							**10**			**10**	**10**
Aphis fabae							8			8	8
Brachycaudus sp.							1			1	1
Parasitoid wasp							1			1	1
Beans			**2**	**1**	**2**	**5**					**5**
Aphis fabae			1	1		2					2
Aphis fabae mordvilkoi			1			1					1
Aphis gossypii					1	1					1
Brachycaudus prunicola					1	1					1
canola					**2**	**2**					**2**
Lipaphis pseudobrassicae					1	1					1
Ropalosiphum padi					1	1					1
Chinese cabbage							**2**			**2**	**2**
Brachycaudus sp.							1			1	1
Myzus persicae							1			1	1
Peach							**5**			**5**	**5**
Hyalopterus pruni							3			3	3
Myzus persicae							2			2	2
Pomegranate tree							**7**			**7**	**7**
Aphis fabae							1			1	1
Aphis gossypii							4			4	4
Brachycaudus sp.							1			1	1
Parasitoid wasp							1			1	1
potato				**1**		**1**					**1**

Parasitoid wasp							1				1											1
Sugar beet							**1 7**	**4**			**21**		**8**	**9**			**17**	**38**
Aphis craccivora							1				1							1
Aphis fabae							4	1			5		7	6			13	18
Aphis fabae							4				4							4
Aphis fabae mordvilkoi							6	1			7							7
Brachycaudus prunicola													1				1	1
Brevicoryne brassicae								1			1							1
Myzus persicae														2			2	2
Parasitoid wasp							2	1			3			1			1	4
United Kingdom	**1**	**8**		**9**										**1 9**	**4**		**23**	**32**
Arabidopsis															**4**		**4**	**4**
Myzus persicae															4		4	4
Barley		**1**		**1**														**1**
Sitobion avenae		1		1														1
Peach		**7**		**7**														**7**
Brachycaudus helichrysi		2		2														2
Dysaphis plantaginea		1		1														1
Myzus persicae		3		3														3
Myzus varians		1		1														1
potato	**1**			**1**										**1 2**			**12**	**13**
Aphis fabae														1			1	1
Macrosiphum euphorbiae														7			7	7
Myzus persicae	1			1										1			1	2
Parasitoid wasp														3			3	3
Sugar beet														**7**			**7**	**7**
Macrosiphum euphorbiae														2			2	2
Myzus persicae														3			3	3
Parasitoid wasp														2			2	2
Grand Total	**15**	**4 7**	**7**	**69**	**5**	**2 7**	**7 1**	**5 4**	**1 2**	**2**	**171**	**9**	**9 2**	**7 5**	**3 6**	**1 2**	**224**	**46 4**

Table S3 Taxonomic annotation details of WGS using mash screen with an internal database of know WGS of aphid species. Information of taxonomic annotation information from 119 WGS samples. Samples are colour coded: Green indicates that samples are identified with strong identity (≥99% identity) to *Myzus persicae* and orange indicates low identity (<95%) to any aphid species in the local database.

