The chemical ecology of host plant associated speciation in the pea aphid (*Acyrthosiphon pisum)* (Homoptera: Aphididae)

By:
David Peter Hopkins

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Sciences

Departments of Animal and Plant Sciences

October 2015

Table of Contents

List of figures

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of tables

Abbreviations

C	Pathway EPG waveform
CNV	Copy number variation
Cr	Chemosensory receptor
E	Phloem EPG waveform
E1	Aphid saliva egestion into phloem EPG waveform
E2	Phloem sap ingestion EPG waveform
EAA	Essential Amino Acid
EPG	Electrical penetration graph
ESI	Electrospray ionisation
F	Stylet derailment EPG waveform
G	Xylem feeding EPG waveform
GLM	Generalized linear model
GLMM	Generalized linear mixed model
Gr	Gustatory receptors
HCAA	Hydroxycinnamoyl amides
JA	Jasmonic acid
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	Liquid chromatography
LDA	Linear discriminant analysis
LM	Linear model
LMM	Linear mixed model
m/z	Mass to charge ratio
MALDI	Matrix assisted laser desorption/ionization
MS aphids	Aphids adapted to *Medicago sativa*
MS	Mass spectroscopy
NPAA	Non-protein amino acid
Or	Odorant receptors
PCA	Principal component analysis
pd	Potential drop EPG waveform
Probe	Leave probe EPG waveform
QTL	Quantitative trait locus
RF	Random forest
rpd	Repetitive potential drop EPG waveform
ROS	Reactive oxygen species
SA	Salicylic acid
Tandem MS	Tandem mass spectrometry
TOF	Time of flight
TOF-MS	Time of flight-mass spectroscopy
TP aphids	Aphids adapted to *Trifolium pratense*
TyDC	L-tyrosine decarboxylase

Abstract

Previous work on pea aphid (*Acyrthosiphon pisum)* host-plant associated races has attributed their divergence to genes involved in chemosensory functions and metabolism of chemicals. In this study the host plant metabolic processes that drive *A. pisum* host plant race formation were investigated. First, profiles of aphid acceptance of plants were developed using the electrical penetration graph (EPG) technique. The acceptance of four *A. pisum* clones from two host races, associated with *Medicago sativa* or *Trifolium pratense,* was profiled across nine *Medicago* and ten *Trifolium* plant species. Acceptance profiles correlated strongly with aphid performance on plants. Aphid acceptance profiles were then compared with untargeted metabolomic profiles of plants, using random forest regression. Analysis revealed a small number of compounds that explained a large proportion of the variation in the *A. pisum* races differential acceptance of plant species. Two of these compounds were identified using tandem mass spectroscopy as L-phenylalanine and L-tyrosine, suggesting a possible link to the expression of a specific plant metabolic pathway. *M. sativa* and *T. pratense* plants were then pre-exposed to two divergent *A. pisum* clones. Aphid responses to pre-exposed and control plants were then profiled using EPG. The results suggested that *M. sativa* and *T. pratense* plants differ in their fixed (constitutive) and dynamic (induced or suppressed) responses to aphid attack. Exposing *M. sativa* plants to *A. pisum* clones appeared to also cause a change in the concentration of L-tyrosine, further suggesting a role of plant metabolic pathways in *A. pisum* divergent acceptance behaviour. The same two aphid clones were tested to see if they responded positively or negatively to diets containing varied concentrations of L-phenylalanine or L-tyrosine, but no conclusive evidence of aphid repulsion or attraction was found. This project identified that elements of plant chemical ecology could underlie divergent selection among *A. pisum* host races.

Acknowledgements

First and foremost I would like to thank my supervisors Professor Roger Butlin and Professor Duncan Cameron for the encouragement and guidance they have given me over the last few years. I particularly appreciate the time Roger spent helping me pick my way through the various layers of statistics involved in my project, and for being patient while he did so. It was also a great pleasure to work alongside Ludovic Duvaux and Isobel Eyres, the other members of the Sheffield aphid group, who never failed to give me the advice and enthusiasm I needed during my project.

I'd like to thank certain individuals who helped my project in crucial ways. Firstly I wish to thank Dr. Jean Christophe Simon, INRA, Rennes, and Dr. Julia Farrrari, University of York, and members of their respective groups for providing me with the aphid clones and the knowledge to set up and maintain my own aphid lab in Sheffield. I also wish to thank Dr. Athole Marshall, IBERS, Aberystwyth University, and his group for donating to Sheffield the seed of most of the plant species I used during this project. I am grateful to Dr. Jeremy Pritchard, Birmingham University, for taking the time to introduce me to the EPG technique and to Dr. Freddy Tjallingii for advising me when it came to setting up my own EPG system. I am also very grateful to Quentin Geissmann, Imperial College London, who introduced and taught me the basics of Random Forest modelling, which became a core component of my analysis.

None of this project would be possible without the help and support of some of the fantastic lab staff who work in the APS department. This includes Irene Jonson who patiently taught me many of the lab techniques needed during my project and put up with my regular invasion of her lab, Heather Walker who time and again showed me the ropes to the Mass spec machines and who's advice was invaluable to my project, Andrew Krupa who always found time to help me when I came calling to him and finally to Maggie Killion for running a tight ship at AWEC site during my PhD (despite at times my lab being the only one in use there) and helping me overcome the many logistical challenges I faced during my project.

I am also very grateful to the students, summer volunteers and other helpers who aided me at various points throughout my project. This is with particular emphasis to Anne-Lise Liabot who was an invaluable and hard working assistant during several periods of my PhD, and who many times helped me achieve what otherwise would have been impossible. I can't thank her enough for the unbelievable number of Petri dishes of seeds, trays of plants, pots of aphids, and Eppendorfs of leaf extracts I made her carry, label, clean, count, water, prepare etc., and doing so always with a smile.

Throughout my PhD I've had the pleasure to meet and work alongside a huge number of friendly, supportive and interesting people, most of whom I discovered

residing in the “intimately” sized box we called a coffee room. It is hard to imagine my PhD being nearly as rich and enjoyable as it was without people like Ben Jackson, Rachel Tucker, Nyle Khwaja, Katy Mayher, Jamie Thompson, Anja Westram and Jen Kaden (to name just a few). They not only gave me the daily lift to get me through but were also a fantastic collection of people to make friends with. In addition I’d like to give thanks to my comrade PhD students, such as Joe Gallagher, Jenny Armstrong, Helen Southern, Mauricio Montaño and Isabel Winney, whom I shared these years with (and in some cases longer) and who I could always relay on for some collective unburdening of our various PhD dilemmas.

I’m extremely grateful to Chris Gibson and Sylvia Kruniger for repeatedly letting me use their small home as my refuge in Sheffield, during the last few months of my PhD (and on top of that always making sure I was stuffed full with amazing food while I stayed there), but also for always being the best friends a person could wish for, in both the best and worse of times.

Большое спасибо очаровательная Венера Тюкмаева, who made my PhD the real adventure it was, and was my rock throughout it. I am also grateful to Venera during my write up for giving me a home to live in, and then with her kindness keeping me sane as I wrote.

I have a huge amount of gratitude to both my parents for sifting through a mass of spelling and grammatical errors and so help turn my thesis into a readable document. I very much appreciate the support they have never failed to give me. My final thanks goes to Henry (the spaniel), who was my obstinate but amusing friend during the final intensive thesis writing days.

Chapter 1: General overview

1.1 Host race ecological divergence and speciation

Since the formulation of the biological species concept it has been widely accepted that for one biological species to evolve into two, three conditions need to be met: i) a source of divergence, ii) a source of reproductive isolation and iii) a genetic basis for divergence and reproductive isolation (Coyne and Orr 2004, Nosil 2012). It is only when gene flow is restricted or entirely prevented can populations diverge by genetic drift and/or differing selection pressures to become true biological species (Rundle *et al.* 2005). Consequently, to understand the very earliest stages of speciation we need to understand both the causes of divergence and the initial barriers to gene flow (Coyne and Orr 2004).

While it is generally accepted that spatial barriers are capable of producing the genetic isolation necessary for speciation (i.e. allopatric speciation) (Hoskin *et al.* 2005), the proposal that species can evolve in the presence of partial gene flow because they co-occur (i.e. sympatric speciation) or have zones of contact (i.e. parapatric speciation) has been a contested issue (e.g. Via 2001, Bush and Butlin 2004, Barton 2010, Gavrilets 2014). This is particularly true for sympatric speciation, because of the difficulty in explaining how an initially panmictic population can develop restrictions to gene flow purely from biological features of the organism (Futuyma and Mayer 1980). We must also explain how populations within a species undergoing divergent selection can both co-exist and maintain their distinct characteristics in the face of recombination under gene flow (Coyne and Orr 2004). In a recent review of empirical and modelling studies Gavrilets (2014) concluded that conditions for sympatric speciation can be met if there is: i) a strong joint effect of both disruptive selection and non-random mating, ii) a high level of genetic variation, iii) close association of traits experiencing disruptive selection and those controlling non-random mating and iv) an absence or a low cost of being choosy. These conditions are relaxed if there is a partial spatial separation of the populations. However, it is important to note, as Fitzpatrick *et al.* (2008, 2009) have argued, that the more important issue is not the precise geography of speciation, but the processes involved, particularly how speciation with gene flow might occur (Smadja and Butlin 2011).

The requirements for speciation under gene flow are most likely to be associated with divergent selection upon resource use. It is therefore most likely to occur under conditions of ecological speciation, defined by Nosil (2012) “as the process by which barriers to gene flow evolve between populations as a result of ecologically based divergent selection between environments”. Under ecological speciation, when gene flow is present, the divergent selection pressure must be sufficiently strong that hybrids of diverging populations incur a severe fitness disadvantage in all parental environments. This then might lead to the reinforcement of gene flow barriers, where selection favours traits that allow individuals to avoid producing costly hybrid offspring (Rundle and Nosil 2005). Ecological speciation is much more likely to occur where traits under divergent selection also contribute to non-random mating (Gavrilets *et al.* 2007, Smadja and Butlin 2011, Nosil 2012).

Host races are often cited as examples where speciation with gene flow could be possible (Coyne and Orr 2004). The term “host race” has several definitions, but the most widely accepted is that of Drès and Mallet (2002): "genetically differentiated, sympatric populations of parasites that use different hosts and between which there is appreciable gene flow". Such races are of interest as they commonly represent a stage in the "speciation continuum" between polymorphic panmictic populations and true biological species (Bush 1969, Via 2001, Drès and Mallet 2002, Peccoud *et al.* 2009a, Nosil 2012).

Drès and Mallet (2002) set out five criteria that populations need to meet to be considered true host races:

i. The populations use different host taxa, with different populations exhibiting fidelity to particular hosts.

ii. Hosts exist in sympatry in at least part of their range.

iii. The populations are genetically differentiated at more than one locus and are spatially/temporally replicable (i.e. are more genetically differentiated from

populations on another host in sympatry than from at least some geographically distant populations on the same host).

iv. The populations display correlations between host choice and mate choice, but still undergo an appreciable level of gene flow (m≥1% per generation).

v. Races exhibit higher fitness on their native hosts than alternative hosts and produce hybrids that are less fit than parental forms.

The best researched examples of host race formation are found in host plant specialists, in particular phytophagous insects (Drés and Mallet 2002). Well studied examples include species from many insect orders, including the apple maggot fly (*Rhagoletis pomonella)* (e.g. Bush 1969; Feder *et al.* 2003), walking sticks (*Timema cristinae) (e.g.* Nosil *et al.* 2002, Comeault *et al.* 2014), the leaf beetle *Neochlamisus bebbianae (e.g.* Funk *et al.* 2002*)* and the pea aphid (*Acyrthosiphon pisum*) (Peccoud and Simon 2010), which is the subject of the current research.

As 25-40% of animal species may be phytophagous (Berlocher and Feder 2002) the process of host race formation may play a major role in the origin of global biodiversity. Indeed, the diversification of the phytophagous insects is strongly linked to the diversification of the plant species they live upon (Farrell 1998, Jaenike 1990, Stork 2007). Amongst phytophagous insects, traits such as host selection, host related performance and assortative mating are often tightly inter-connected, thus predisposing them to speciation under gene flow (Bush, 1975; Gripenberg *et al.* 2010; Powell *et al.* 2006). However, a challenge with host race formation is explaining how separate traits are both co-selected and maintain their genetic association during early stages of speciation, when gene flow is extensive and recombination tends to break down the genetic associations between co-adapted alleles and weaken trait associations (Smadja and Butlin 2011). Trait correlations can be maintained if relevant loci are tightly linked, favouring linkage disequilibrium, or if alleles have pleiotropic effects (Smadja and Butlin 2011). In addition, some traits may possess multiple effects, for example where insects reproduce on the same plant they feed on. In this case, traits for choosing hosts automatically cause non-random mating; defined by Gavrilets (2004) as the ‘habitat mechanism’.

It is important to note that host driven speciation occurs as a continuum of divergence from panmictic polymorphic populations to fully reproductively isolated biological species. Thus, organisms undergoing host race formation represent the intermediate stage in speciation, in which divergence and reproductive isolation between populations are still developing. They are therefore valuable study systems to investigate the selective forces and isolation mechanisms that initiate speciation; factors that are difficult to identify in full species.

As an example of ecological speciation, a key question concerning host plant specialisation is which features of different hosts select for divergent traits in insects. It is estimated that plants produce more than 100,000 different organic compounds (Walters 2011) and, although other characteristics such as morphology play a role, host plant chemistry is a particularly important source of divergent selection in phytophagous insects (Jaenike 1990). In addition to their primary metabolites, plants produce a huge array of secondary metabolites that vary greatly between plant species (Wink and Mohamed 2003, Wink 2013). This exposes specialist phytophagous insects to complex and dynamic chemical environments to which they must adapt. As primary producers, a multitude of organism use plants as a source of food and so the chemistry of host plant nutritional quality is a potential source of divergent selection (Awmack & Leather 2002). In addition, many plant secondary metabolites are thought to provide defences that insects must overcome (Walters 2011,Wink 2003, Wink and Mohamed 2003, Agrawal 2011, Wink 2013). Finally, plant chemicals may act as chemical fingerprints to allow insects to locate and identify suitable host plants (e.g. Picket *et al.* 1992, De Bruyne & Baker2008, Bruce and Picket 2011). Confirmation of the importance of plant chemistry in phytophagous insect adaptation comes from genomics. For example, in *A. pisum*, chemosensory, salivary and detoxification genes appear to play important functional roles in host adaptation and are undergoing rapid, host-related divergence (Jaquiéry *et al.* 2012, Duvaux *et al.* 2015, Simon *et al.* 2015). Understanding the way in which physiological and behavioural traits that vary between host races and are influenced

by plant chemistry is likely to provide further insight into early stages of speciation and is the main focus of this study.

1.2 *A. pisum as* a model system for studying ecological divergence of host-specialist races

In the exploration of the early processes underlying ecological speciation and host race formation, the *A. pisum* (Homoptera: Aphididae) species complex is increasingly becoming an important model (Drès and Mallet 2002, Peccoud and Simon 2010). Consequently it was chosen for this study into the role of chemical ecology in host race adaptation. *A. pisum* is a small sap sucking insect native to the Palaearctic region (Peccoud *et al.* 2009b) but introduced to North America and other temperate parts of the world (www.gbif.org/species /2077503). It is a specialist parasite of the Fabaceae (Pecoud and Simon 2010). Typical of most aphid species, *A. pisum* clones are polyphenic with the same genetic lines producing several morphs during the year. Most of the time, in the warm season, *A. pisum* females reproduce by apomictic viviparous parthenogenesis, rapidly producing genetically identical clones. However, in autumn sexual morphs living on the same host plants produce hardy eggs that enter overwinter diapause (for more details on the *A. pisum* life cycle see Figure 1.1). It is generally thought *A. pisum* races use the same host during winter diapause as summer (Hawthorn and Via 2001). However, there is little formal experimental or observational evidence on the winter host plants so far been published, which is an important gap in our understanding of the *A. pisum* model system. However, considering it is common for one both sexual forms to be wingless (Frantz et al. 2010) suggest *A. pisum* lays its egg on or near food host plants. For *Medicago* and *Trifolium* aphid races this particularly true, as there seem to be a higher proportion of winless males than other host races (Frantz et al. 2010). In addition the frequency of purely asexual aphid lineages on *Medicago* and *Trifolium* aphid races from western France also point to a particularly close fidelity to these host plants throughout the year (Frantz et al. 2006). Most summer asexual adults are wingless and live their entire lives on a single plant. At higher population densities (and sometimes in male morphs) clones produce winged (alate) aphids allowing dispersal to new host plants (Figure 1.1).

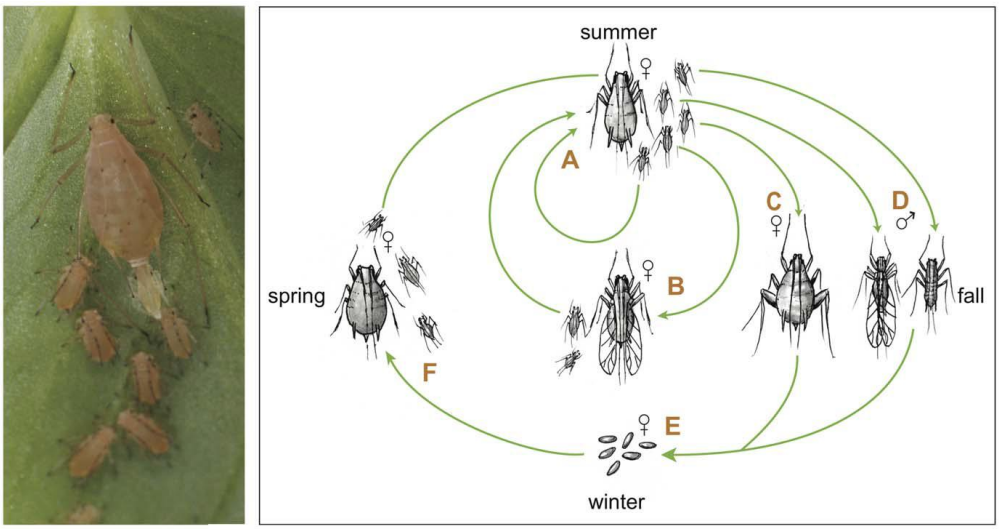


Figure 1.1: Life history of the pea aphid (Acyrthosiphon pisum). Females hatch in the spring to produce live wingless asexual females by parthenogenesis (F). Over summer these aphids (A) in turn produce multiple generations of wingless parthogenetic females that are able to reproduce at a rapid rate. During warmer parts of the year females continue to produce asexually but occasionally produce asexual winged (alate) parthogenetic females, particularly when under stress or at high population density (B). Alates allow aphid genetic lines to disperse to new plants when host plants become unsuitable. As day length shortens and temperatures drop parthenogenetic female aphids produce both the sexual female (C) and smaller male morphs (D). Sexual morphs then produce eggs (E) which stay in diapause over the winter months. In some clones male morphs are winged to aid dispersal. The eggs not only allow winter survival but provide recombinant genetic forms in an otherwise asexual reproductive system. In some clones, normally from places without cold winters, the sexual and egg-producing stages are missing entirely and there is continuous asexual reproduction. (Figures sourced from International Aphid Genomics Consortium 2010)

A. pisum is particularly useful for laboratory studies. Its small size, short generation time, an ability to maintain parthenogenetic morphs in incubators which mimic summer conditions and a rapid reproduction rate means large populations of single genotypes can be maintained easily (Brisson and Stern 2006). Nearly all *A. pisum* laboratory cultures can be maintained throughout their life cycle on a universal host plant *Vicia faba* meaning clones of different biotypes can be reared in identical conditions pre-experimentation. The presence of a universal host suggest that *A. pisum* races likely to share some host plant ranges, on which could mean in nature different races may occasionally co-occur and thus allow for gene flow to take place. The refinement of the electrical penetration graph (EPG) technique (Tjallingii 1987) means interaction between individual aphid stylets and specific host tissues

can be analysed in great detail. *A. pisum* was also the first hemimetabolous insect to have its genome sequenced and this information has become an invaluable resource for genetic research (International Aphid Genomics Consortium 2010).