Sample ID	Identity	shared hashes	total hashes	multiplicity	P-value	Closest species annotation
BE86	0.93	232	1000	7	0	*Acyrthosiphon pisum*
C11	0.95	347	1000	9	0	*Aphis fabae*
C18	0.95	353	1000	10	0	*Aphis fabae*
C19	0.95	353	1000	11	0	*Aphis fabae*
C5	0.95	348	1000	9	0	*Aphis fabae*
RUS14	1.00	913	1000	7	0	*Aphis fabae*
S116	1.00	967	1000	8	0	*Aphis fabae*
UK35	1.00	970	1000	8	0	*Aphis fabae*
BE2_B1	1.00	904	1000	7	0	*Aphis fabae*
C13	0.94	280	1000	11	0	*Aphis_gossypii*
IT102	1.00	905	1000	10	0	*Aphis_gossypii*
S135	1.00	967	1000	10	0	*Brachycaudus_helichrysi*
ES14	0.91	140	1000	8	0	*Brevicoryne brassicae*
HUN12	0.85	30	1000	3	2E-54	*Eriosoma lanigerum*
S258	0.93	228	1000	8	0	*Metopolophium dirhodum*
FR6	1.00	961	1000	7	0	*Metopolophium dirhodum*
NL53	0.93	218	1000	6	0	*Metopolophium dirhodum*
UK37	0.93	233	1000	6	0	*Metopolophium dirhodum*
UK32	0.93	227	1000	9	0	*Metopolophium dirhodum*
BE87	0.93	227	1000	5	0	*Metopolophium dirhodum*
FR44	0.93	233	1000	7	0	*Metopolophium dirhodum*
BE84	0.93	223	1000	7	0	*Metopolophium dirhodum*
BE85	0.93	219	1000	2	0	*Metopolophium dirhodum*
IT68	0.93	216	1000	8	0	*Metopolophium dirhodum*
A102	1.00	983	1000	9	0	*Myzus persicae*
A105	0.95	320	1000	1	0	*Myzus persicae*
A151	0.81	11	1000	1	2E-26	*Myzus persicae*
A166	1.00	977	1000	7	0	*Myzus persicae*
I1	1.00	971	1000	9	0	*Myzus persicae*
K16	1.00	973	1000	10	0	*Myzus persicae*
K40	1.00	971	1000	9	0	*Myzus persicae*
K43	1.00	971	1000	8	0	*Myzus persicae*
S196	1.00	975	1000	10	0	*Myzus persicae*
S204	1.00	979	1000	8	0	*Myzus persicae*
S232	1.00	975	1000	8	0	*Myzus persicae*
S472	1.00	973	1000	9	0	*Myzus persicae*
S481	1.00	979	1000	10	0	*Myzus persicae*
UK2	1.00	980	1000	8	0	*Myzus persicae*
A138	1.00	973	1000	8	0	*Myzus persicae*
C25	1.00	925	1000	3	0	*Myzus persicae*

Crespys	1.00	973	1000	7	0	*Myzus persicae*
ES01	1.00	980	1000	10	0	*Myzus persicae*
FR15	1.00	967	1000	9	0	*Myzus persicae*
FRC09	1.00	963	1000	8	0	*Myzus persicae*
FRC	1.00	970	1000	9	0	*Myzus persicae*
Generac	1.00	976	1000	9	0	*Myzus persicae*
Lierida	1.00	978	1000	9	0	*Myzus persicae*
MG1107	1.00	967	1000	8	0	*Myzus persicae*
NL1	1.00	980	1000	8	0	*Myzus persicae*
NL21	1.00	980	1000	7	0	*Myzus persicae*
NL5	1.00	984	1000	11	0	*Myzus persicae*
Q1200	1.00	970	1000	8	0	*Myzus persicae*
S121	1.00	962	1000	5	0	*Myzus persicae*
S152	1.00	971	1000	10	0	*Myzus persicae*
S2196G	1.00	970	1000	10	0	*Myzus persicae*
S2B01	1.00	971	1000	9	0	*Myzus persicae*
UKW3	1.00	983	1000	9	0	*Myzus persicae*
MISC27	1.00	980	1000	8	0	*Myzus persicae*
MISC28	1.00	982	1000	7	0	*Myzus persicae*
MISC30	1.00	974	1000	7	0	*Myzus persicae*
MISC31	1.00	976	1000	7	0	*Myzus persicae*
MISC81	1.00	976	1000	7	0	*Myzus persicae*
MISC33	1.00	985	1000	8	0	*Myzus persicae*
MISC34	1.00	970	1000	8	0	*Myzus persicae*
MISC38	1.00	984	1000	9	0	*Myzus persicae*
MISC4	1.00	975	1000	8	0	*Myzus persicae*
MISC42	1.00	980	1000	8	0	*Myzus persicae*
MISC46	1.00	978	1000	8	0	*Myzus persicae*
MISC47	0.99	789	1000	3	0	*Myzus persicae*
MISC48	1.00	974	1000	7	0	*Myzus persicae*
MISC51	1.00	982	1000	8	0	*Myzus persicae*
MISC53	1.00	979	1000	9	0	*Myzus persicae*
BE1	1.00	984	1000	9	0	*Myzus persicae*
BE2_b2	1.00	982	1000	10	0	*Myzus persicae*
BE23	1.00	988	1000	9	0	*Myzus persicae*
BE6	1.00	983	1000	10	0	*Myzus persicae*
K66	1.00	969	1000	8	0	*Myzus persicae*
MISC56	1.00	984	1000	8	0	*Myzus persicae*
UK20	1.00	955	1000	4	0	*Myzus persicae*
UK25	1.00	981	1000	8	0	*Myzus persicae*
UK21	1.00	981	1000	8	0	*Myzus persicae*
ES146	1.00	972	1000	5	0	*Myzus persicae*
ES149	1.00	959	1000	7	0	*Myzus persicae*
ES88	1.00	979	1000	8	0	*Myzus persicae*
ES92	1.00	982	1000	8	0	*Myzus persicae*
FR433	1.00	975	1000	11	0	*Myzus persicae*