The most important feature of *A. pisum* as a model of host race divergence is the occurrence of 11 or more host races and putative subspecies in Europe, each specific to one or a narrow range of Fabaceae species (Peccoud *et al.* 2009a). It appears this host race radiation is relatively recent, with the majority of diversification occurring within the last 3,600 to 9,500 years (Peccoud *et al.* 2009b) and *A. pisum* host races have been shown to be both ecologically separated and genetically divergent (e.g. Via 1999, Peccoud *et al.* 2009a, Via 2009). In Europe, Peccoud *et al.* (2009a) identified amongst 11 host associated biotypes and 3 possible sub-species of *A. pisum,* exhibiting a continuum in divergence from partial to complete reproductive isolation. Significantly, the availability of numerous *A. pisum* host races allows for multiple-contrast studies of host-race evolution (e.g. Peccoud *et al.* 2009a, Duvaux *et al.* 2015). *A. pisum* is, in fact, one of only a few cases that meet nearly all, if not all, of the five stringent criteria of a plant host race defined by Drès and Mallet (2002):

i. *The races use different host taxa, with each population exhibiting fidelity to a particular host:* Each of the known races is associated with a different species or group of closely related species in the field (Ferrari *et al.* 2006, Peccoud *et al.* 2009a). In choice tests, races exhibit clear preference for their native host (Caillaud and Via 2000, Via *et al.* 2000, Ferrari *et al.* 2006).

ii. *Hosts exist in sympatry in at least part of their range:* Alate aphid morphs are known to have large dispersal ranges (Loxdale *et al.* 1993, Compton 2002), possibly up to 300 km in *A. pisum* (Smith & MacKay 1989). Where host plants grow in sympatry they are therefore likely to be accessible to their respective *A. pisum* races. Field collection and genotyping of *A. pisum* biotypes in France and Germany confirms that they occur in sympatry in all, or parts of their range (Peccoud *et al.* 2009a), as is the case with the two host races in North America (Via 1999).

iii. *Races are genetically differentiated at more than one locus and are spatially / temporally replicable:* Hawthorne and Via (2001) located four complexes of quantitative trait loci (QTLs) linked to preference and performance traits in North American *Medicago sativa* and *Trifolium pratense* races, two of which only occur in the *M. sativa* race. European biotypes are genetically differentiated at multiple loci (Peccoud *et al.* 2009a, Duvaux 2015). Based on sampling at three sites in France and Germany, nine of the 11 *A. pisum* biotypes had greater genetic similarity between sites than with biotypes on other hosts found in sympatry (Peccoud *et al.* 2009a).

iv. *The races display correlation between host choice and mate choice, but still undergo gene flow (m≥1%):* It is reported that *A. pisum* mating only occurs on the host (Via 1999, Hawthorne and Via 2001). Given that in most instances both male and female forms are wingless (although winged males occur in some races (Braendle *et al.* 2005), it seems highly likely that most matings are on native host plants, although there appears to be no systematic study of this in the field. Between 11 European biotypes, the rate of hybridisation is variable but in eight it exceeds the arbitrary ≥1% threshold of gene flow (Peccoud *et al.* 2009a).

v. *Races exhibit higher fitness on their natal hosts than alterative hosts and produce hybrids that are less fit than parental forms: A. pisum* incurs a large fitness cost when transplanted reciprocally between host plants (Via 1991, Via *et al.* 2000, Ferrari and Godfray 2008, Peccoud *et al.* 2014). F1 hybrids of North American *T. pratense* and *M. sativa* host races also have lower fitness than each parent race on their native host plants (Via 2000), and hybrid inviability has also been demonstrated for F1 hybrids of five other *A. pisum* biotypes (Peccoud *et al.* 2014).

An additional critical issue is that for host race divergence to occur under gene flow *A. pisum* trait associations between host choice and performance on hosts need to resist the effect of recombination. Hawthorne and Via (2001) found, in two *A. pisum* races, tight linkage or pleiotropy of QTLs associated with fecundity, a measure of performance, and choice of host plants. Additionally, these QTLs are thought to have

antagonistic effects between plant species on both the host preference and performance (Hawthorne and Via 2001,Via and Hawthorne 2002, Via and West 2008). Overall *A. pisum* seems to be a ideal model species to ask whether host plant chemistry acts as an important selective force on traits associated with ecological divergence and early stages of speciation.

1.2.1 Evidence that plant chemistry is a source of divergent selection on *A. pisum* races

As sap sucking plant parasites aphids have an intimate relationship with their host plant through their feeding stylets and it has long been recognised that plant chemistry plays an important role in aphid interactions with plants (Pickett *et al.*1992, Dixon 1998 pp32-38, Powell *et al.* 2006). However, the plant chemistry involved in host preference, host acceptance and aphid performance is currently poorly understood (Peccoud and Simon 2010).

While visual cues could play a role in plant location (Powell *et al.* 2006), winged forms of *A. pisum* do not identify their native host plant at a distance, and reject alternate hosts only after a few minutes probing the leaf surface (Caillaud & Via 2000), a pattern of stereotypical behaviour encountered in other aphids (Powell *et al.* 2006) (Box 1.1). *A. pisum* responses to plant metabolites on and within the leaf are therefore likely to be the most important cues for plant discrimination, with the metabolites acting either as attractants or deterrents (Pickett *et al.* 1992, Caillaud and Via 2000, Powell *et al.* 2006).

It has been hypothesised by Smadja and Butlin (2009) that divergent selection on animal chemosensory systems plays an important role in speciation, including host selection. *A. pisum* behaviour provides some support for this. In a study using EPG, of feeding penetration by six *A. pisum,* clones on *Vicia faba* and *Pisum sativum,* Wilkinson and Douglas (1988) concluded that specific chemical cues inform host choice. Del Campo *et al.* (2003) found application of a crude extract of *M. sativa* to *T. pratense* leaves stimulated feeding by an *M. sativa* specialist *A. pisum* race, with a

similar stimulatory effect of the reciprocal treatment on the *T. pratense* race. To what degree the crude extracts represent intact plant chemistry and the identities of the chemicals involved are, however, unclear.

Insect perception of chemical stimuli is via a super-family of ligand-gated ion channels known as chemoreceptors (Cr) (International Aphid Genomics Consortium 2010). Two sub-families of Cr thought to be involved with aphid host acceptance are i) odorant receptors (Or) and ii) gustatory receptors (Gr) (Smadja *et al.* 2009, International Aphid Genomics Consortium 2010). Smadja *et al.* (2009) identified a total of 79 Or genes and 77 Gr genes in *A. pisum*, most of which were aphid-specific genes and many recently diverged. Smadja *et al.* (2009) also found the most recent duplications of Or and Gr genes had the highest rates of non-synonymous substitutions and concluded these genes had evolved under positive selection. This suggests that the *A. pisum* complex, perhaps even the Aphidoidea superfamily as a whole, has been positively selected to broaden its chemo-reception, presumably to discriminate between wider ranges of plant compounds during host selection ~~specialisation~~ (Smadja *et al.* 2009). In organisms with differently adapted race types different selection criteria are needed to choose the right host. It is possible selection on chemoreceptors in many animals could be a potential source of divergent selection necessary for ecological speciation (Smadja and Bultlin 2009). As further evidence of sensory divergence, targeted screening revealed that, between three *A. pisum* clones (adapted to *Lotus pedunculatus, M. sativa*, and *T. pratense)*, out of 175 candidate genes, only a handful of Cr genes had higher levels of between-clone differentiation than would be expected under neutrality (Smadja *et al.* 2012). Jaquiéry *et al.* (2012) also suggested a role for Cr genes in *A. pisum* diversification: out of 390 microsatellites only 11 loci had higher than neutral levels of differentiation, two of which were close to Or genes. These same two Or genes (along with only two other loci), were found to display higher levels of differentiation between a further eight *A. pisum* races (Nouhaud *et al.* 2014). More recent analysis of gene copy number variation (CNV) between *A. pisum* races has also suggested that Cr genes play an important role in *A. pisum* evolution. In a comparison of 434 genes across eight *A. pisum* host races, higher CNV variation between races was found in the Or, and to a lesser extent Gr genes compared to other

gene families, with the exception of the P450 gene family discussed below (Duvaux *et al.* 2015).

Non-sensory mechanisms in which plant chemistry could underlie divergent selection in *A. pisum* may also be important. Jaquiéry *et al.* (2012) found three of 11 divergent microsatellites were loci associated with aphid salivary proteins. This is interesting as aphid gelling and watery saliva play separate but highly critical roles in the aphid-plant interaction. Aphid gelling saliva contains cellulase and pectinase, creating enzyme products such as oligogalacturonides that are elicitors of plant defence (Will and van Bel 2008). Watery saliva is a mixture of a large number of proteins and metabolites some of which may act as elicitors of defence in plant tissues (Giordanengo *et al.* 2010, Carolan *et al.* 2011). As aphid stylets pass through the apoplast to the phloem, they regularly penetrate cells and watery saliva is thought to sample cell chemistry (Figure 1.2 and 1.3) (Tjallingii and Esch 1993). Salivary proteins could also have a chemosensory purpose as they could be transporting chemical cues form the plant tissues to aphid sensory organs. However, there is a growing body of evidence that many aphid salivary proteins do have multiple direct roles in aphid-host plants interactions (Bos et al. 2010, Hogenhout & Bos 2011).

Aphid watery saliva also acts as a medium to introduce effector proteins that manipulate and suppress plant responses (Carolan *et al.* 2011, Sharma *et al.* 2014) and differences in watery saliva composition have been observed between two *A. pisum* clones of the same race (Will *et al.* 2009). There is much potential for *A. pisum* salivary proteins to function as adaptations to deal with plant chemical defences (Carolan *et al.* 2011). For instance, enzymes such as M1-zinc metalloprotease and CLIP-domain serine protease have been identified in *A. pisum* saliva (Sharma *et al.* 2014), enzymes that are known to alter host plant-defence mechanisms (Carolan *et al.* 2009, Carolan *et al.* 2011). Aphid watery saliva has also been linked to the suppression of a plant protein coagulation process known as phloem occlusion, which acts as a way to seal compromised phloem vessels.

Box 1.1: Host selection and acceptance behaviour in *A. pisum*

The majority of host location and acceptance is performed by winged (alate) morphs. Although alate flight is weak it can be sustained for up to 16 hours (Caillaud and Via 2000, Powell *et al.* 2006). This allows aphids to reject plants of the wrong type and continue foraging until the appropriate host species is found. There is a conserved sequence of behaviour that leads to acceptance or rejection of plants (Pickett 1992, Wilkinson and Douglas 1998). It is described by Powell *et al.* (2006) in the following stages.

1. PRE-ALIGHTING: During flight alates can perceive and respond to both visual cues and plant volatiles to locate hosts.
2. PLANT CONTACT AND ANTENNATION: As aphids land they come into contact with possible leaf surface cues and surface volatiles.
3. STYLET PROBING: Aphid stylet penetration of leaves is a stereotyped reflex triggered by tarsal contact with any surface. At this point aphids penetrate the leaf surface briefly (<1min) several times, before aphids either proceed to enter the sieve element of the phloem or reject the plant.
4. STYLET PATHWAY ACTIVITY: Stylet penetrations of 0.5 to 1 minute pass through the plant mesophyll and parenchyma in search of the phloem. During this period the aphid excretes gelling saliva which hardens around the stylet to protect it (Figure 1.2).
5. SIEVE ELEMENT PENETRATION: The stylet eventually reaches the sieve element where it punctures the cell wall and membrane to enter the phloem. At this point aphids exude watery saliva into the phloem which is thought to suppress plant defences, notably phloem occlusion (Figure 1.3).
6. FEEDING FROM THE PHLOEM: Once reached aphids may continue to feed from the phloem for many hours.

Several studies of *A. pisum* and other species have shown that host acceptance occurs during different parts of the plant probing stage (Wilkinson and Douglas 1998, Pickett *et al.* 1992**,** Caillaud and Via 2000, Gao *et al.*2008, Powell *et al.* 2006). As acceptance can happen before the stylet enters the phloem, plant compounds that influence host acceptance may not be exclusively found in the phloem (Caillaud and Via 2000). This suggests *A. pisum* may respond not just to plant nutritional cues but also to metabolites in other parts of the leaf (Caillaud and Via 2000).

As aphid stylet moves through the apoplast contact with the cell wall chemistry is likely (Figure 1.2). Studies by Tjallingii and Esch (1993) on *Aphis fabae*, show that stylets regularly penetrate into the cytoplasm of plants cells. It is possible that sampling of host cytoplasm may also occur. This cell sampling could be the stylet searching for the sieve element (Will and van Bel 2008), though aphids may also be suppressing the plant defence cascade through the leaf (Powell *et al.* 2006). During probing aphids may also perceive host surface cues (Pickett *et al.* 1992). For instance, aphid rejection may be a reaction to defensive volatiles that many plants rapidly produce in response to attack (Arimura *et al.* 2009).

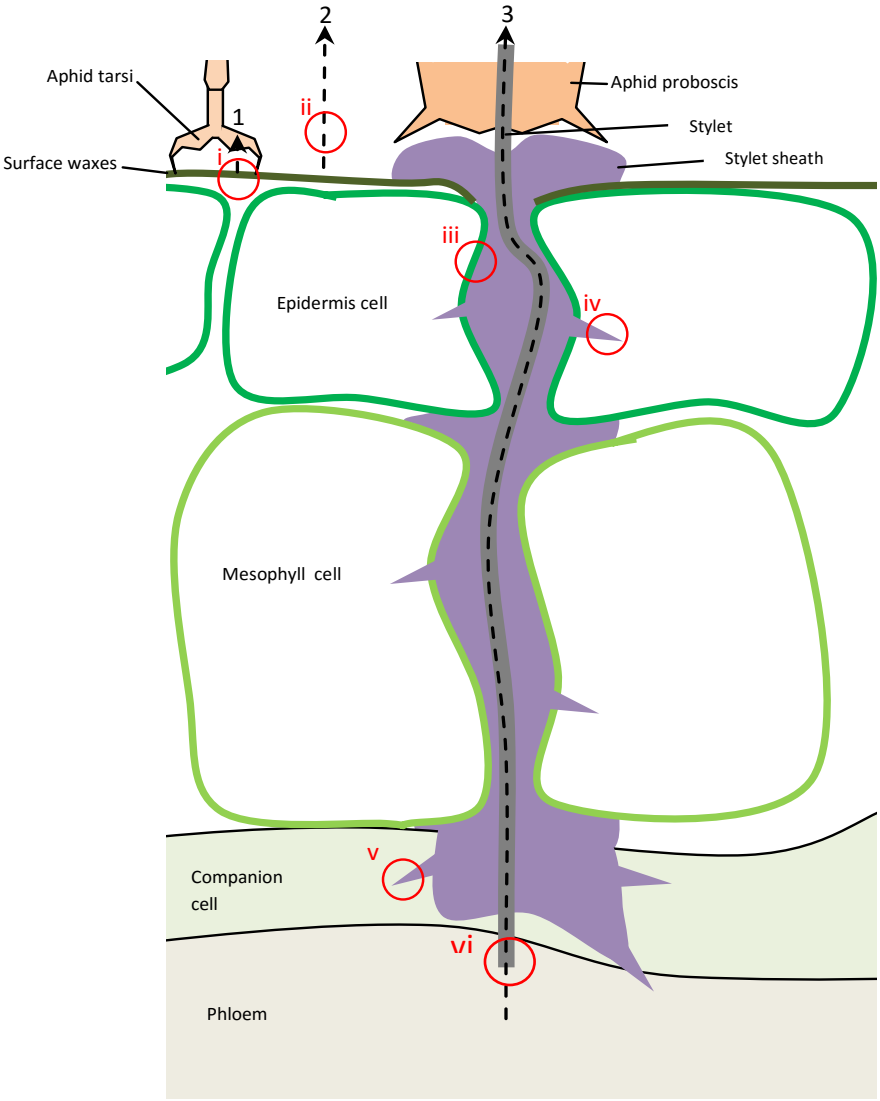


Figure 1.2: Path of aphid stylet through the plant leaf apoplast during plant probing. The diagram shows the sheath (purple), formed by the secretion of gelling saliva to encase the stylet, and the stylet itself (grey line) as it passes through plant tissues. Red circles indicate possible sites of aphid-plant interactions during the probing period including i) surface waxes, ii) released volatiles, iii) cells walls, iv) inside epidermal cells, v) companion cells and vi) in the phloem. Dotted arrows indicate the possible path of plant cues to chemosensory organs of the 1) aphid tarsi, 2) aphid antennae and 3) gustatory organ. Diagram based on Tjallingii (2006).

.

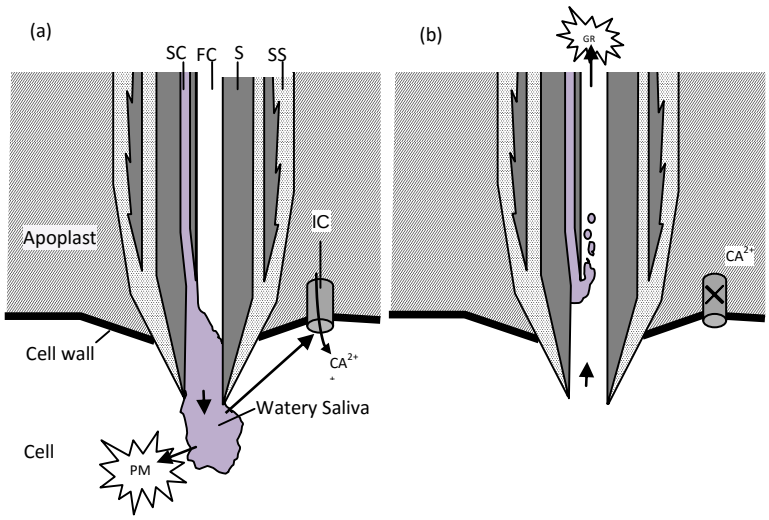


Figure 1.3: Aphid stylet as it penetrates a cell. The diagram shows the stylet (S), stylet sheath (SS), salivary canal (SC) and the food canal (FC). Arrows indicate direction of flow. Aphids first (a) secrete watery saliva into a plant cell where it interacts with plant metabolites (PM) and/or inhibits plant signalling via plant ion channels (IC). The watery saliva is then (b) re-ingested by the aphid, where PM can be perceived by the gustatory receptors (GR). Diagram based on Dixon (1998).

Phloem occlusion is triggered by a calcium ion (Ca^{2+}) flux mediated by calcium channels in the phloem cell membrane (Will and van Bel 2008). As the aphid stylet enters the phloem, it is likely to cause sufficient mechanical and chemical disturbance to trigger a Ca^{2+} flux and phloem occlusion must be counteracted if aphids are to feed. Proteins such as C002 in the watery saliva are thought to act as Ca^{2+} binding proteins (Figure 1.3) preventing protein coagulation (Sharma *et al.* 2014) and silencing the C002 gene in *A. pisum* reduced the ability of aphids to colonize *V. faba* (Mutti *et al.* 2008). Whether abilities to subvert plant defences are divergent between *A. pisum* races is currently untested.

Cytochrome P450 proteins are involved in detoxification of plant allelochemicals and could also aid aphid adaptation to host plants (Simon *et al.* 2015). Ramsey *et al.* (2010) compared the *A. pisum* genome to the partial genome of a host generalist, the

peach-potato aphid (*Myzus persicae)* and found *A. pisum* possessed around 40% fewer P450 genes. This suggests *A. pisum* races have a reduced capacity to detoxify as an evolutionary consequence of narrower host plant ranges and so exposure to fewer allelochemicals. This last study only looked at a single *A. pisum* genome but in the comparison of CNV in eight *A. pisum* biotypes by Duvuax *et al.* (2015) completely duplicated or deleted variants in P450 gene family were also found to be highly polymorphic. This CNV variation in P450s was interpreted as potentially contributing to divergence between *A. pisum* races in their resistance to plant toxins (Duvaux *et al.* 2015).

Taken together, the above evidence indicates that through divergence in chemoreceptors and also possibly alongside salivary and P450 proteins, *A. pisum* race formation involves multiple trait adaptation to host chemistry. It further supports the recommendation of Huang *et al.* (2011) that *A. pisum* should be developed as a model for the study of the interaction between insects and plant chemical defences. However, the genetic studies mentioned above have yet to identify the precise features of plant chemistry, nor the mechanisms they act on in aphids, to cause the divergent selection necessary for host race formation.

1.2.2) Fabaceae as diverse chemical environments for *A. pisum* races

The Fabaceae ("legumes") is third largest angiosperm family with over 19,400 species and is globally widespread (Wojciechowski *et al.* 2004). After the grasses (Poaceae) the Fabaceae is the most economically important plant family and so has been extensively researched (Wojciechowski *et al.* 2004). The Fabaceae produce an exceptionally wide range of chemicals; over 4,000 are recorded (Bisby 1994). This is significantly due to a diversity of secondary metabolites, many of which are known to have pharmacological or toxicological properties against vertebrates, insects and microbes and many are thought to serve in plant defence (Walters 2011). These include: i) alkaloids and amines including non-protein amino acids (NPAAs), ii) peptides (including lectins and protease inhibitors), iii) simple and compound phenolics, iv) terpenoids, v) carbohydrates and vi) organic acids (Wink 2013). The distribution of many secondary metabolites amongst species and genera of the

Fabaceae shows only weak correspondence to phylogeny (Wink 2003, Wink 2013). Given this, *A. pisum* radiation across the Fabaceae is not particularly strongly linked to the Fabaceae phylogeny (Peccoud *et al.*2009b), it seems *A. pisum* divergence is less likely to be the result of a co-evolution with host plant divergence but instead a response to divergent selection by highly diverse plant chemical environments. In this context the divergence of the *A. pisum* P450 gene family involved in detoxification and chemoreceptor gene families appears particularly interesting (Duvaux *et al.* 2015).

Due to most Fabaceae species' ability to fix nitrogen, secondary metabolites containing nitrogen, (e.g. quinolizidines, cyanogenic glycosides and NPAAs) are a distinctive feature of the Fabaceae. For instance, approximately 250 NPAAs have been found in plants and they are especially diverse in the Fabaceae (Vranova *et al.* 2011). Whilst some NPAAs (e.g. L-ornithine, L-homoserine) function in primary metabolism others are toxic to insects, either as analogues of protein amino acids which disrupt protein synthesis (e.g. L-cavanine, L-mimosine) or because they interfere with neurotransmisson (e.g. L-DOPA, GABA) (Huang *et al.* 2011). In addition, some alkaloid families are abundant within the Fabaceae (Wink and Mohamed 2003, Wink 1992). This includes the quinolizidines, such as sparateien, lupinine and cystine, which are known to act as potent feeding deterrents to *A. pisum* aphids (Wink 1992). Another class of defence compounds found across the Fabcease is cyanogenic glycosides, which are rapidly metabolised to release cyanide-based compounds upon wounding, with strong anti-herbivore effects (Ballhorn *et al.* 2010). However, as cyanogenesis requires damage to vacuole membranes and aphids cause typically only minor cell disruption this may not be relevant to aphid-plant interactions. Nonetheless the diversity of unusual toxic compounds found in the Fabaceae means an ability to tolerate or detoxify specific allelochemical blends may play an important role in the divergence of *A. pisum* host races (Huang *et al.* 2011).

Fabaceae secondary metabolites also act as signalling molecules (Wink 2013), including plant hormones used in the regulation of plant homeostasis and defence. Widely found in plants, the hormone jasmonic acid (JA) mainly triggers plant defences against insects, whilst salicylic acid (SA) is mainly involved in signalling

defence against pathogen attack (Walters 2011). *A. pisum* is thought to interact with host chemical defences by the manipulation of plant signalling pathways, specifically the JA and SA pathways (Gao *et al.* 2008, Takemoto *et al.* 2013, Schwartzberg and Tumlinson 2014). In this way, the aphids elicit decoy defences to divert JA-regulated defences into less harmful SA-regulated defences (Thompson and Goggin 2006, Will and van Bel 2008, Gao *et al.* 2008). Takemoto *et al.* (2013) found that in *V. faba* plants, pre-exposure to *A. pisum* not only improved performance of a new *A. pisum* inoculum but also decreased levels of JA concentration in exposed plants. This facilitation effect is negated when plants were pre-treated with JA (Takemoto *et al.* 2013) presumably due to the negative co-regulation of JA and SA. Gao *et al.* (2008) also found that *A. pisum* infestation of aphid resistant *M. truncatula* plants did not induce large changes in the expression of genes of the octadecanoid pathway, which leads to the production of JA. This was in contrast to the response to blue green aphid (*A. kondoi*) infestation, where genes involved in the JA pathway were exclusively or predominantly induced (Gao *et al.* 2008).

Fabaceae species also have a unique phloem occlusion mechanism involving coagulation proteins known as forisomes (Will & van Bell 2008). There is evidence that forisome coagulation depends on Ca2+ ion channels, located on the endoplasmic reticulum and plasma membrane of sieve elements (Hafke *et al.* 2009). It has been suggested that *A. pisum* Ca2+ binding proteins found in aphid saliva could be associated with suppression of forisome coagulation (Will and van Bel 2008). *A. pisum* co-evolution with different host induced chemical defences such as JA/SA defence pathways and forisome coagulation could be one source of chemically based divergent selection among *A. pisum* host races.

In addition to defence compounds *A. pisum* races could also be discriminating betwen Fabaceae based on their nutritional quality. For example *A. pisum* feeding on *V. faba* caused more nutritionally beneficial changes to host plant protein amino acid content than the generalist vetch aphid (*Megoura viciae*) (Leroy *et al.* 2011). Carrillo *et al.* (2014) also showed that pea (*Pisum sativum*) plants exposed to native or non-

native *A. pisum* aphids caused differential expression of various proteins involved in amino acid and carbohydrate metabolism, photosynthesis and stress response signalling. This evidence suggests native aphids have co-evolved with the chemistry of at least some Fabaceae plants to improve host quality, while non-native aphids may instead induce plant defences and reduce host quality (Carrillo *et al.* 2014). Further as *A. pisum* clones have been shown to differ in their requirements for particular essential amino acids (Vogel and Moran 2011), divergence in host plant species nutritional qualities may also be an important source of differential selection among *A. pisum* races.

On balance, the evidence currently available supports the view that aphids that adopt the Fabaceae as hosts require multiple adaptations to host chemistry. The striking discontinuities in chemistry between Fabaceae genera and species have likely played an important part in *A. pisum* host race divergence. This makes *A. pisum* clones and their various Fabaceae host plants ideal models for further study on the chemical ecology of host race formation

1.3 Overview of thesis

In the first half of this thesis the acceptance profiles (i.e. the willingness of an aphid to feed upon a given host) of four *A. pisum* clones (belonging to two races) on nine *Medicago* and ten *Trifolium* species are correlated with the metabolic content of the plant species. In the work described in Chapter 2, various EPG measurements were developed into metrics to profile acceptance of the two *A. pisum* races across the 19 host plants. These acceptance profiles are then compared to performance measures (fecundity, aphid condition score) to test if they correlate strongly. In Chapter 3 the acceptance profiles of plants were then compared using random forest regression to metabolic profiles, collected using MALDI-TOF mass spectrometry, of leaf extraction from the same plant species to determine which putative masses are strongly associated with aphid acceptance. Where possible masses were then identified using online data bases (www.biocyc.org, www.genome.jp/kegg/pathway). Putative masses were subjected to tandem mass spectrometry and their fragmentation

patterns compared to those of synthesised metabolomic standards of putatively identified compounds. Identified compounds were then assessed for their relationship to each other and their metabolic significance, via pathway analysis.

The second half of the thesis sets out to explore the role of plant compounds in divergent host adaptation of *A. pisum* races. In Chapter 4, *M. sativa* and *T. pratense* plants were exposed to a native *A. pisum* clone, a non-native *A. pisum* clone or neither, to test the effect of pre-exposure on plant acceptability to a second inoculum of aphids. Aphid-induced changes to target mass in the plants were also measured. This was to test whether plant responses to two different clones were fixed (constitutive), activated by aphid attack (induced) or prevented by aphids (suppressed) and identify the effect of plant exposure to aphids on the concentration of compounds of interest, identified in Chapter 3. Finally, in Chapter 5, two plant compounds correlated with aphid acceptance were offered in artificial diets to see whether they have any direct effect as feeding deterrents or attractants.

Chapter 2: Profiling host plant acceptance and performance by two *A. pisum* races across multiple species of the plant genera *Medicago* and *Trifolium*.

2.1 Chapter summary

Theoretically for divergent host plant races to evolve in phytophagous insects there needs to be a tight association between insects acceptance of host plant and its abilities to perform well on that same plant. To test this association requires a reliable measure acceptance and performance. However, measuring host acceptance in aphids has proved challenging, resulting in a wide variety of methods to measure. This makes comparing results of aphid acceptance across studies challenging, particularly as some methods may use poorer proxy measurements of acceptance behaviours than others. The aims of this research are to develop a reliable method to profile pea aphid (*Acyrthosiphon pisum)* acceptance using electrical penetration graphs (EPG) and to test the association between acceptance and performance, across a range of typical and non-typical host plants in the genera *Medicago* and *Trifolium*.

Aphid acceptance was assessed using five individual waveform measurements: i) duration of E1, ii) duration of E2, iii) number of rpd , iv) number probes and v) time to E2. In addition acceptance was measured using the first two axes of a principle component analysis (PCA) and a linear discriminant analysis (LDA) of 60 non-correlated EPG waveform measurements. Acceptance profiles were recorded for four *A. pisum* clones, two from the *Medicago sativa* adapted race (MS aphids) and two from the *Trifolium pratense* adapted race (TP aphids). This involved testing acceptance on 19 different species of *Medicago* and *Trifolium*. With the exception of duration of E1, PC2 and LD2 scores, the EPG profiles used recorded broadly similar patterns of acceptance or rejection of plants species. Each aphid race exhibited a continuum from high levels of acceptance to high levels of rejection across the 19 plant species, with striking differences between the two races. EPG profiles were then correlated with measure of performance (adult fecundity and adult quality), on 16 of the *Medicago* and *Trifolium* species. This identified a consistent tight association between acceptance and performance.

2.2 Introduction

A fundamental challenge in explaining ecological speciation with gene flow is how traits that select for fitness can co-evolve with traits that cause reproductive isolation. In the context of host race formation this question is specifically how physiological adaptations to perform well on a host can evolve with behavioural traits to select the correct host to live on. However, while measuring features that constitute performance on a host plant are usually fairly clear (e.g. number of young in a life time and longevity), finding the appropriate measurements to accurately quantify a behavioural traits like acceptance are less obvious.

Quantifying phytophagous insect host plant specificity is particularly challenging because measurements of the behavioural traits associated with *acceptance* of a host plant (i.e. traits that allow an insect to recognise a suitable plants before or during early feeding stages) are often difficult to separate from traits associated with insect *performance* on host plants (i.e. adaptations that allow an insect to survive and reproduce on a host). This difficulty in separating acceptance and performance is especially the case for plant phloem-feeding insects, such as aphids, where most feeding behaviour occurs at a fine scale within the leaf and is consequently difficult to observe (Powell *et al.* 2006).

These considerations are relevant to research on the *A. pisum*, where it is thought differences in host plant choice act as a pre-mating isolation barrier between diverging host races (Peccoud *et al.* 2010, Peccoud and Simon 2010). Four complexes of quantitative trait loci (QTL) have been identified in *A. pisum* that are associated with both acceptance and performance (Hawthorne and Via 2001), which suggests genes for acceptance and performance are under tight genetic linkage or pleiotropic alleles are involved (Hawthorne and Via 2001). Consequently it is theorised that *A. pisum* race divergence is a consequence of host plant specialisation through the genetic trade-off of these two traits in tandem (Hawthorne and Via 2001, Via and Hawthorne 2002).

Many studies that have investigated *A. pisum* host specificity use measurements of performance such as fecundity and adult survival (Table 2.1). These measurements are relatively easy to collect, analyse and interpret. However, some acceptance decisions are likely to occur before or at the very beginning of the feeding stage, well before performance measurements are recorded (Wilkinson and Douglas 1988, Pickett *et al.* 1992, Caillaud and Via 2000, Gao *et al.* 2008, Schwarzkopf *et al.* 2013). Because host acceptance occurs so early, studies that measure aphid performance may not be adequate proxies for measuring acceptance traits. Further acceptance and performance are processes which are likely to involve quite different mechanisms.

While some studies present performance measurements alongside acceptance measurements, these experiments use varied methods to measure acceptance (Table 2.1). This diversity of approaches makes it difficult to compare acceptance results across studies since it is unclear which measurements for acceptance are reliable. In addition, these studies rarely test the strength of the relationship between acceptance and performance quantitatively.

Several studies have used fecundity rates of adult aphids after the first 24 hours or more they are exposed to plants, as a measure of acceptance (Del Campo 2003, Ferrari and Godfray 2006, Ferrari *et al.* 2007, Ferrari *et al.* 2008). As Powell *et al.* (2006) discuss, studies have shown that early fecundity rates are linked to the acceptance of a plant, and that reproduction in aphids is likely to be triggered by plant leaf sub-epidermal chemistry (Tosh *et al.* 2002) . The measurement periods for acceptance used in these studies (varying from 24-70 hours) are assumed to be before any negative effect of feeding can impact on the aphid's ability to produce young (Ferrari *et al.* 2008). However, using fecundity rate over more than 24 hours risks overlooking key critical aphid decisions about acceptance that occur in the first few minutes to hours after aphids are exposed to a plant. In addition, early fecundity measurements overlook the effect of rapidly induced defences that might not affect early acceptance but impact on aphid performance in the longer term (Arimura *et al.* 2011, Walters 2011). An exception to using fecundity after 24 hours is work by

Table 2.1: Summary of studies on A. pisum that measure aphid acceptance of different plant types

				Acceptance measures			
Studies Of *A. pisum*	Plant used	Performance Measured (what measured)	No. young produced (over what period)	Choice (number of plant types)	Artificial diets	Behaviour on leaf (how measured)	EPG
Via 1999	*Medicago sativa* *Trifolium pratense*			Yes (2)			
Caillaud and Via 2000	*M. sativa* *T. pratense*		Yes (30 min)	Yes (2)		Yes (leaf Penetration, + tracking software)	Yes
Hawthorn and Via 2001	*M. sativa* *T. pratense*	Yes (demography + fecundity)	Yes (70 hrs)				
Gao *et al.* 2008	*M. truncatula* (2 varieties)	Yes (survival and growth)		Yes			Yes
Wilkinson and Douglas 1998	*Pisum sativum* *Vicia faba*						Yes
Ferrari and Godfray 2003	*Lotus. uliginosus* *T. pratense*	Yes (survival +fecundity)	Yes (24 hrs)				
Ferrari and Godfray 2006	*5 species ǂ*	Yes (survival +fecundity)	Yes (24 hrs)				
McLean *et al.* 2009	*L. pratensis* *V. faba*	Yes (fecundity)		Yes (2)			

Table 2.1: Continued

			Acceptance measures				
Studies Of *A. pisum*	Plant used	Performance Measured (what measured)	No. young produced (over what period)	Choice (number of plant types)	Artificial diets	Behaviour on leaf (how measured)	EPG
Ferrari *et al.* 2006	8 *species* †		Yes (24 hrs)	Yes (8)			
Ferrari *et al.* 2007	*T. pratense*	Yes (survival + fecundity)	Yes (24 hrs)	Yes(2)			
Ferrari *et al.* 2008	8 *species* †	Yes (survival + fecundity)	Yes (24 hrs)				
Del Campo *et al.* 2003	*M. sativa* *T. pratense*		Yes (48hrs)	Yes (2 diets)	Yes	Yes (visual observation)	
Schwarzkopf *et al.*(2013	*M. sativa* *P . sativum* *T. pratense* *V. faba*	Yes (survival + weight)					Yes
Kordan *et al.* 2011	*P. sativum* (2X treatments)						Yes
Sauvion *et al.*2004	Artificial diets used			Yes (2 diets)	Yes		Yes (on diets)

† *V. faba, Lathyrus pratensis, P. sativum ,T. pratense , M. sativa , Lotus pedunculatus, Cytisus scoparius ,* and *Ononis spinosa.*
‡ *V. sativa, L. pratensis, T. pratense, L. pedunculatus, C. scoparius.*

Caillaud and Via (2000) who recorded the presence or absence of offspring in the first 30 minutes of plant exposure.

Artificial diets can also be used to assess *A. pisum* host acceptance, for example by examining ingestion rates or preference between artificial diets (Sauvion *et al.* 2004) or plant extracts (Del Campo *et al.* 2003). Using the latter method Del Campo *et al.* (2003) were able to show that extracts from *Trifolium pratense* and *Medicago sativa* elicited discriminatory acceptance behaviour between divergent aphid clones. However the use of plant extracts is questionable as many plant chemicals are likely to be altered by the extraction process and are removed from their natural context within the plant.

Del Campo *et al.* (2003) also used observational data of aphid behaviour on leaves. These observations included the frequency of aphid leaf probing, the extent of aphid movement on leaves and the tendency of aphids to leave the leaf. Caillaud and Via (2000) also recorded aphid activity using automated tracking software to quantify movement. A problem with these observational approaches remains that many aphid choice behaviours occur beneath the plant cuticle (Powell *et al.* 2006) making it difficult to know how these externally observed measures actually relate to aphid host acceptance mechanisms.

It is possible to measure acceptance for *A. pisum* more directly by providing a choice between various plant species and ranking the acceptance of these options, either as paired choices (Klingler *et al.* 2005, Gao *et al.* 2008) or as choices among several plants within the confined space of a choice arena (Caillaud and Via 2000, Ferrari *et al.* 2006, Ferrari *et al.* 2007). This is by far the most naturalistic experimental test that can be done in the laboratory. However, even this technique has limitations. Firstly, it does not distinguish between: i) pre-landing preference caused by visual and volatile chemosensory cues, ii) post-landing acceptance caused by surface and internal plant chemistry cues, and iii) early feeding success as influenced by plant defence mechanisms and food quality. A further limiting factor with choice tests is

that they can only provide a score in relation to the plant species and aphid clones used in a given experiment. Consequently, it can be difficult to relate results to wider contexts and other data types or choice experiments that do not use the same plant species and aphid types. Finally, choice tests become difficult to carry out when dealing with an increasing number of treatments, i.e. when a large number of plant species and aphid clones are tested. The number of tests that needs to be performed increases rapidly in paired designs and it is impractical to run an experiment where aphids have to choose between large numbers of plants within a restricted physical space. As a consequence, laboratory based choice studies tend not to measure acceptance beyond a limited number of plant species at a time (with a maximum of eight used by Ferrari *et al.* 2006). This is a particular issue with *A. pisum* as it has a minimum of 11 biotypes, described from field collection, that exhibit genetic differentiation ranging from host races to possible cryptic species (Peccoud *et al.* 2009a).

An additional scientific weakness of the above methods is that they offer limited possibilities to explore the behavioural mechanisms involved in acceptance. Further, the ideal measurement of acceptance would be quantitative, reliably repeatable and versatile when using different types of statistical analysis. This would allow comparison of results to other experiment types and give the felxibilty to contrast results across studies. The electrical penetration graph (EPG) technique (Tjallingii 1978) provides such data. EPG is ideally suited to studying the initial stages of plant probing, as it continuously records the activity of insect stylets during these critical early acceptance processes. Consequently, EPG not only can be used to quantitatively measure acceptance but also offers opportunities to explore the behavioural mechanisms involved in acceptance. This mechanistic information can be very useful in understanding aphid-plant interactions as they provide evidence t when and where aphid acceptance to plants might occur. In terms of investigating ecological speciation demonstrating different mechanisms of acceptance between plants could reveal possible sources of divergent selection.

The EPG technique works by measuring changes of voltage in an electrical circuit that passes between a plant and an aphid (Figure 2.1). It records the changes in circuit resistance as an aphid (or any other sucking insect) stylet passes through plant tissue (Figure 2.2). This creates distinct and stereotypical patterns in voltage changes over time, known as *waveforms*, as characterised in Tjallingii and Esch (1993) (see Appendix 1 for full details of EPG waveforms patterns). These waveforms have been interpreted as reflecting specific aphid behaviour patterns (Tjallingii 1978, Tjallingii and Esch 1993, Tjallingii 2006).

Waveform patterns recorded by EPG include:

i. probing of the leaf with the aphid stylet (probe)

ii. aphid stylet searching through the apoplast (C)

iii. stylet entering and feeding from the xylem (G)

iv. stylet derailment(F)

v. stylet within the phloem (E)

vi. salivation into the phloem (E1); the first stage of waveform E.

vii. phloem feeding (E2); the second stage of waveform E.

viii. repeated probing of the phloem also known as repetitive potential drops (rpd). Rpd is associated with repeated penetration of the phloem sieve element before entering it. Rpd is a poorly understood behaviour common in *A. pisum* but thought to represent an early interaction with the phloem just before waveform E occurs (Tjallingii and Gabrys 1999, Schwarzkopf *et al.* 2013).

EPG is suitable for quantitative analysis as it is possible to take measurements of the individual waveforms in terms of their duration, frequency and time to onset. Of the numerous potential EPG waveform measurements possible, many have been identified as having importance to host plant acceptance by aphids in previous research (e.g. Via and Caillaud 2000, Gao *et al.*2008, Tjallingii and Gabrys 1999,

Schwarzkopf *et al.*2013). This is because they represent important stages in the feeding behaviour of aphids. For example, Gao *et al.*(2008) were able to show with EPG that *A. pisum* feeding on aphid resistant strains of *M. truncatula* had reduced durations of E2 and number of rpd. Caillaud and Via (2000) used EPG to show that waveform behaviour on *M. sativa* and *T. pratense* such as plant probing durations, rpd and E successfully predicted *A. pisum* clones' host plant biotypes. Schwarzkopf *et al.* (2013) compared the EPG waveform data of three host races of *A. pisum* across four putative host species. This study showed that predominantly early probing and E associated behaviours were important in *A. pisum* selection between host and non-host plants. However, other probing behaviours, such as rpd, appeared to be linked to divergence acceptance by *A. pisum* (Schwarzkopf *et al.* 2013). EPG is therefore a versatile and simple technique that can be used in a wide range of experimental designs to give quantitatively comparable results of aphid acceptance and a wealth of information to use for interpretation.