K87	1.00	970	1000	8	0	*Myzus persicae*
K88	1.00	971	1000	8	0	*Myzus persicae*
A161	1.00	971	1000	10	0	*Myzus persicae*
NL93	1.00	967	1000	6	0	*Myzus persicae*
S17	1.00	972	1000	9	0	*Myzus persicae*
UK1	1.00	991	1000	9	0	*Myzus persicae*
UK19	1.00	970	1000	5	0	*Myzus persicae*
MISC60	1.00	985	1000	9	0	*Myzus persicae*
MISC64	1.00	981	1000	8	0	*Myzus persicae*
MISC65	1.00	978	1000	8	0	*Myzus persicae*
MISC67	1.00	988	1000	8	0	*Myzus persicae*
MISC69	1.00	987	1000	8	0	*Myzus persicae*
MISC76	1.00	978	1000	9	0	*Myzus persicae*
MISC78	1.00	987	1000	8	0	*Myzus persicae*
MISC79	1.00	986	1000	8	0	*Myzus persicae*
BE33A	1.00	981	1000	9	0	*Myzus persicae*
BE44	1.00	981	1000	9	0	*Myzus persicae*
BE49	1.00	979	1000	8	0	*Myzus persicae*
a456BE	1.00	979	1000	8	0	*Myzus persicae*
MISC5	1.00	982	1000	8	0	*Myzus persicae*
MISC10	1.00	978	1000	9	0	*Myzus persicae*
MISC14	1.00	988	1000	10	0	*Myzus persicae*
MISC15	1.00	981	1000	8	0	*Myzus persicae*
MISC18	1.00	973	1000	7	0	*Myzus persicae*
MISC19	1.00	988	1000	12	0	*Myzus persicae*
MISC20	1.00	981	1000	8	0	*Myzus persicae*
MISC21	1.00	985	1000	9	0	*Myzus persicae*
MISC23	1.00	984	1000	11	0	*Myzus persicae*
MISC26	1.00	983	1000	10	0	*Myzus persicae*
C16	1.00	994	1000	9	0	*Myzus varians*
S145	1.00	980	1000	11	0	*Phorodon humuli*
C4	1.00	959	1000	9	0	*Rhopalosiphum padi*
NL13	1.00	950	1000	6	0	*Sitobion miscanthi*