EPG does have some disadvantages. First, during recording aphids are tethered to a thin wire (Figure 2.1c) which can stress aphids over long periods. Although tethered aphids have been shown to produce similar numbers of offspring (Powell *et al.* 2006) and perform a similar number of probes compared to non-tethered aphids (Powell and Hardie 2001), it is generally advised that EPG recordings should be no longer than 6 hours. Fortunately, in studies that are interested in the very early stages of feeding, this is not a significant issue. In addition, as an aphid is tethered to a single plant it is likely that an aphid on a non-host that would normally leave the plant is instead forced to perform behaviours that would not naturally occur. This is why in the interpretation of EPG needs to be considered in relation to the behaviour of other plants which have expected good and poor acceptance behaviours. Another challenge is that EPG produces a large number (<100) of measurement types (Sarria *et al.*2006). With so many measurements, many highly correlated, EPG data can be difficult to interpret. Further, attempting to analyse too many waveform measurements causes a statistical issue if the number of parameters is higher than number of observations, resulting in a multiple comparisons problem, i.e. increasing the chance of Type 1 error. As EPG waveforms vary in their biological significance depending on the insect or host plant species used (Tjallingii and Gabryś 1999),

selecting waveforms for analysis may not always be straightforward. Consequently, it could be useful to develop a method to reduce the multitude of waveform measurements to a single value which encompasses as much of the information EPG provides as possible, rather than selecting a subset of measurements. This value would be conceptually less biased towards any particular treatment in a given experiment or in cross-study comparisons.

To address some of these issues in handling EPG data, the objective of this study was to compare five pre-selected waveform measurements of functional significance with a multivariate analysis of 60 waveform measurements, using PCA and LDA.

The five individual measurements were:

i. The total duration of time that the aphid spends salivating into the phloem (duration of E1).

ii. The total duration of time an aphid spends feeding from the phloem (duration of E2).

iii. The total number of times the aphid's stylet probed the leaf (number of probes) .

iv. The time from the first probe to when the first sustained E2 lasts longer than 10 minutes (time to E2 >10 min).

v. The number of rpd behaviours (number of rpd).

The first objective was to assess the consistency of the various EPG measurements as a predictor of host plant acceptance. The second objective was to test the prediction that reliable EPG measurements of acceptance correlated with measures of performance, in terms of fecundity and adult quality. This was to test the hypothesis that fitness on a given host plant requires these adaptations are strongly correlated. Such evidence would further support the theory that traits for performance and preference in *A. pisum* are linked (Hawthorn and Via 2001), as necessary for host races to evolve under gene flow.

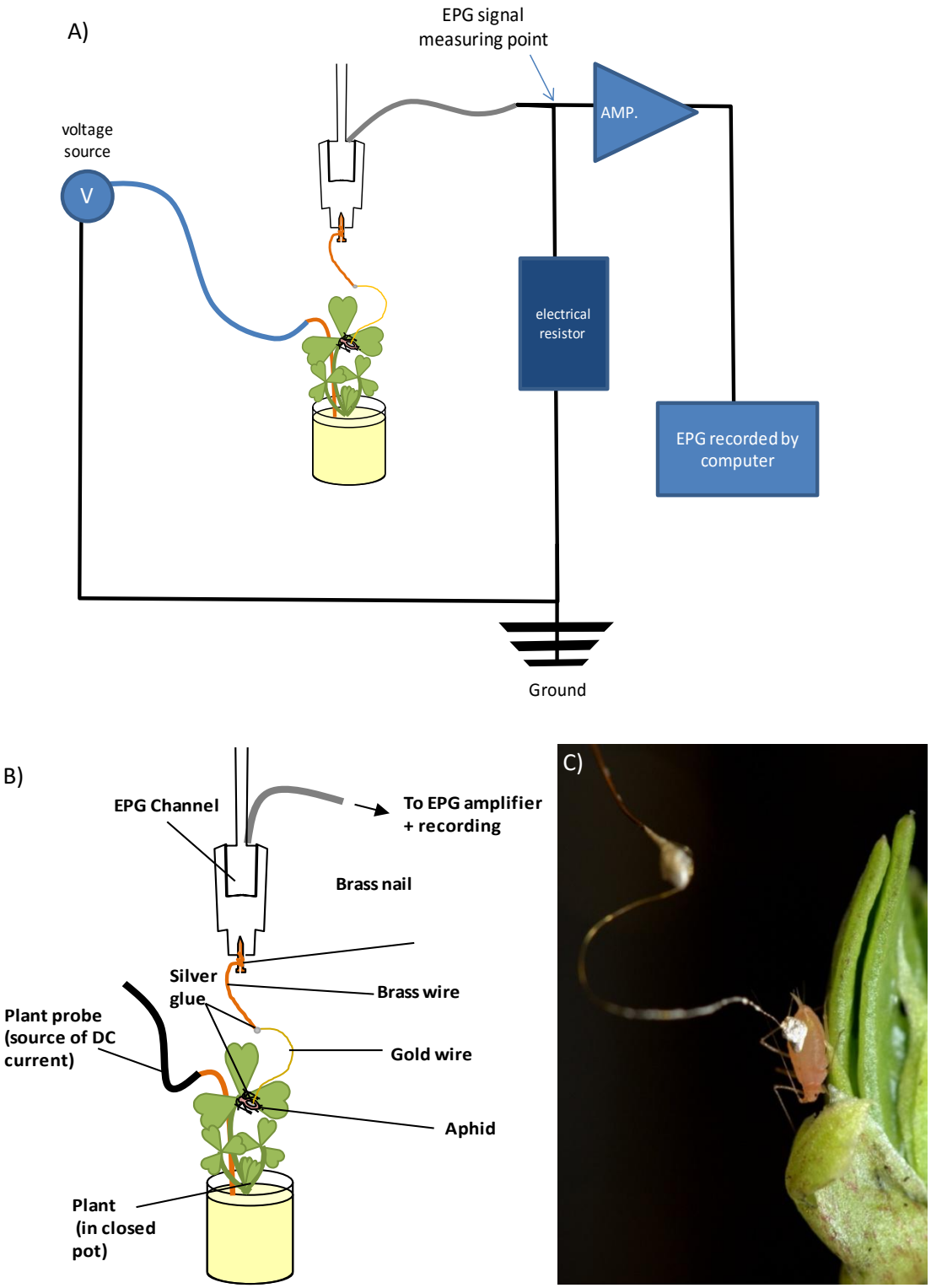


***Figure 2.1: Typical set up of Electronic Penetration Graph (EPG) equipment*.** A) The EPG circuit in which a current is passed from the plant to the aphid to allow changes in resistance to be measured. The amplifier (amp) amplifies the signal. The output is connected to a computer for recording. The plant output voltage is adjustable by the voltage source electrode (diagram adapted from Tjangilli 2006). B) A detailed diagram of how an aphid is wired to the EPG apparatus via a thin piece of gold wire to allow aphids to move freely. C) Photograph of a wired-up aphid on a *V. faba* plant.

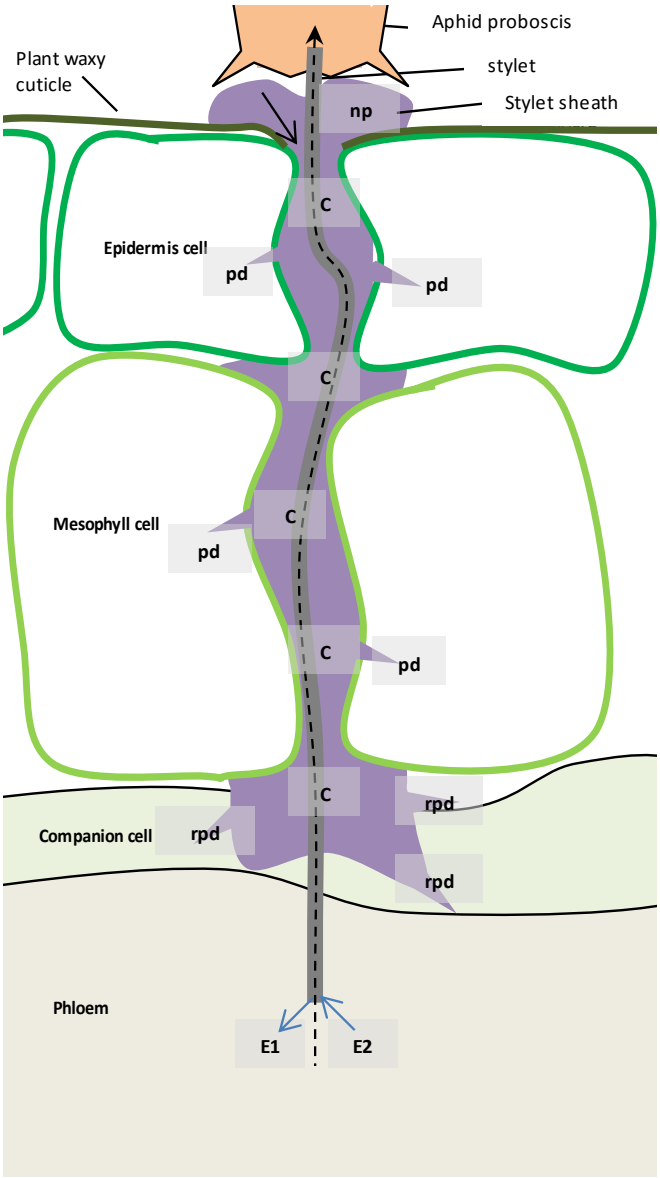


Figure 2.2: EPG waveform sequence within the leaf. The EPG signal starts when the aphid probes the leaf (shown by arrow) and moves through the apoplast (C) of the plant. Occasionally the stylet enters the cell to cause potential drops (pd). At, or close to, the phloem there are multiple cell punctures, causing repetitive potential drops (rpd), then the stylet enters the phloem and egests saliva (E1) before ingesting sap (E2). Xylem feeding (G) and stylet derailment (F) are not shown.

Four aphid clones were used from two genetically divergent host races within the *A. pisum* species complex. These were specialised upon either *M. sativa* (MS aphids) or *T. pratense* (TP aphids). The host acceptance and performance of these two races was compared across 19 species of *Medicago* and *Trifolium*.

2.3 Methods

2.3.1 Plant material

In total, 19 species of plant were used from the genera *Medicago* and *Trifolium* (*M. arabica, M. orbicularis, M. littoralis, M. tornata, M. turbinata, M. laciniata, M. lupulina, M. truncatula, M. sativa, T. ambiguum, T. striatum, T. nigrescens, T. repens, T. pratense, T.ochroleucum, T. rubens , T. semipilosum, T. dubium , T. pallidum*). Details of seed sources can be found in Appendix 2. Levels of preference or performance on most of these plants species were previously unknown for any *A. pisum* race. However, *A. pisum* host races distinct from those used here have been described as specialists to *M. lupulina, T. dubium* and *T. repens* (Peccoud and Simon 2010). Seeds were sterilised by soaking them in a saturated solution of calcium hypochlorite for two minutes and then agitating the solution for one minute. This was with the exception of *M. sativa* and *T. pratense* which were soaked for three minutes and agitated for three minutes because of higher levels of fungal infection seen in these seed stocks. The larger of the seeds (*M. truncatula*, *M. turbinata*, *M. orbicularis* and *M. tornata)* had their seed coats scarified by nicking with a sterilised scalpel to help overcome dormancy.

Seeds were then plated into petri-dishes containing 1.2% plant agar with the plant hormone gibberellin added at a concentration of 0.05g/l. Seeds were kept for one week in a climate control cabinet, set at 20^0C by day and 15^0C at night, with a 16 hour day-light period. For *M. turbinata*, M. *orbicularis* and *M. tornata*, seed coats were completely removed after two days of swelling in the agar to further increase germination rates. One week after this, germinated seedlings were potted into 24 celled seed trays (each cell 50 mm x 48 mm) filled with moist soil (four parts fine sand: one part John Innes no.2 compost) and covered with a transparent lid to retain

humidity for three weeks and without a lid for a fourth week. Plants were kept in a temperature and pressure controlled greenhouse bay set at 20^0C day and 15^0C night, with supplementary lighting to give a 16 hour day length. Over this period plants were watered twice a week with distilled water. Plants were fed during watering in the fourth week after initial seed sterilisation with 40% Rorison's solution (Appendix 3 and a 20% Rorison's solution in the fifth week.

2.3.2 Aphid material

Four asexually reproducing lineages (clones) of *A. pisum* were used. For the MS aphid race these were LSR1 (International Aphid Genomics Consortium 2010) and L9Ms_052 (source SE France, supplied by JC Simon, INRA institute, Rennes). For the TP aphid race these were clones YR2 (Simon *et al.* 2011) and L7Tp_232 (source SE France, supplied by JC Simon, INRA, Rennes). Stocks of all clones used were maintained on 10 day old broad bean plants *(Vicia faba* var. The Sutton Dwarf) sealed within culture pots (Appendix 4). Aphid material for experimental use was prepared by placing a single adult aphid of the clone of interest on an individual bean plant 14 days before the experiment and removing it 24 hours later. The resulting age-controlled first instar aphid nymphs were then left to mature for 10 days. The newly formed adults to be used in the experiment were transferred to fresh bean plants to prevent the host plant becoming overstressed and left for a further four days before being used experimentally. On the day of experimentation, healthy adults were selected, taken off the bean plants and starved by being left to rest in a Petri dish for one hour.

2.3.3 EPG equipment and data capture

The EPG equipment used was a DC Giga-8 channel EPG, sourced from EPG systems© (www.epgsystems.eu). Individual plants were re-potted into 70ml pots at the start of the experiment. A plant electrode was inserted into the soil as close to the plant as possible but without the risk of touching the aphid during the experiment (Figure 2.1). There were eight plant electrodes in total (one for each plant being used), providing an electrical connection to the plant from the source of the electrical

current. Each of the plant probes had a corresponding aphid channel to which an aphid was attached via ~2 cm of gold wire using a small amount of water-based glue. The glue used was a non-toxic water-based glue (provided by EPG systems©) impregnated with silver particles to aid electrical conduction. Each of the eight aphid channels connected an aphid to the circuit so that, when the aphid stylet entered the plant, the circuit was closed and the change in resistance could be measured. All eight aphids were slowly lowered onto the plant leaf to allow some flexibility in the wire so that they could move a short distance and find their preferred feeding positions. Aphids were always placed between the petiole and underside of a leaflet of the youngest fully formed leaf, where personal observations suggested aphids in culture tended to feed.

The equipment was installed within a Faraday cage constructed from wire mesh (<1cm mesh size) soldered to a steel frame. The frame was constructed using length 120cm x height 90cm x depth 60 cm steel struts held together using solder. The EPG set up and control box with amplifier were kept in the centre of the cage and earthed to the cage. The cage front consisted of two mesh panels that opened as doors and closed during recording. Finally the cage was lined with tin foil to reduce electrical interference (Appendix 5). Inside this cage the 8 EPG channels were attached across a bar held up by two clamp stands and connected to the EPG control box (Appendix 5). Data were captured with the "EPG Stylet+d" software, avalible from the EPG Systems© website (www.epgsystems.eu/downloads). Each run of the experiment lasted six hours.

A five-week balanced block design was used to test one aphid clone at a time. Within each week, individual plant species–aphid clone combinations were run in duplicate within a single day, so that occasional recording failures did not interfere with the block design. EPGs could fail for reasons including aphids falling off or leaving the test plant, aphids breaking the gold wire, aphids causing their connecting wire to touch the plant probe, extreme changes to very high or very low output voltage during recording making fine voltage change unreadable, and occasionally interference from other electronic devices. In each week each plant species was

represented twice. At the end of each five week period MS and TP clones were alternated in the order: LSR1, TP_232, MS-052 and YR2.

Due to losses of replicates in each EPG run, most frequently due to aphids falling off plants before the six hour period was complete or insufficient plant material being available, there was some variance in the number of replicates for each aphid-plant species combination. However there was an average of 8 EPG replicates for each *A. pisum* clone-plant combination and an average of 15 replicates for each *A. pisum* race-plant combination (detailed sample sizes are given in Appendix 6).

2.2.4 Interpretation of raw EPG data

The raw data consisted of current output potential over time, which changed as the level of electrical resistance changed, giving the stereotyped patterns known as waveforms. These patterns emerged as a consequence of the aphid stylet passing though the plant with a small contribution from the muscle and neural potentials of the aphid (www.epgsystems.eu/epg/measuring-systems). Waveforms were classified into eight different forms: non-probing (np), pathway (C), potential drops (pd), repetitive potential drops (rpd), stylet within the phloem (E), xylem feeding (X) and stylet derailment (F) (Tjallingii and Gabryś 1999, Tjallingii 2006). The E phase is further divided into two distinct waveforms: E1 associated with egestion of saliva into phloem and E2 associated with phloem ingestion. Examples of each wave form are given in Appendix 1. To improve analytical efficiency, time consuming analysis of the pd waveform was omitted but was incorporated into the C waveform. Instances of pd behaviour were extremely brief and only increased C duration very slightly.

2.3.5 Processing EPG waveform data and analysis of individual waveform measurements

EPG waveforms were processed using an Excel macro, made by Sarria *et al.* (2009), which converts waveform changes over the recording period into a range of measurements in terms of: i) total duration of a behaviour, ii) time until the first

onset of a behaviour or iii) the number of separate instances of a behaviour. In total, the Sarria *et al.* (2009) Excel macro produces 119 individual waveform measurements. The five waveform measurements preselected for analysis were total duration of E1, total duration of E2, total number of probes, time to E2 >10 minutes and the total number of rpd. In the analysis of individual waveforms, all measurement replicates where a waveform did not occur were counted as zero.

Firstly, the similarity of the two clones of each *A. pisum* (MT & TP) race was tested. To do this, Spearman's rank correlation coefficient was computed between the mean of each EPG score of each clone pair using the cor.test() function in R (version 3.2.1). Secondly, each waveform was compared by calculating the Hedges' g effect size for the difference in acceptance between the MS and TP aphid races across all plants species using the R function cohen.d() found in the *effsize* package (Torchiano 2015). The significance of the difference between MS and TP aphids foreach plant species was tested using a generalised linear model (GLM) with the two pairs of clones that made up a race nested within the race identifier. This was done using the glm() function in R. The distribution family used of each waveform measurement were as follows: tweedie for E1 and E2, gamma for rpd and probing data and Gaussian for time to E2>10. In addition, to reduce heteroscedasticity in E2, rpd and probing data was transformed before analysis by ranking E2 data and taking the square root of rpd and probing data.

2.3.6 Interpretation of waveform measurements with multidimensional analysis

One objective of this study was to explore the use of multidimensional summary values derived from EPG data that can be reliably used as a measure of aphid acceptance. For this, principal component analysis (PCA) and linear discriminant analysis (LDA) were performed using a cleaned set of EPG waveform measurements.

Prior to the analysis, waveform measurements were removed that were not suitable for analysis because they had incomplete or strongly correlated data sets. Firstly, waveforms that had 50% or more values as zero or as a missing value were removed.

The remaining 90 waveform measurements were compared using a correlation matrix, generated with the R function cor() from the MASS library (Venables & Ripley 2002), to detect any highly correlated variables that might inflate the importance of two measurements that are essentially measuring the same attribute (e.g. measurements of total time spent probing and total time spent not probing). One of each pair of waveform measurements that had a positive correlation greater than 0.80 or negative correlation less than -0.80 was removed from the data set. This gave a final total of 60 different measurements of EPG waveform measurements (Appendix 7).

To observe how well the data fitted to patterns of aphid acceptance, all cleaned waveform measurements were analysed with PCA using the base R function princomp(). This was done to examine the overall pattern in aphid behaviour and identify which waveform measurements contributed most to differences between TP and MS aphids.

In addition LDA was run separately on each aphid race, treating plant species as the classes to be distinguished. This analysis was carried out using the R function LDA() from the *MASS* library (Venables & Ripley 2002). The patterns between MS and TP aphid first (LD1) and second (LD2) LDA axis were compared to first (PC1) and second (PC2) components of the PCA. The significance of the difference between the MS and TP aphids for each plant species was calculated using a GLM with a Gaussian distribution and with clone nested within in race. The PC1 scores were also transformed before analysis to reduced heteroscedasticity by taking the rank of scores. Then an LDA profile score of acceptance discrimination between the clones (LDA profile) was created by subtracting MS aphid LDA scores from TP aphids LDA scores. This meant that that plants accepted by MS aphids and rejected by TP aphids had positive LDA profile scores, plants accepted by TP aphids and rejected by MS aphid had negative LDA profile scores and plants accepted or rejected by both races had LDA profile scores close to zero. Finally, as with the individual waveform measurements, the difference between the clones in both their PC1 and LD1 scores were correlated using Spearman's rank correlation coefficient.

2.3.7 Comparison of EPG acceptance profiles with aphid performance

To test whether acceptance and performance were strongly correlated, the performance of MS and TP races on 16 species of *Medicago* and *Trifolium* was measured. Of the 19 species listed above *T. ochroleucum, T. pallidum* and *T. semipilosum* were not used in this part of the study. This was due to poor germination leading to insufficient plant material in these species. Performance of aphid races, measured as fecundity and adult quality after seven days of feeding on plants, was compared to EPG profiles. The plants and aphids were grown in identical conditions to the EPG experiment.