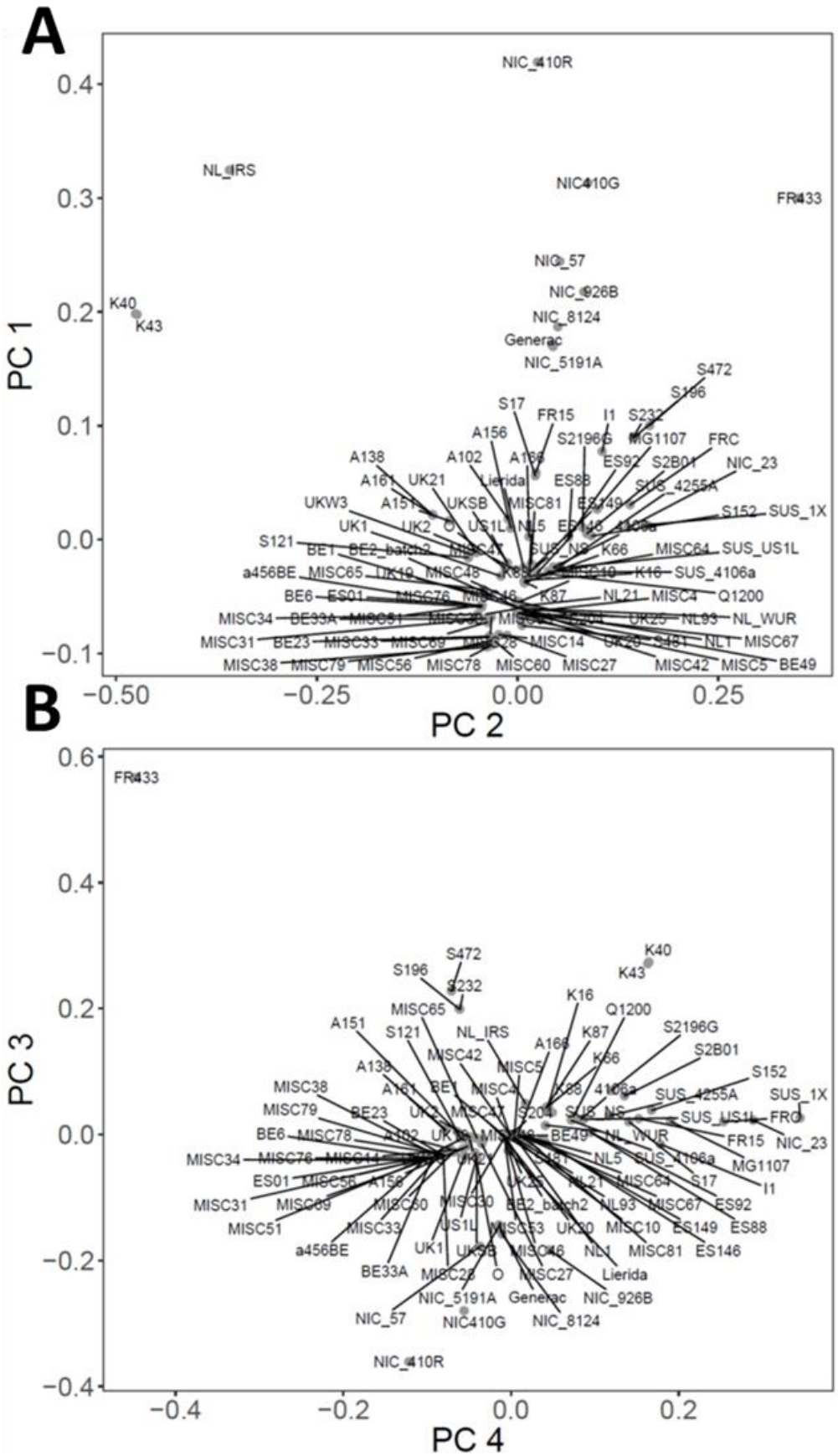


Figure S1. Principle component analysis with labels (figure 7.2.3.1). PCA showed around 15.5% of the variance between the genomic sample set of 2.8 million SNPs. **A.** PC1=4.2%, PC2=3.9%, **B.** PC3=3.8% and PC4=3.6% respectively.