Five-week-old plants of each species, growing in a culture pot, were inoculated with a single 14 day-old adult aphid. Once inoculated plants were placed in the climate control room in which the aphid stocks were raised (at 20^0C with a 16 hour daylight cycle). At exactly seven days after inoculation culture pots were opened and the numbers of living offspring were recorded as a measure of fecundity. At this stage the offspring were at most in their third instar and could easily be distinguished from the adult inoculum.

The quality of the adult aphids was recorded by visual inspection using the following scoring system: 0) *dead*: aphid does not move, 1) *poor*: adult aphid pale in colour with none of the original red or green pigment and extremely undernourished, 2) *good*: aphid has some of its original red or green colour and only a slight reduction of abdomen size (abdomen less elliptical in shape), 3) *very good*: aphid has barely changed in condition with full colour and a rounded abdomen. For both performance measures there were three to eight replicates for each aphid clone-plant combination and six to 16 replicates for each aphid race-plant combination.

The mean values of the five pre-selected individual waveform measurements (total duration of E1, total duration of E2, total number of probes, time to E2>10 min and

the number of rpd) and PC1, PC2 and LD1 acceptance scores for both MS and TP aphids were compared to the mean values of the two performance measures using the Spearman's rank correlation coefficient in the cor.test() function in base R.

2.4 Results and discussion

19 species of plant had their acceptance profiled for both MS aphids and TP aphids using EPG waveform measurements E1, E2, time until E2, rpd and number of probes. In addition, the first two axes of a PCA and a LDA of EPG data were taken as multivariate profiles of overall acceptance. Trends between individual waveform measurements in the acceptance by the divergent *A. pisum* races were then studied to see which of them represent consistent measures aphid acceptance or rejection. Multivariate acceptance profiles of EPG data were then compared to see if they maintained the same patterns of divergent acceptance by races. Finally all EPG profiles were compared to performance measures to see if acceptance traits closely correlate with performance across multiple plant species.

2.4.1 Correlation between clones within races

For each of the five individual waveform measurements (duration of E1; duration of E2, number of rpd, number of probes, time to E2<10 min) the two clones that make up each race were compared to check if EPG scores were consistent within races. For each waveform measurement there was a moderate (rho> 0.4) to very strong (rho>0.8) significant correlation between clones of the MS races (LSR1 and L9Ms_052) (Figure 2.3) and TP races (YR2and L7Tp_232) (Figure 2.4). Given this pattern of correlation, data from the two clone pairs was combined in order to focus analyses on differences in behaviour between the MS and TP races.

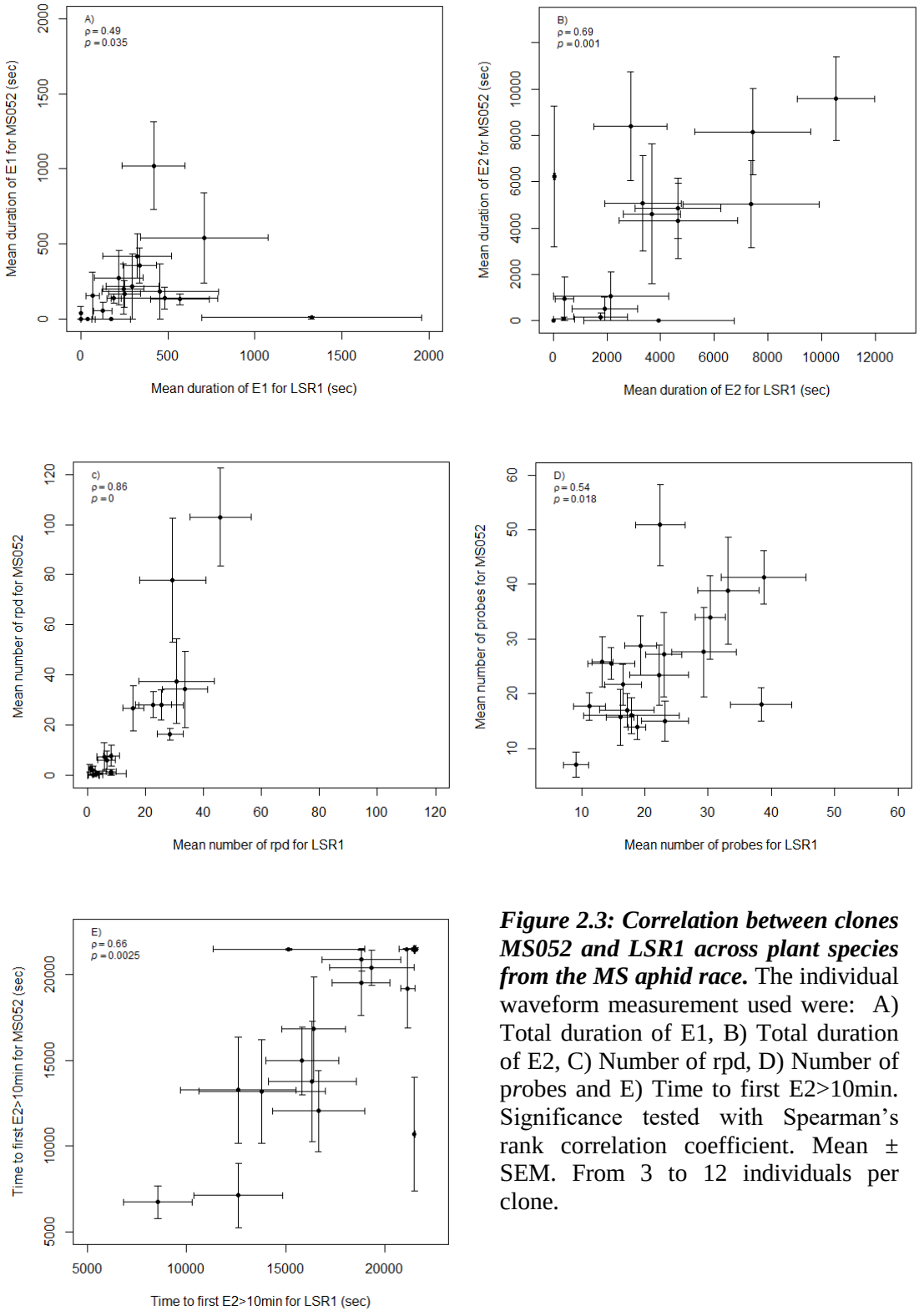


Figure 2.3: Correlation between clones MS052 and LSR1 across plant species from the MS aphid race. The individual waveform measurement used were: A) Total duration of E1, B) Total duration of E2, C) Number of rpd, D) Number of probes and E) Time to first E2>10min. Significance tested with Spearman's rank correlation coefficient. Mean ± SEM. From 3 to 12 individuals per clone.

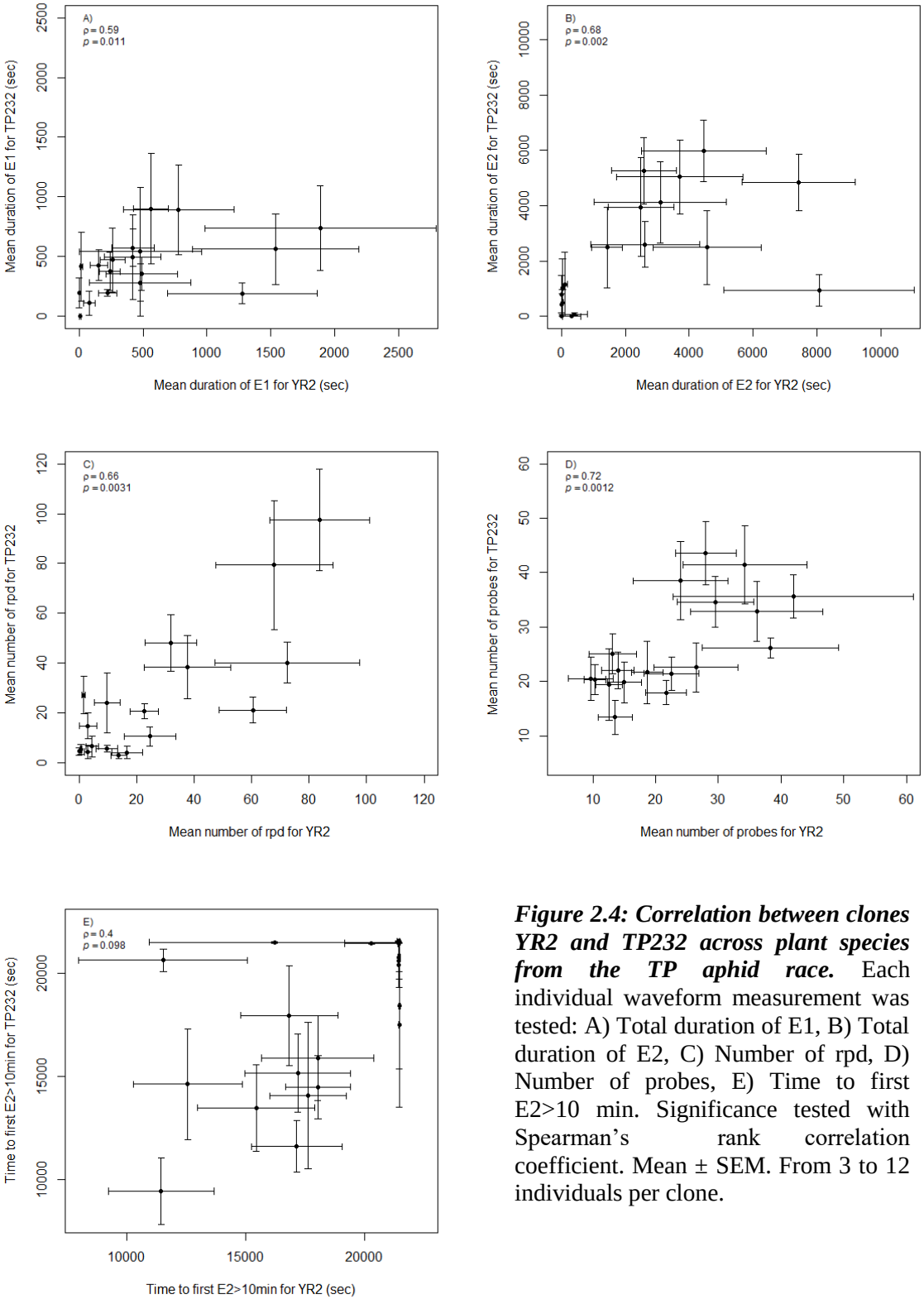


Figure 2.4: Correlation between clones YR2 and TP232 across plant species from the TP aphid race. Each individual waveform measurement was tested: A) Total duration of E1, B) Total duration of E2, C) Number of rpd, D) Number of probes, E) Time to first E2>10 min. Significance tested with Spearman's rank correlation coefficient. Mean ± SEM. From 3 to 12 individuals per clone.

2.4.2 Analysis of individual EPG waveforms

Direct comparison of aphid EPG *M. sativa* and *T. pratense* show significance differences in nearly all EPG waveform measurements (Table 2.2), with the notable exception of total duration of E1 on *M. sativa* plants and time to E1 >10 min on M sativa plant. However caution is needed for the interpretation rpd waveform on *M. sativa* and Probe waveform on *T. pratense*, as there was also a significant difference between clones within a race suggesting there is some within clone variation (Table 2.2. there was also marginally non-significant difference between clone for time to E2 >10 min on *T. pratense.* On the whole these results are in line with expectations that aphid on their adapted host would express more phloem associated behaviours and less probing and searching behaviours than on the non-host plant (Figure 2.6). This suggest EPG can be used effectively to distinguish aphid host preferences, though the importance of an individual waveforms is plant and to some degree aphid specific.

A strong significant difference in acceptance between *M. sativa* and *T. pratense* by both MS and TP aphids was also seen for PC1 and LD1 scores (Figure 2.9 and Table 2.2). This suggests that both these scores are good measures of overall host acceptance. However, both PC2 and LD2 scores did not consistently discriminate between plants species (Table 2.2) suggesting these values may be less effective at predicting the race acceptance. Again there some significant and marginally significant within race variance in the multivariate scores on *T. pratense.*

Comparison of the five waveforms E1, E2, time until E2, rpd and probing across all species showed both negative and positive co-linear relationships between waveforms (Figure 2.5), although this co-linearity was much weaker for the number of probes compared to the other the waveforms. This suggests that probing was driven by different mechanisms other than waveform measurements. Of these waveform measurements, total duration of E2 and number of rpd appeared to represent positive acceptance behaviours towards plants, while waveforms time to E2>10min and number of probes represented negative acceptance behaviours (Figures 2.6 and 2.7B:E). For instance, MS aphids compared to TP aphids on *M. littoralis*, *M. sativa* and *M. orbicularis* plants had longer durations of E2, a higher number of rpd, a lower number of probes and shorter durations to E2 >10 (Figure 2.7

and 2.7 B:E). These characteristics are associated with host plant acceptance hence these three *Medicago* plant species are subsequently referred to as *MS accepted* plants. In contrast, TP aphids compared to MS aphids, on *T. pratense, T. nigrescens* and *T. repens* had longer durations of E2, higher number of rpd, lower number of probes and shorter durations to E2 >10 (Figures 2.6 and 2.7); these three plant species are subsequently referred to as *TP accepted* plants. These observations are consistent with results seen in other studies that compared EPG measures from divergent *A. pisum* clones on *M. sativa* and *T. pratense* (Caillaud and Via 2000, Schwarzkopf *et al.* 2013).

Table 2.2: Difference between MS aphid and TP aphids EPG score. Significance tested with glm with clone identifier nested within race.

		M. sativa			***T. pratense***		
Measure	Difference tested	F	DF	*P*-value	F	DF	*P*-value
Individual EPG waveform							
E1	plant	0.27	1,54	0.604	**10.74**	**1,54**	**0.002**
	clone pair	0.03	2,52	0.968	0.58	2,52	0.567
E2	plant	***20.34***	***1,54***	***<0.001***	**9.12**	**1,41**	**0.005**
	clone pair	0.15	2,52	0.865	1.20	2,39	0.313
Rpd	plant	***11.20***	***1,54***	***0.002***	**27.64**	**1,54**	***<0.001***
	clone pair	***11.20***	***2,52***	***0. 006***	1.92	2,52	0.159
Probe	plant	***11.90***	***1,54***	***0.001***	**8.78**	**1,54**	**0.005**
	clone pair	2.93	2,52	0.062	**6.88**	**2,52**	**0.003**
Tim to	plant	***16.16***	***1,54***	***<0.001***	3.99	1,54	0.052
>E2 10min	clone pair	0.251	2,52	0.778	1.34	2,52	0.275
Multivariate score							
PC1	plant	**12.67**	***1,54***	***<0.001***	**11.30**	**1,54**	**0.002**
	clone pair	0.59	2,52	0.556	**4.09**	**2,52**	**0.024**
PC2	plant	1.27	*1,54*	0.261	**13.33**	**1,54**	***<0.001***
	clone pair	2.56	2,52	0.087	3.11	2,52	0.056
LD1	plant	**102.06**	***1,54***	***<0.001***	**48.41**	**1,54**	***<0.001***
	clone pair	2.32	2,52	0.108	**3.30**	**2,52**	**0.047**
LD2	plant	**16.27**	***1,54***	***<0.001***	2.18	1,54	0.147
	clone pair	0.11	2,52	0.892	3.22	2,52	0.051

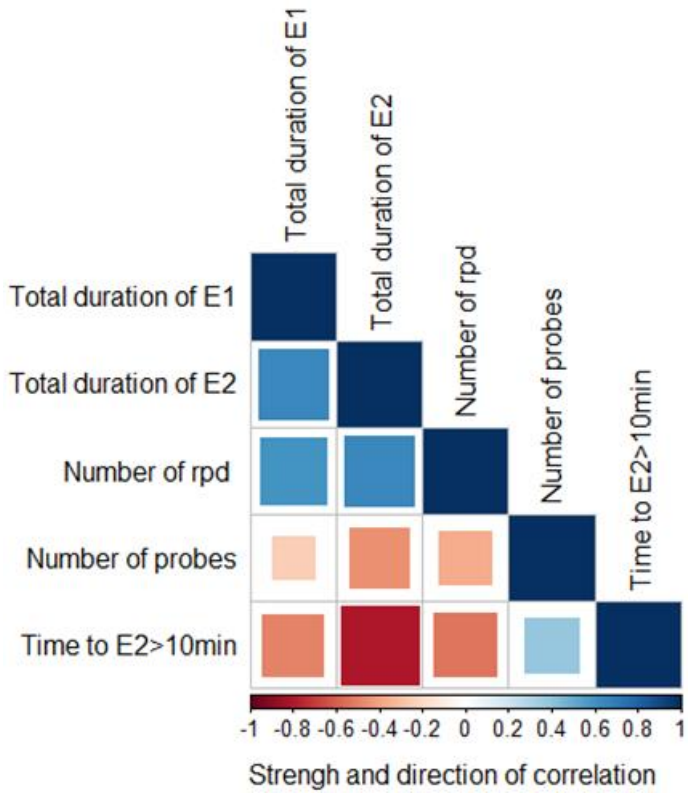


Figure 2.5: Correlation matrix comparing the different waveform measurements. Size of squares represents the strength of correlation between waveforms and colour the direction of the correlation.

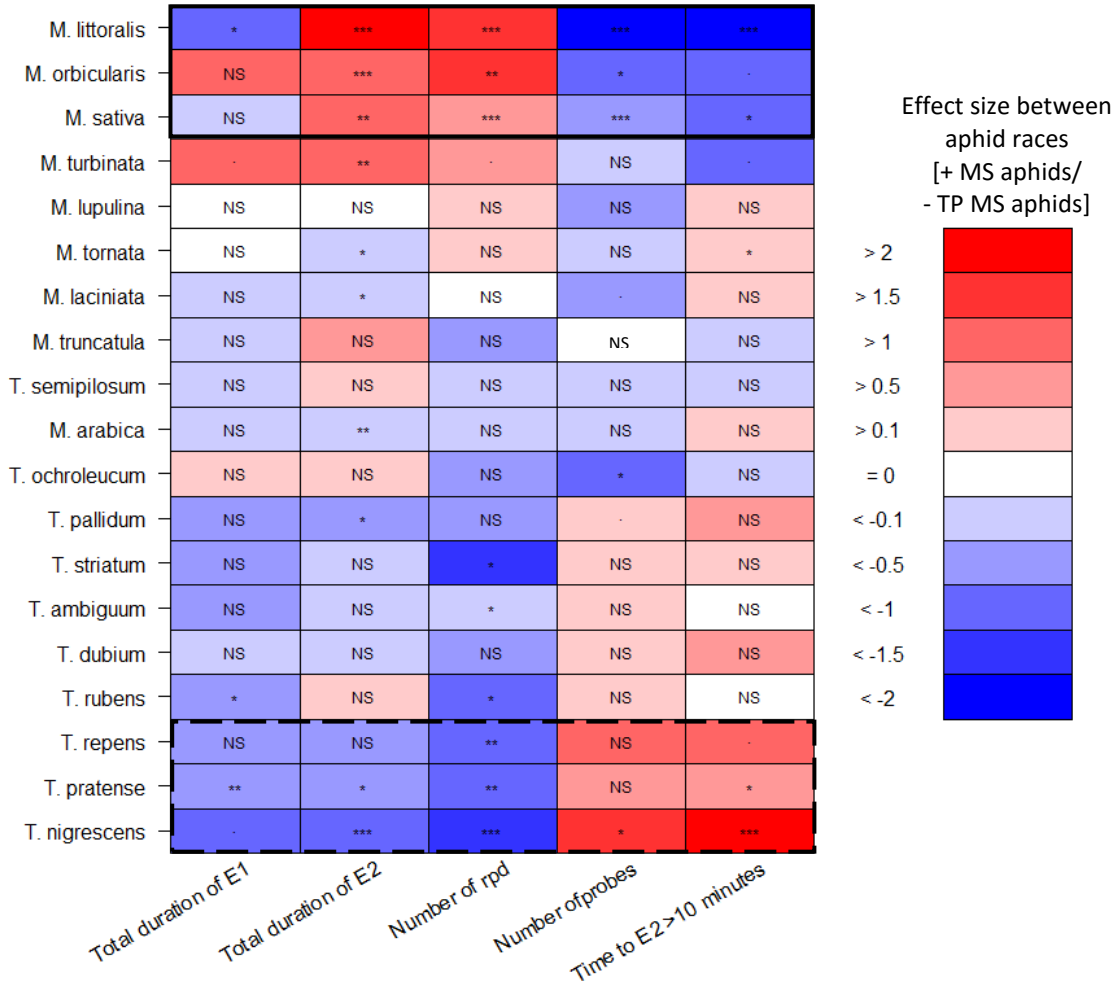


Figure 2.6: Effect size difference between MS and TP aphids on each plant species for five waveforms. Positive values (red) indicate a larger score for MS aphids and negative values (blue) indicate larger score for TP aphids. ·P<0.10, * $p < 0.05$, ** $p <$ 0.01, *** $p < 0.001$. Significance between aphid race scores was tested for plant species using GLM, with nested clone identifiers. Species within solid box represent the MS accepted plants and species within dashed box represent TP accepted plants.

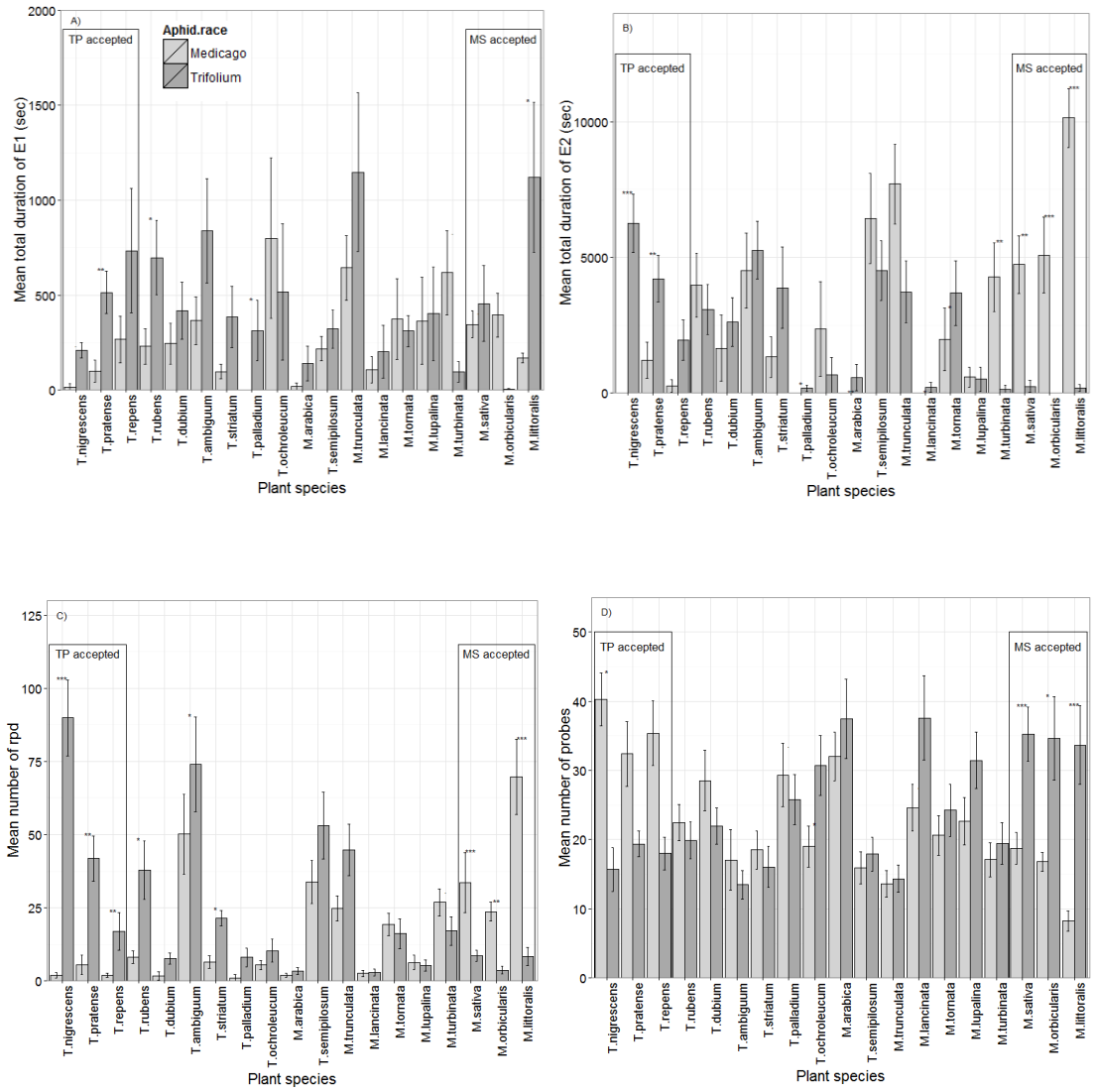


Figure 2.7: Mean individual EPG waveform measurements for each plant species for MS and TP aphid races. Measurements shown are A) total duration of E1, B) total duration of E2, C) number of rpd, D) number of probes , E) time from the first probe to the first sustained E2. Plant species within the solid rectangles are those that show the most consistent difference between clones, i.e. the MS accepted plants and TP accepted plants. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Significance tested using generalised linear model with clone identifiers nested within the race. Mean ± SEM.

In contrast to the other four individual EPG waveform measurements the duration of E1 waveform measurement had an inconsistent pattern on MS accepted plants (Figure 2.6 and 2.7A). This suggests the duration of E1 is a relatively poor measure of acceptance by these two aphid races. This result diverges from previous findings by Caillaud and Via (2000) and Schwarzkopf *et al.* (2013). It could be that when comparing results across many species, aphid salivation during the E1 stage is a less important factor in the process of acceptance or that its duration is unrelated to aphid ability to accept plants. It is likely that the mixed pattern in E1 durations could be explained by the fact that E1 can occur without E2, but E2 cannot occur without E1 preceding it. Therefore, a poorly accepted plant, with no E2, could have E1 durations that are comparably long, while a highly accepted plant, with a small number of very long instances of E2, would have only a few short E1 periods. The lack of a clear pattern in E1 durations in this study presents a noteworthy issue for future research, since it shows that preselecting any one waveform for analysis in isolation, has the potential to bias results and their interpretation.

Other more complex patterns of variation can be observed among the aphid waveform measurements. For instance, *M. truncatula, T. ambiguum, T. semipilosum* show high levels of acceptance for both aphid races across waveforms, while *M. arabica, M. lupulina, M. laciniata* exhibit low levels of acceptance for both aphid races (Figure 2.6). In addition, there are some inconstancies in the difference between waveform measurements in the divergence between MS and TP aphid scores across plants. For example, the E2 durations, associated with aphid phloem feeding, are more divergent in the MS accepted plants than the TP accepted plants (Figure 2.6). Conversely, divergence between aphid races in the number of rpd, associated with a pre-phloem stylet interaction with plant cells, is greater in TP accepted plants than MS accepted plants. Mechanistically, this suggests that on MS accepted plants MS aphid acceptance and TP aphid rejection decisions tend to occur at a later stage, when the stylet is actually within the phloem. However, on TP accepted plants the decision to accept seems to occur earlier, just before the aphid has fully entered the phloem when the rpd waveform is recorded. These differences in waveform measurements suggest individual waveforms need to be interpreted in the context of other waveforms, if they are to be used as measures of acceptance.

EPG studies that use only a limited number of EPG waveforms risk presenting a misleading picture of host acceptance.

Previous studies have placed particular importance on the interpretation of individual waveform measurements as a way of describing divergent acceptance between *A. pisum* races on different host plants (Wilkinson and Douglas 1998, Calliud and Via 2000, Schwarzkopf *et al.* 2013, Gao *et al.* 2008). However, none of these studies have used as large a range of plant species as this study. Using so many plant species allows us to observe waveform variation across a wider range of aphid-plant combinations. This diversity gives the potential for a more extensive comparative analysis of the role of various waveforms and improves our ability to compare waveform data to other data types. The small discrepancies between waveform measurements, as well as the notable difference seen in E1 data, suggest caution is needed when comparing results between studies that use different waveform measurements. However, taken together, most of the EPG waveforms used here were consistent measures of acceptance and rejection, despite the fact they record behaviours that occur at different sites within the leaf tissues.

2.3.3 Multivariate analysis of acceptance using EPG data.

From a methodological perspective, the *a priori* selection of a limited number of waveforms of inferred mechanistic origin in EPG analysis runs the risk of introducing bias into the results. This is because each individual waveform represents just a small part of the aphid acceptance process. This may be practically important if EPG is being used to inform untargeted profiling studies, such as genome or metabolomic profiling, since biased and incomplete waveform analysis could markedly affect end results. A multivariate approach that encompasses as much of the waveform information as possible could resolve this issue. This multivariate profile of acceptance would consist of one or two orthogonal variables that explain a high proportion of the variance. Ideally these multivariate axes would be driven by measurements with meaningful biological interpretations. In this study, multivariate analyses of acceptance using EPG were conducted using both PCA and

LDA. For the PCA, PC1 and PC2 explained the majority of the variance (60.7%) (Appendix 8). Of these, PC1 accounted for much more variance in the data (46.4%) than PC2 (14.3%).

Inspection of PC1 and PC2 shows there are two clusters within the data (Figure 2.8). The first cluster, on the left of the PCA plot and with more negative PC1 scores, consists of a tight cluster of points exhibiting similar EPG waveform profiles (Figure 2.8; Solid ellipse). The second cluster, with more positive PC1 scores, consists of a less tightly clustered spread of points (Figure 2.8; Dotted ellipse). Driving PC1 in the positive direction are waveform measurements associated with acceptance, such as total durations of E and time the aphid stylets spend within a plant (Appendix 9). Conversely, waveform measurements driving PC1 in the negative direction are associated with host plant rejection, such as the time to E2>10min which measures time the aphid spend searching before it establishes a successful feeding event (Table 2.3.

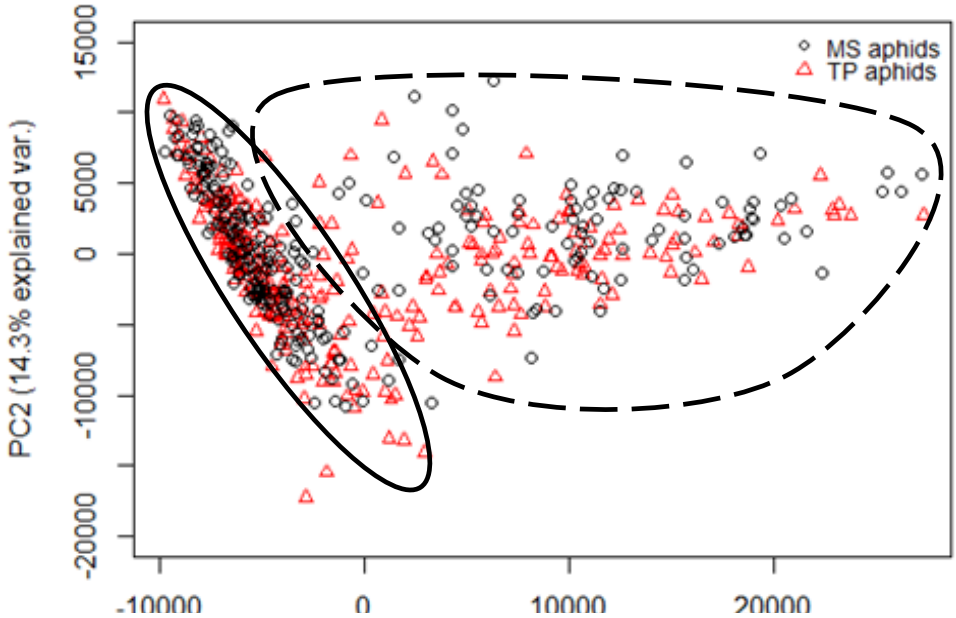


Figure 2.8: PCA plot of cleaned EPG waveforms. Black circles represent MS aphids and red triangles represent TP aphids. The black solid line ellipse encloses a group that exhibit behaviours linked to low acceptance (i.e. aphids perform little or no positive behaviours, such as feeding). The broken line ellipse encloses aphids that exhibit some level of acceptance (i.e. aphids perform positive behaviours, such as feeding).

Table2.3: Top 5 and bottom 5 PC1 and PC2 loading scores

EPG measurement	loadings	Interpretation
PC1		
Total duration of E	0.455	Duration in phloem across all hours
Total probing time	0.411	Duration inside the plant across all hours
Duration of 1st E	0.243	Duration of 1st phloem probe
Time from the beginning of E2 to the end of the EPG	0.195	Time searching that contains feeding behaviour
Duration of 2nd probe	0.048	Duration of the second time entering a plant
...		
Total duration of np during the 4th hour	-0.049	Time spend outside of the plant in 4th hour of recoding
Total duration of np during the 6th hour	-0.050	Time spend outside of the plant in 6th hour of recoding
Total duration of np during the 3rd hour	-0.057	Time spend outside of the plant in 3rd hour of recoding
Total duration of C	-0.073	Time spent in the apoplast
Time to from start of EPG to first E2 >10 minutes	-0.709	Time spend searching the plant until a successful feeding is reached
PC2		
Total probing time	0.709	Tota l time inside the plant
Total duration of C	0.353	Total time inside the apoplast
Total duration of F	0.334	Total time in F waveform
Time to from start of EPG 1st sustained E2 > 10 minutes	0.187	Time to the first sustain E2 >10min
Duration of 2nd probe	0.119	Duration of the 2nd probe and aphid make into a plant
...		
Total duration of np during the 2nd hour	-0.133	Total duration stylus outside of the plant between in the 2nd hour
Time from the beginning of E2 to the end of the EPG record Z	-0.145	Duration from first E2 tot eh end of the EPG?
Total duration of np during the 1st hour	-0.154	Total duration stylus outside of the plant between 1st and 2nd hour
Total duration of E	-0.186	Duration in phloem across all hours
Duration of 1st E	-0.241	Duration of 1st phloem penetration

Clearly lower PC1 scores indicate a lower level of host plant acceptance and the tight cluster of aphid-plant combinations with low scores are exhibiting rejection-like behaviours. To support this proposition, PC1 scores closely match the trends observed in the individual waveform measurements, with a clear difference between aphid races' acceptance of MS accepted and TP accepted plants and a continuum of variation across the other plant species (Figure 2.9 and 2.10A).

In contrast, PC2 describes less separation between the two clusters (see figure 5). PC2 also showed a much less clear pattern of difference between plants, in terms of MS aphid and TP aphid acceptance (see figure 9 and 10B). It would appear the PC2 scores are a poor measure of aphid acceptance and this claim is supported by the relatively low level of variance PC2 explains (14.3%). In addition, the waveforms driving PC2 in the positive direction, such as time in the plants, and in the negative direction, such as duration of the first penetration, are harder to interpret in terms of aphid acceptance (Appendix 9).

To improve on PCA scores, which explain the maximum variance across all the data set, an LDA was performed to find the maximum differences between the plant species. By doing this the LDA controlled for the less informative within plant species variance. As the LDA was performed on each aphid race separately this created two profiles of acceptance. Observation of the LDA showed that much more between plant variance was explained by LD1 (MS aphids 24.5%, TP aphid 22.2%) than LD2 (Ms aphids 9.9%, TP aphids 12.6%) with other LD axes becoming progressively less informative (appendix 10).

Overall mean LD1 scores of the LDA strongly correlated with PC1 scores for both MS (r=0.85) and TP aphids (r=0.1), suggesting that these scores are driven by similar features of aphid acceptance (Figure 2.11A). In addition LD1 was strongly influenced by phloem associated behaviours in the positive direction and a mix of non-phloem behaviours (F and G) in the negative. The occurrence of interpretable pattern further supports LD1 scores as a measure of aphid acceptance (Appendix 11).

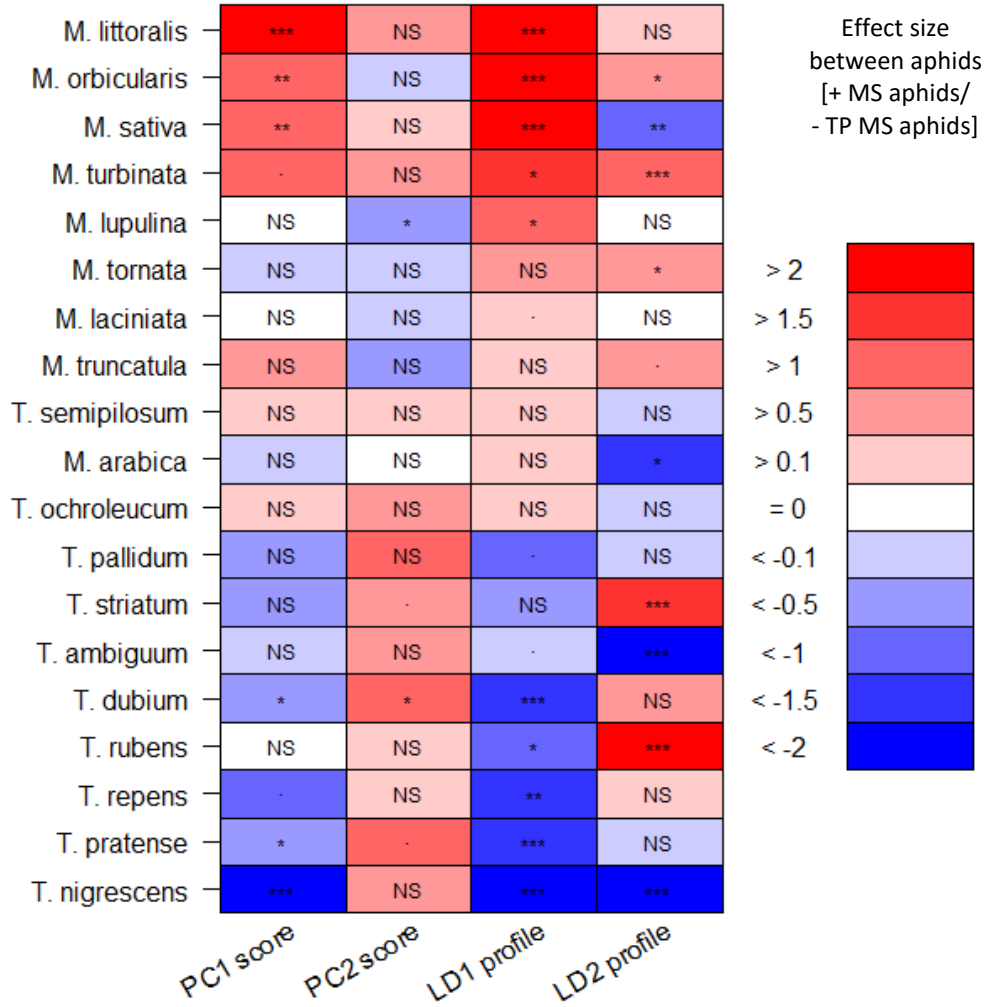


Figure 2.9: Effect size for the difference between the multivariate profile scores of MS and TP aphids. LD1 and LD2 profiles are calculated by taking the difference between score of an LDA performed on MS aphids and an LDA performed on TP aphids. Positive values indicate a larger score for MS aphids and negative values indicate a larger score for TP aphids.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Significance of the difference between aphid races was tested for each plant species using generalised linear model with clone identifiers nested within the race.

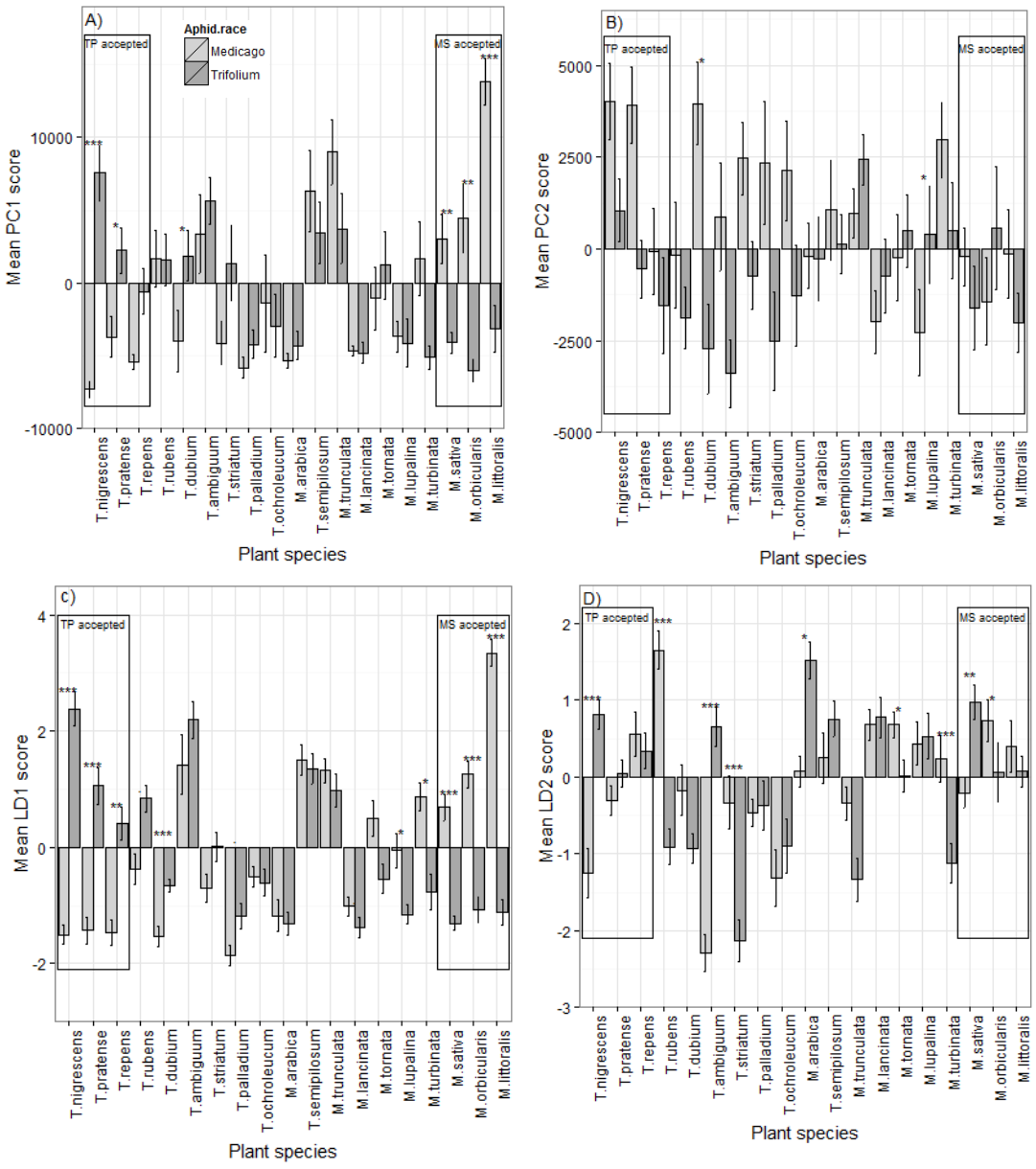


Figure 2.10: Multivariate scores based on EPG waveforms for each plant species for MS and TP aphid races. Mean score ± SEM for A) PC1, B) PC2, C) LD1 and D) LD2against plant species. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Plant species within the solid rectangles are those that show the most consistent difference between clones, i.e. the MS accepted and TP accepted plants. Significance tested using generalised linear model with clone identifiers nested within the race. ± SEM.