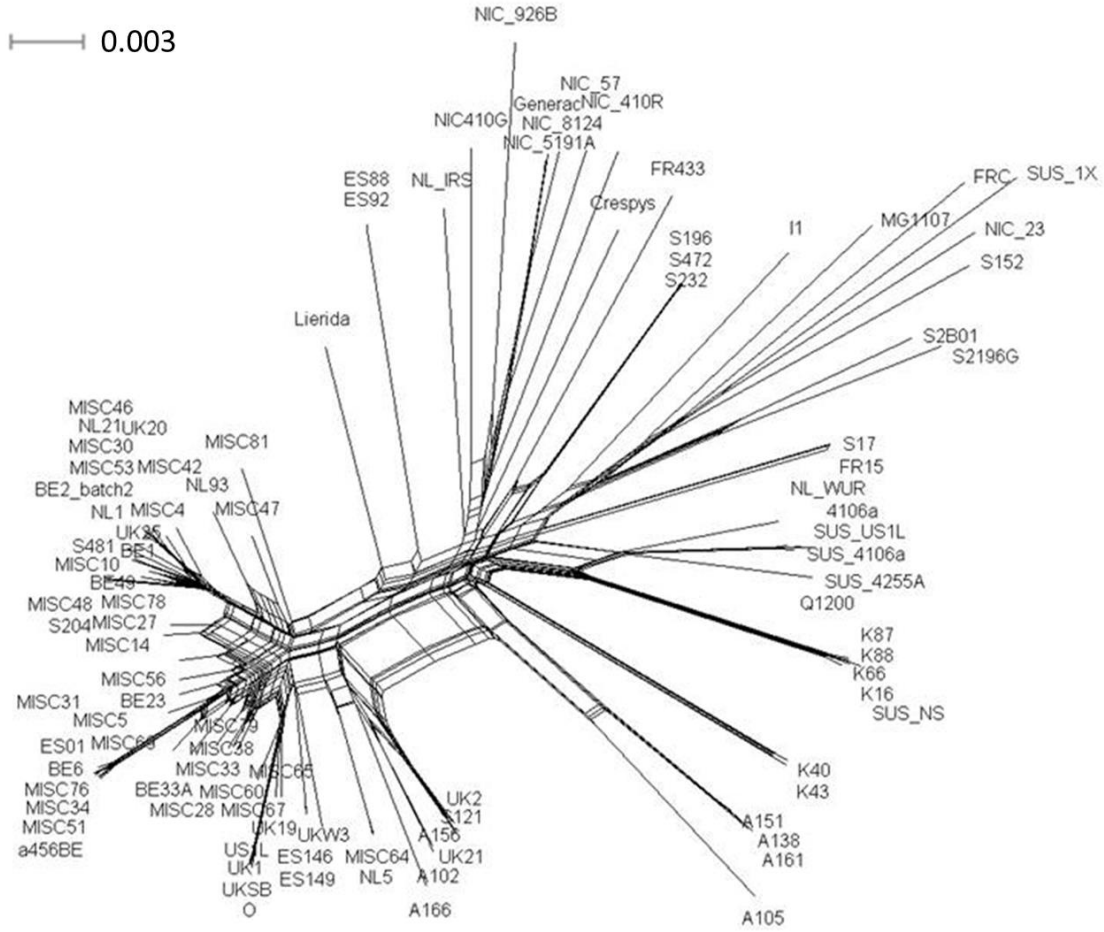


Figure S2 Network analysis with labels (figure 7.2.3.2). Neighbour-Net network analysis of 99 WGS samples. The scale represents distances estimated using the uncorrected p-distance.