In contrast, there was no significant relationship between LD2 and PC2 scores, this implies that these two scores are unrelated (Figure 2.11B). The LD2 loadings appear to be influenced in both directions by waveform measurements associated with F, indicating LD2 is likely to be influenced strongly by noise in the data (Appendix 11).

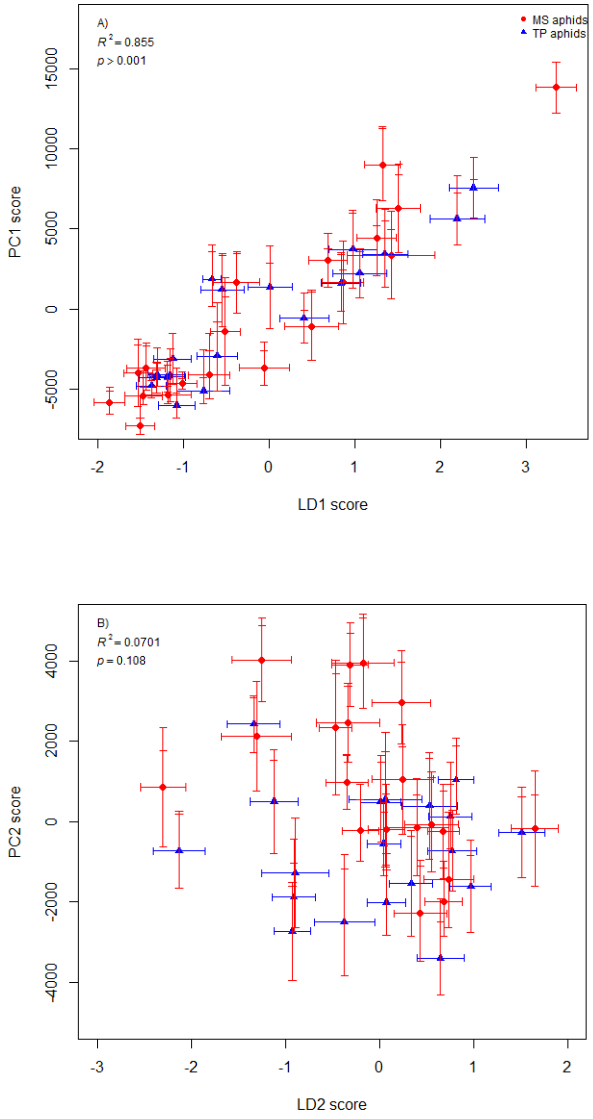


Figure 2.11: Correlation of multivariate PCA against LDA EPG scores for each aphid clones. There is a highly significant relationship between A) LD1 scores and PC1 scores (t = 14.5, df = 36, P>0.001) but no such correlation between B) LD2 scores and PC2 scores (t = -1.6, df = 36, p = 0.108). Mean ± SEM. Significance tested with Pearson's correlation coefficient.

The comparison between the LD1 scores of the two aphid races exhibits an interesting pattern in acceptance (Figure 2.12). One group of plant species exhibit a continuum of acceptance between high MS and low TP accepted plants, through to plants with low MS and high TP acceptance (figure 12A: blue circles), then a set of plants that both aphid races either seem to accept (Figure 2.12A: red triangles), or reject (Figure 2.12A: black squares). This reflects similar patterns seen in the individual waveform measurements. Combining MS and TP scores into an *LD1 profile* shows a smooth continuum of acceptance similar to that of PC1 scores (Figure 2.9). The same comparison between MS aphid and TP aphid LD2 with PC2 scores reveals no obvious pattern (Figure 2.12B) and LD2 profiles do not match the PC1 scores or LD1 profiles (Figure 2.9), or the other individual waveform measurements (Figure 2.6).

The advantage of using LD1 score over PCA is that uninformative within plant species variance in the LD1 profile is reduced to give a better separation between the aphid-plant species combinations. This may be useful in profiling studies that would benefit from acceptance profiles with uninformative variance controlled for. Finally, as with the individual waveform measurements, the between-clone PC1 and LD1 scores were significantly correlated (PC1: MS aphids rho = 0.64, PC1:TP aphids rho = 0.61, LD1: MS aphids rho = 0.90, TP aphids rho = 78) (Appendix 12).

2.3.4 Comparison of EPG profiles with fecundity and adult quality results.

A number of authors have inferred the relationship between aphid acceptance and performance (e.g. Del Campo *et al.* 2003, Ferrari *et al.* 2008, Schwarzkopf 2013). This study has shown quantitatively that there is a consistently strong correlation between most EPG measures of acceptance and both aphid fecundity (Figure 2.13, Table 2.4) and adult quality (figure 2.14, Table 2.5) across 16 plants species in two genera. For duration E2 and number of rpd, there was a positive relationship with performance, while for time to E2>10 min and number of probes there was a negative relationship. It seems behaviour patterns related to feeding are linked to higher performance, while time spent on behaviours associated with plant

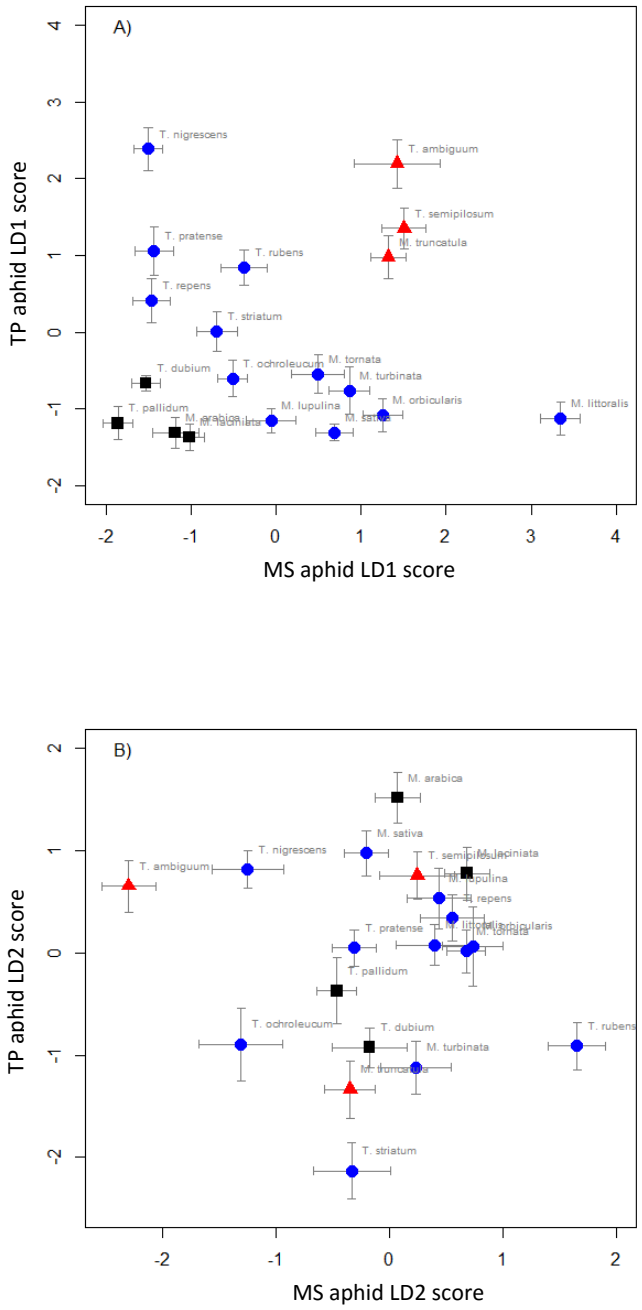


Figure 2.12: Relationship between MS and TP aphid LDA scores. A) LD1 scores exhibit a clear pattern of plant acceptance with a continuum of acceptance between i) high MS and low TP accepted plants and low MS and high TP accepted plants (blue circles), ii) plants both aphid races accept (red triangles) and iii) plants both aphid races reject (black squares). Conversely B) LD2 exhibits no obvious patterns and bears no relationship to the LD1 scores (as shown with the same blue circles, red triangles and black squares). Mean. ± SEM.

assessment and searching for the phloem, is linked to reduced fitness. This further supports the conclusion that most EPG measurements closely represent divergent acceptance by aphid races. E1 was again the exception as it had a a weak correlation to the fecundity and adult quality performance measures (Figure 2.13-2.14, Table 2.4-2.5). The lack of conformity of total duration of E1 with the other EPG waveform measurements supports the claim that it is a poor measure of acceptance.

Comparisons of correlations between EPG scores and either fecundity (Table 2.4) or adult quality (Table 2.5) showed limited differences between PC1 and LD1 multivariate scores and individual EPG waveforms. PC2 and LD1, on the other hand, had very poor correlation, suggesting again it is a poor measure of acceptance (Table 2.3). This result reinforces the validity of citing these multivariate scores as reliable holistic measures of acceptance complimented by information from selected waveforms to provide insight into the role of individual acceptance mechanisms.

The close link between acceptance and performance on the characteristic host species for MT and TP aphids is not surprising. Firstly, acceptance is likely to influence performance and aphids incur a cost from their unwillingness to feed promptly on a host presented to them (Bernays and Funk 1998, Powell *et al.* 2006). Secondly, for phytophagus insects it is expected that traits for performance and acceptance need to co-evolve if either trait is to remain advantageous (Matsubayashi *et al.* 2010). The close correlation between acceptance and performance across multiple host and non-host plant species seen in this study suggests that the features of the plant that influence acceptance and performance must also be co-evolved. A close relationship between *A. pisum* traits for host acceptance and traits for performance on known host plants was concluded by Hawthorne and Via (2001) to be due to pleiotropic alleles or strong genetic linkage. That there is a strong link between acceptance and performance, even on plants not currently recognised as *A. pisum* host plants in the wild, indicates that traits are directly associated with a continuous feature of plant phenotypes.

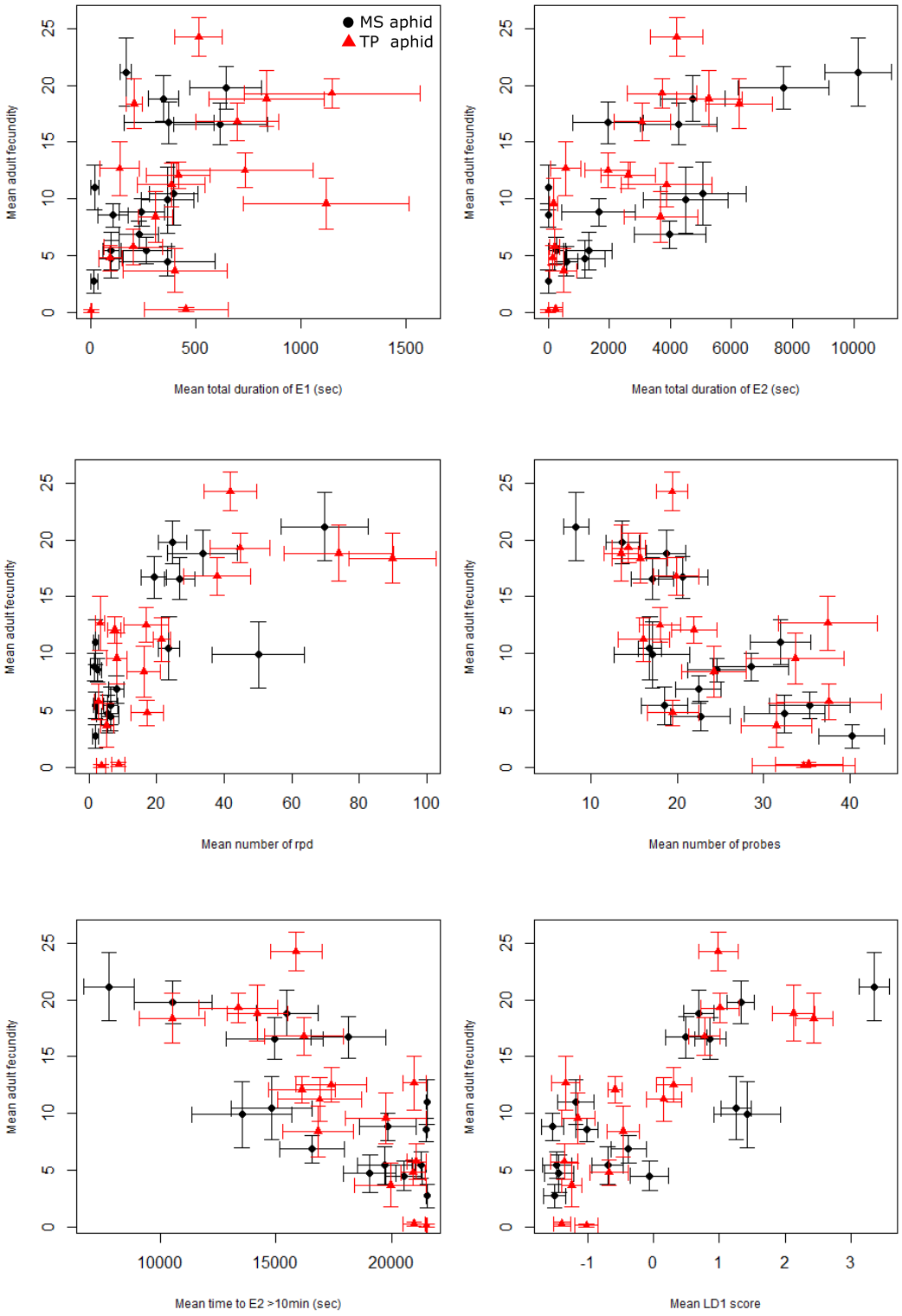


Figure 2.13: Correlation of individual waveform measurements and LD1 scores against mean aphid fecundity. ± SEM. Significance is given in Table 2.4.

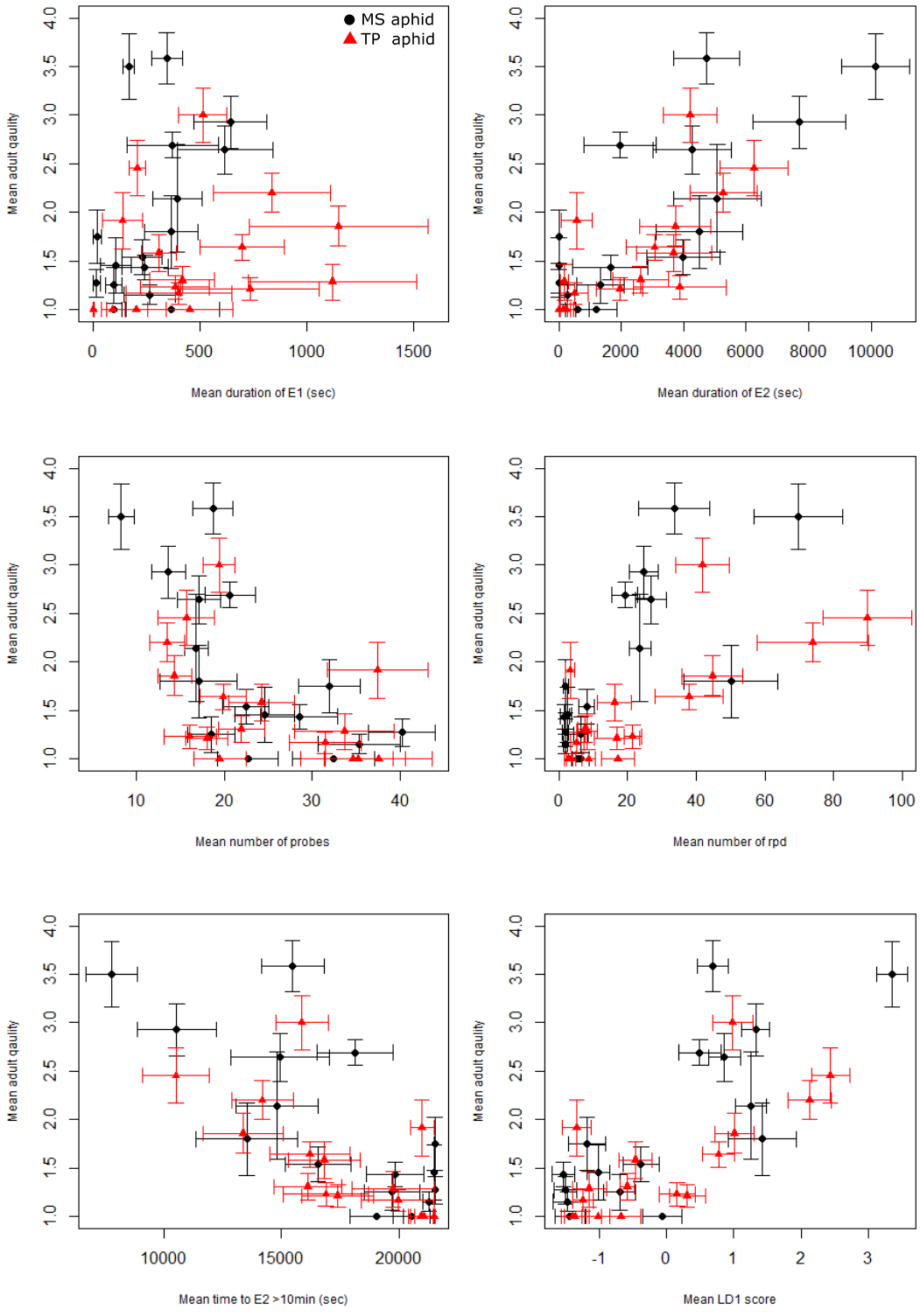


Figure 2.14: Correlation of individual waveform measurements and LD1 scores correlated against mean aphid quality scores. ± SEM. Significance in given in Table 2.5.

Table 2.4: Correlation between all EPG measurements and fecundity. Values in bold are significant (>0.05). Significance tested with Spearman's rank correlation coefficient.df = 14

Aphid Races	Total duration of E1	Total duration of E2	Number of rpd	Number of probes	Time to E2>10 min	PC1	PC2	LD1 profile
MS aphids								
ρ	0.51	0.71	0.65	-0.69	-0.68	0.68	-0.21	0.68
P-value	**0.048**	**0.002**	**0.007**	**0.004**	**0.005**	**0.005**	0.443	**0.005**
TP aphids								
ρ	0.53	0.80	0.69	-0.67	-0.80	0.82	-0.09	0.78
P-value	**0.038**	**<0.001**	**0.004**	**0.006**	**<0.001**	**<0.001**	0.755	**0.001**

Table 2.5: Correlation between all EPG measurements and adult quality. Values in bold are significant (>0.05). Significance tested with Spearman's rank correlation coefficient, df =14.

Aphid Races	Total duration of E1	Total duration of E2	Number of rpd	Number of probes	Time to E2>10 min	PC1	PC2	LD1 profile
MS aphids								
ρ	0.48	0.74	0.74	-0.71	-0.69	0.69	-0.21	0.73
P	0.057	**0.001**	**0.001**	**0.002**	**0.003**	**0.005**	0.458	**0.001**
TP aphids								
ρ	0.37	0.82	0.61	-0.53	-0.78	0.80	-0.01	0.71
P	0.160	**<0.001**	**0.013**	**0.035**	**<0.001**	**<0.001**	0.957	**0.002**

Intriguingly the plants with the most divergent aphids acceptance and performance results, *T. nigrescens* for TP aphids and *M. littoralis* for MS aphids, were not the host plants from which the aphid races are known from in the wild. Interestingly, there is little evidence aphid acceptance reflects the phylogenetic relationships of the plants concerned. *T. nigrescens* (Sect. Trifoliastrum) is not closely related to *T. pratense* (Sect. Trifolium) (Ellison *et al.* 2006, Visnevschi-Necrasov *et al.* 2013), while *M. littoralis* (Sect. Spriocarpos) is not closely related to *M. sativa* (Sect. Medicago) (Maureira-Butler *et al.* 2008). Given this weak phylogenetic relationship between the plants with similar aphid acceptance it appears the plant-aphid interactions may be understood best in terms of plant ecology and chemistry. In aphids where host acceptance takes place on or below the plant epidermis, the obvious candidates for the drivers of acceptance and performance are the chemical properties of the host plant (Pickett *et al.* 1992, Del Campo *et al.* 2003, Powell *et al.* 2006). For this reason it is important that the chemical ecology of host acceptance is investigated fully.