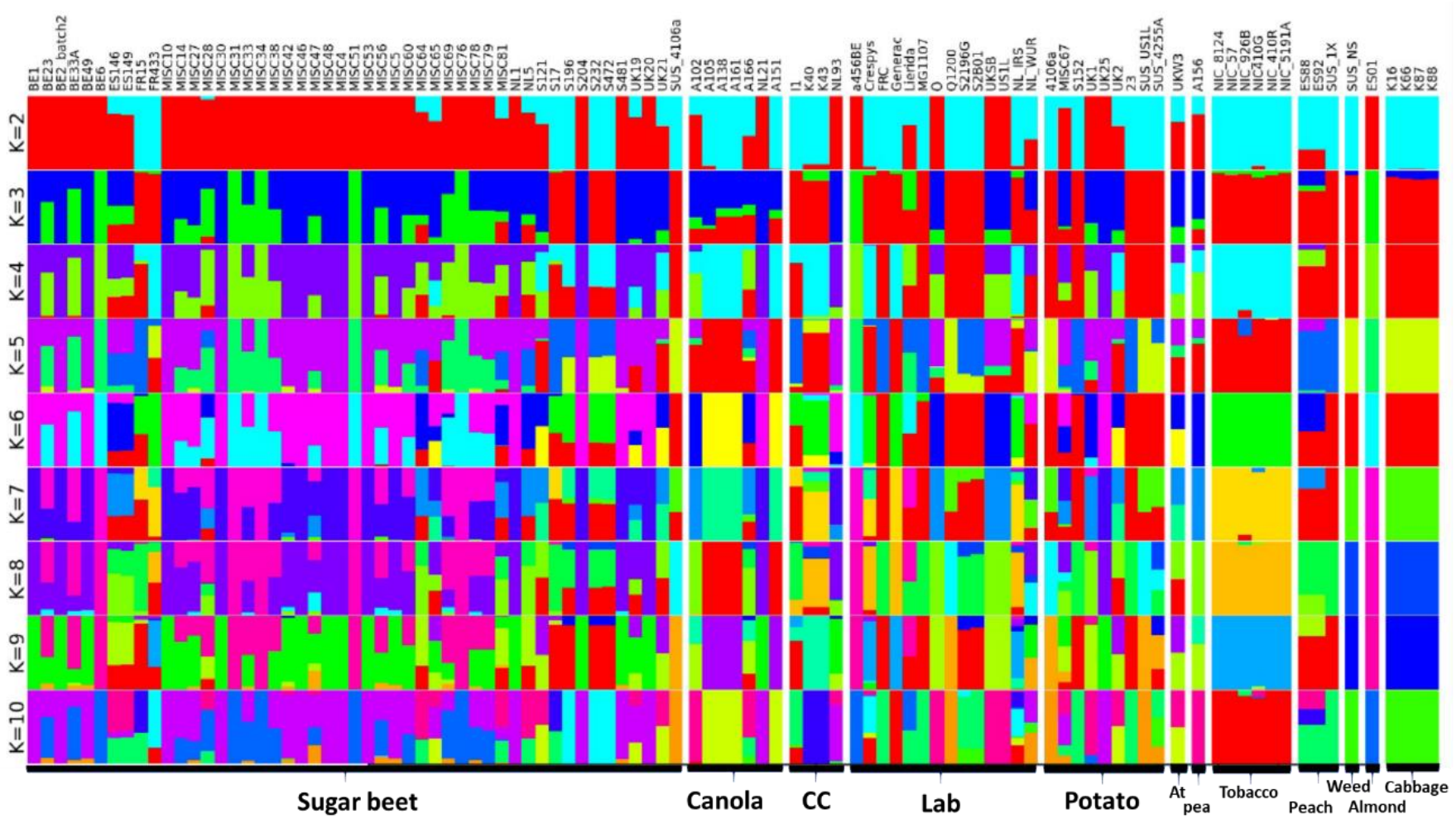


Figure S3 Admixture analysis with K=2 up to K=10 grouped per host. Genotypes were colour coded by predicted population genetic ancestor (K). Grouped per host (CC = Chinese cabbage, At = Arabidopsis thaliana).

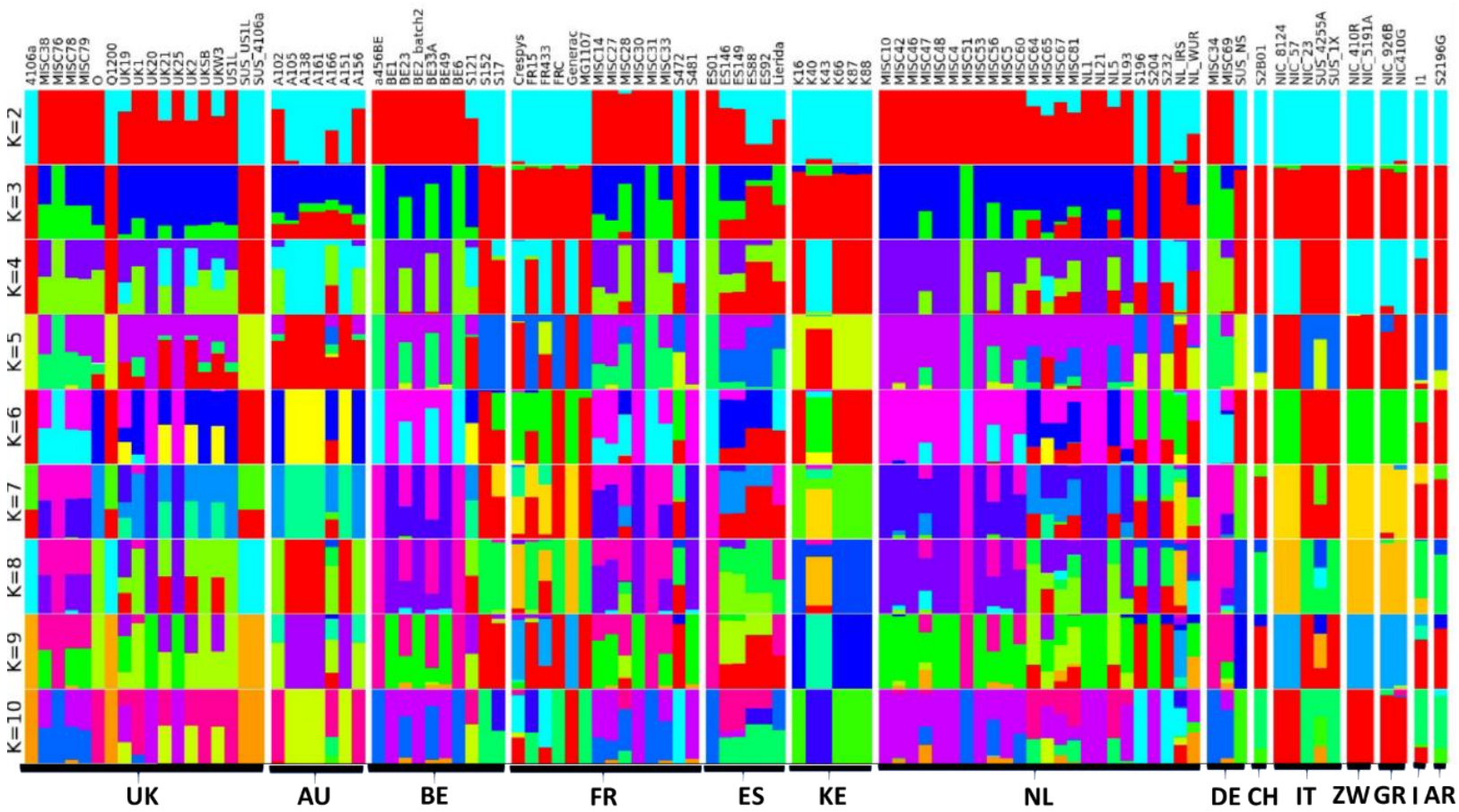


Figure S4 Admixture analysis with K=2 up to K=10 grouped per country. Genotypes were colour coded by predicted population genetic ancestor (K). Grouped per country (country ISO codes used for each country, except Israel that is represented as "I"). The additional admixture plots from K=2 up to K=7 with genotype label are found in the appendix (grouped per host: S3 and grouped per country: S4).