The comparative approach adopted here also sheds some interesting light upon the evolution of host plant specificity. A clear separation of aphid races in terms of acceptance and performance on TP and MS accepted plants might be interpreted as providing a clear barrier to co-occurrence. As aphids are thought to mate on their preferred host plants, acceptance specificity likely acts as a pre-mating barrier to gene flow (Caillaud and Via 2000, Smadja *et al.* 2009, Peccoud *et al.* 2009a). This survey of acceptance and performance across 16 plant species suggests the situation may not be so clear cut, because the continuum of variation amongst these species may present opportunities for co-occurrence and gene flow. This is most notable on *T. semipilosum* and *M. trunacatula,* where both MT and TP aphids have high levels of acceptance and performance. Interestingly, while *T. semipilosum* does not co-occur with *T. pratense* or *M. sativa*, it is unlikely to act as an alternative host in the field, but *M. trunacatula* does co-occur (www.gbif.org). As the level of divergence between *A. pisum* races is a continuum, with genetic introgression observed between races (Peccoud *et al.* 2009a), it would be interesting to carry out

field investigations to test whether intermediate plant host species in the wild do harbour MS and TP aphids and, if so, what role they might play in gene flow between the aphid races. If co-occurrence on intermediate hosts is common in nature, it would raise interesting questions about how TP and MS aphids are able to maintain distinct host-associated races.

2.5 Conclusions

In accordance with previous research (Schwarzkopf *et al.* 2013), comparisons of EPG waveform measurements have been found here to be an effective measure of host acceptance by aphids. However, assessment of two *A. pisum* races across 19 plant species has shown significant variation in waveform measurements. This indicates aphid acceptance decisions in different plant species might be occurring at different sites within the plant and the use of any one individual waveform measurement as an overall measure of acceptance is unreliable.

Profile scores created by PCA and LDA produce comparable and reliable holistic measures of aphid acceptance. These multivariate scores offer a versatile and conceptually less biased alternative to analysing a large number of individual waveform measurements separately. However, the use of individual waveforms to study acceptance is still recommended, as they provide useful information about the mechanisms of acceptance.

It has been demonstrated quantitatively here that there is a strong correlation between acceptance and performance, as was inferred in several earlier studies but not tested. However, the use of a wide range of typical and non-typical host plants shows that this relationship is not associated with the unique features of a given host plant species, but instead relates to traits that vary between the plants. This opens up interesting questions concerning the chemical ecology and evolutionary dynamics of host specialisation and speciation in *A. pisum*.

Chapter 3: Using aphid responses to multiple plant species to identify the chemical signatures behind host plant acceptance by two specialised races of pea aphid (*Acyrthosiphon pisum*)

3.1 Chapter summary

The huge diversity of phytophagous insects is largely attributable to speciation involving shifts between host plants. These shifts are mediated by the close interaction between insects and plant metabolites, which may act as feeding stimulants or repellents, or influence insect performance, for example through toxicity. However, there has been only limited progress in understanding the chemical signatures that underlie insect preferences. Here, we use the pea aphid (*Acyrthosiphon pisum*) to address this question. Host-associated races of pea aphid discriminate between plant species and some of their chemosensory genes appear to have diverged under selection. We have combined untargeted metabolomic profiling of multiple plant species with tests of differential acceptance by two *A. pisum* races, using random forest regression to identify metabolites that explain variation in acceptance. The identity of some of these compounds was confirmed using tandem mass spectrometry.

Our results reveal that a small number of compounds explain a large proportion of variation in the differential acceptability of plants to *A. pisum* specialised on *Medicago sativa* or *Trifolium pratense.* Two of these compounds were identified as the metabolites L-phenylalanine and L-tyrosine. They indicate a possible deterrent role for alkaloids derived from L- L-tyrosine. The set of compounds implicated in differential acceptability is not related to the set correlated with general acceptability of plants to aphids, regardless of host race. This suggests that small changes in response to common metabolites may underlie host shifts and is consistent with the identification of a small number of divergent chemosensory receptor loci. The results, and the underlying approach, open up new opportunities for understanding the mechanistic basis of host discrimination and host shifts in insects.

3.2 Introduction

Phytophagous insects are extremely diverse and often feed on restricted ranges of host plants (Jaenike 1990). Co-speciation of host plants and insects is common in some taxa (e.g. fig wasps (Cruaud *et al.* 2012) but the majority of speciation events

in phytophagous insects involve shifts in their narrow host range (Matsubayashi *et al.* 2010). Consequently, understanding how host shifts occur is critical in explaining a major component of biodiversity (Matsubayashi *et al.* 2010). Since many phytophagous insects are serious pests of crops (Oerke 2006), understanding what determines the range of host plants acceptable to an insect population, also has important practical implications.

Host acceptance, host-related performance and assortative mating are often tightly inter-connected, especially for species that spend their whole lives on the host plant (Bush 1975, Gripenberg *et al.* 2010). Changes in acceptance may be the first stage in a host shift, and so in host-associated speciation, because they lead automatically to assortative mating (Bush & Butlin 2004). Reproductive isolation may then be reinforced by selection to increase performance on the new host (Bush & Butlin 2004, Drès & Mallet 2002). Therefore, understanding how host acceptance evolves in the early stages of speciation is critically important. While insects may utilise a variety of cues when making feeding decisions, chemical cues (either volatiles detected before feeding or compounds detected during feeding initiation) are very frequently involved (Smadja & Butlin 2009). This focuses attention on the insect chemosensory system, including chemosensory genes, and on differences in plant chemistry among potential hosts.

Feeding stimulants and repellents have been identified in many insect-plant interactions (Bruce and Pickett 2011, Nishida 2014). Feeding stimulants show a wide range of chemistry, for example flavonoids stimulate feeding and oviposition in *Spodoptera* species (Simmonds 2003), nicotine at low concentrations stimulates feeding by peach-potato aphid (*Myzus persicae)* (Ramsey *et al.* 2014), catechol extracts stimulate oviposition for cigarette beetle (*Lasioderma serricorne*) in potential larval food resources (Nagasawa *et al.* 2014), and acylated flavonol glycosides from *Vicia angustifolia* act as probing stimulants for the bean aphid (*Megoura crassicauda*) (Takemura *et al.* 2002). Specific blends of plant volatiles stimulate antennal sensillae in many herbivorous insects indicating that they are used in host location and acceptance (Bruce and Pickett 2011). Examples of antifeedants

are just as varied, including naturally occurring quinones that deter cabbage looper (*Trichoplusia ni*), (Akhtar *et al.* 2012), the lectin 'concanavalin A' that deters *A. pisum* (Sauvion *et al.* 2004) and glucose-derived cycloalkanes widely encountered in plant metabolism, which deter some *A. pisum* clones (Kordan *et al.* 2011). Koul (2008) discusses the role of a wide range of potential antifeedants, including chromenes, polyacetylenes, saponins, quassinoids, cucurbitacins, cyclopropanoid acids, phenolics, alkaloids, terpenes and many more.

These interactions are of interest in pest control but in most cases they do not explain insect specificity, i.e. why insects of closely-related species or host races accept distinct ranges of host species. Cases where a compound, or mixture of compounds, has been shown to be attractive or to stimulate feeding for one insect population but have the opposite effect for a related population, are scarce. One example is the use of volatile blends to discriminate between hosts by divergent races of *Rhagoletis pomonella* (Linn *et al.* 2003). Another, is the divergent chemical content of necrotic host cactus tissue that is associated with the genetic divergence between *Drosophila mojavensis* populations (Matzkin *et al.* 2006).

Host races of *A. pisum*, provide an excellent model to study speciation (Peccoud and Simon 2010) and chemically-induced host plant discrimination. In pea aphids, chemical recognition occurring after penetration of plant tissue is critical to host discrimination (Schwarzkopf *et al.* 2013) and leads to performance differences and assortative mating (Caillaud and Via 2012, Peccoud *et al.* 2014). These conditions result in reproductive isolation and genetic differentiation among races (Caillaud & Via 2012, Peccoud & Simon 2010). There is evidence that divergence in genes involved in recognition via chemoreception (Smadja and Butlin 2009) manipulation via salivary proteins (Jaquiéry *et al.* 2012) and detoxification via P450 proteins (Duvaux *et al.* 2015) has been associated with host shifts. Host acceptance occurs when aphid stylets penetrate plant epidermal layers suggesting that interactions with compounds within plant leaves are important (Powell *et al.* 2006, Schwarzkopf *et al.* 2013). However, the identities of the key plant compounds involved in host discrimination by different races of *A. pisum* are not known.

Here, we have employed untargeted metabolomic analysis, using MALDI mass spectrometry of metabolites extracted from leaves, to characterise variation in potential chemical cues among host plants and related species. Host acceptance by aphids was measured using electrical penetration graphs (EPG). Four clones of the pea aphid, two that are adapted to feed on *Medicago sativa* ('MS aphids' from now on) and two that are adapted to *Trifolium pratense* ('TP aphids') were tested on 19 plant species in the genera *Medicago* and *Trifolium*. This range of comparisons provided us with the statistical power to identify candidate metabolites capable of explaining aphid discrimination from amongst the large number of compounds in the metabolomic profiles.

Acceptance was summarised for each combination of host species and aphid races as either the time spent in the E2 phase, a measure of sap ingestion during phloem contact ('E2 profile'), or as the first linear discriminant axis ('LD1 profile') based on 60 variables extracted from EPG traces. In either case, we derived scores for 'discrimination' by MS vs TP aphids and for 'overall acceptability' by all aphids. The relationship of these scores to aphid performance were tested. Random forest (RF) regression was then used to search the polar and non-polar fractions of the metabolomics data for the best predictors of discrimination and acceptability.

3.3 Methods

Methods used for aphid, plant culturing and EPG data collection are the same as for Chapter 2. These are outlined in brief below.

3.3.1 Aphid culture

Four asexually-maintained lineages (clones) of *A. pisum* were used; the *Medigcao sativa* specialised clones LSR1(International Aphid Genomics Consortium, 2010), and L9Ms_052 (source SE France, supplied by JC Simon, INRA, Rennes) and *Trifolium pratense* specialised clones YR2 (Simon *et al.* 2011) and L7Tp_232 (source SE France, supplied by JC Simon, INRA, Rennes). Aphids were kept at a density of 10-15 individuals on a single 10 day old *V. faba* plant. Age-controlled

aphids were produced by exposing plants to adult aphids for 24 hours then removing adults. Progeny were then left to develop for 14 days. On the day of use aphids were taken off the bean plants and starved for 1 hour before starting experimentation.

3.3.2 Plant culture

In total 19 species of plant were used from the genera *Medicago* (*M. arabica, M. orbicularis, M. littoralis, M. tornata, M. turbinata, M. laciniata, M. lupulina, M. truncatula, M. sativa*) and *Trifolium* (*T. ambiguum, T. striatum, T. nigrescens, T.repens, T. pratense , T.ochroleucum, T. rubens ,T. semipilosum, T. dubium, T. pallidum*)(for seed origins see Appendix 2). Seeds were sterilised by soaking in saturated calcium hypochlorite solution for 2 minutes and then plated out in Petri-dishes containing 1.2% plant agar containing 50 mg/ml giberrellin (source Sigma-Aldrich UK©). These were left to germinate for one week at 20^0C day and 15^0C night temperatures, with 16 hour day length. Resulting seedlings were then transferred into seed trays containing 4:1 sand and John Innes no.2 compost mix and covered with a lid to retain humidity. Plants grew for five weeks in total and were watered twice weekly with distilled water. Plants were fed twice with Rorison's solution (Appendix 4): in the 4^{th} week with 40% of full strength solution, and in the 5^{th} week with 20% of full strength solution.

3.3.3 Measuring aphid host preference using EPG

Aphid acceptance was measured by the electrical penetration graph (EPG) method (Tjallingii, 1978) using a DC Giga-8 sourced from EPG systems © (www.epgsystems.eu). This technique records the changes in the potential difference (measured as output voltage) as the resistance to a weak electrical current is effected by the progression of an aphid stylet through a leaf to the phloem (Tjallingii 1978). Using this information, changes in the patterns can be interpreted into "waveforms" that represent particular aphid behaviours (Tjallingii 1978). Behaviour for each aphid was recorded for 6 hours. Eight EPG recordings were performed per day. A blocked design was used, with each plant species represented during each five week block, and on each week-day within a block (Appendix 13). Plant species –aphid clone

combinations were set up in duplicate within days so that occasional recording failure (e.g. due to the aphid leaving the test plant) did not interfere with the block design. Five to 11 aphids were recorded successfully for each plant species and aphid clone combination.

3.3.4 EPG profile of acceptance

EPG traces were interpreted into "waveforms" by manually using the Stylet+ software (www.epgsystems.eu) and the waveform key found in Sarria *et al.* (2009) . Annotated EPG recordings were then entered into the MicroSoft Excel macro "workbook for EPG parameter calculations of EPG data: version 4.4" (Sarria *et al.* 2009) which was used to calculate 119 separate behavioural measurements for each recording. Missing values were imputed using the RFimpute() function in R (Liaw and M. Wiener (2002). Uninformative variables were cleaned from the data set by removing waveforms with 50% or more values equal to zero and then removing one variable from any pair of variables with correlation >0.80. After this 60 variables remained (Appendix 7).

Two summary statistics were then created for each recording and then averaged for each aphid race-plant combination. The first ("E2 profile") was the mean of the "Total duration of E2 waveform", which represents the time an aphid passively ingests phloem sap (Tjallingii 2006). E2 duration was chosen as it can easily be interpreted as a measure of acceptance of the plant for feeding. The second statistic (LD1 profile) was the mean score on the first axis of a linear discriminant analysis of all 60 EPG waveforms, calculated for each aphid race separately, with plant species as the grouping factor.

The difference between the means of the E2 and LD1 profiles for the two aphid races was calculated to provide a measure of discrimination and the sum of the means was used to measure overall acceptability of the plant species. The EPG data summaries were then compared to the data on aphid performance on plants growing under the

same conditions using Pearson's product-moment correlation coefficient across the means of all plant species. Performance was measured as the number of live young produced by a single 10 day old adult aphid (reared on *V. faba*, as detailed above) over a seven day period, following transfer to a fresh test plant. Due to poor germination there was insufficient plant material of *T. ochroleucum*, *T. pallidum* and *T. semipilosum* to use in this part of the study. For the performance measure there were three to eight replicates for each clone-plant combination and six to 16 replicates for each race-plant combination.

3.3.5 Plant metabolomic profile

Half-way through the EPG data collection period the first fully formed leaf from each of 5-7 plants of each species, was cut, weighed and then immersed in liquid nitrogen. Metabolites were then extracted from the frozen leaf material using the cold extraction methanol-water-chloroform method, as described in Field and Lake (2011). From this extraction two phases, polar and non-polar, were separated for analysis.

The concentration of each extract was adjusted according to the original leaf weight to account for a large difference in leaf sizes between species. Extracts were then diluted by 50% with methanol. Metabolic profiles were recorded using MALDLI TOF (full instrument settings in Appendix 14). Metabolite profiles for individual plants were created by binning the crude m/z values into 0.2-unit bins (m/z bin) and the mass abundances [% total ion count (%TIC)] for each bin were summed (Field and Lake 2011).

3.3.6 Plant metabolome by aphid phenotype comparison using Random Forest regression

The discrimination and overall acceptability scores derived from the EPG data were used as response values to regress against the metabolic profiles of individual plants using the "randomForest()" package in R (Chen *et al*. 2004). To account for

uncertainty in the discrimination and acceptability scores, we repeated each RF analysis 500 times. For each repeat, random score for each plant species were drawn from distributions defined by their observed means and standard errors. We then recorded the rank value of the RF importance (measured as mean decrease in Gini) for each m/z bin. Median ranks were then used to sort m/z bins and inter-quartile ranges of the ranks were used to assess consistency of variable importance. The highest ranking m/z bins were included in linear regression models, to assess the proportion of variance they explained in aphid discrimination between or overall acceptability of plants.

In total eight RF models were analysed, one for each combination of: discrimination and overall acceptance, LD1 and E2 profiles and polar and non-polar plant metabolomic data. Results were then compared across the RF models to identify m/z values with high importance held in common. For bins of high importance, we examined correlations between the behavioural scores m/z bin values.

3.3.7 Putative mass identification and tandem mass spectrometry

The putative identities of compounds in m/z bins were investigated using the comprehensive Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.genotome.jp/kegg) and MetaCyc Compound (www.biocyc.org) databases. Putative compounds (sourced from Sigma-Aldrich UK©) were then obtained and used as standards alongside m/z bins in ESI TOF tandem mass spectrometry (tandem MS) in order to compare fragmentation patterns (full instrument settings in SI). Finally, the KEGG pathway (www.genome.jp/kegg/pathway) was used to investigate the relationships between the different putative compounds in plant metabolism.

3.4 Results

3.4.1 Analysis of EPG profiles

Both LD1 and E2 profiles from the EPG showed a continuum in overall acceptability and discrimination across the 19 host plant species (Figure 3.1 and Appendix 15) and these measures were uncorrelated (r^2 = 0.11, P = 0.093). The LD1 and E2 profiles were strongly correlated (MS aphids r^2 = 0.82, P < 0.001; TP aphids r^2 = 0.84, P < 0.001) and aphids of different clones, within races, showed very similar profiles (MS aphid clones r = 0.88, P < 0.001; TP aphid clones r = 0.75, P < 0.001). There were significant correlations between aphid performance on each plant species, measured as fecundity of single adults over seven days. For both E2 (MS aphids r^2 = 0.59, P < 0.001; TP aphids r^2 = 0.63, P < 0.001) and LD1 profiles (MS aphids r^2 = 0.58, P < 0.001, TP aphids r^2 = 0.65, P < 0.001) (Chapter 2: figure). This is in line with previous observations that EPG provides meaningful measures of host acceptance (Caillaud and Via 2000; Gao *et al.* 2008; Schwarzkopf *et al.* 2013).

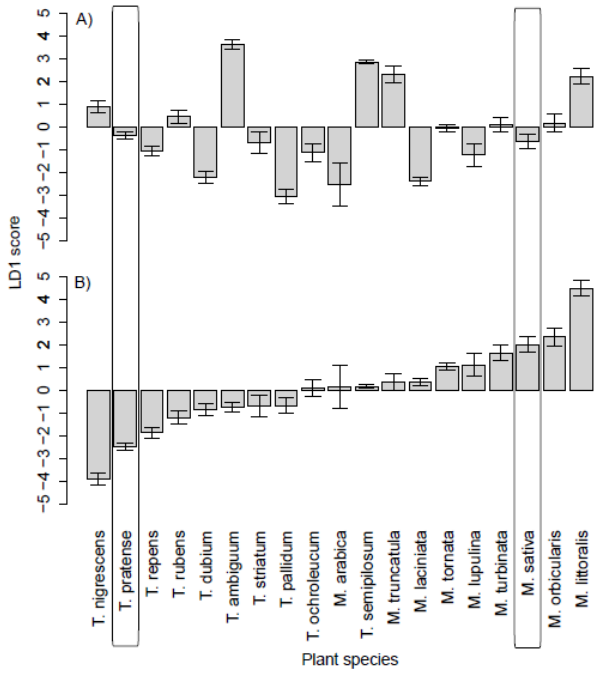


Figure 3.1: EPG profiles (first linear discriminant axis) of MS and TP aphid races for each plant species. A) Accumulated acceptance profile with positive LD1 values indicating preferences and negative values indicating rejection of plants. B) Discrimination profile with positive values indicating overall MS aphid preference, and negative overall TP aphid acceptance. Solid outlines indicate the natural host species of MS aphids and TP aphids. There were 2 clones per race, 5-11 replicates per clone and plant species. Mean ± SEM,

3.4.2 Random Forest models and identification of candidate masses

Metabolomic profiles of plants showed considerable overlap among host plant species in composition of both polar and non-polar fractions and only weak separation between the plant genera (Appendix 16). However, RF regression for discrimination profile, using either 955 polar or 965 non-polar m/z bins, identified a small number of bins with consistently high explanatory power, as indicated by importance rank (Figure 3.2). The 8 top scoring m/z bins, from the polar and non-polar data sets together explained, 44% of the variation in the LD1-based discrimination score. In contrast, RF regression for the LD1-based overall acceptance score revealed no specific masses with consistently high explanatory power: the top 8 m/z bins together explained less than 1% of variation.

Comparison of RF models of aphid discrimination with those for overall acceptance, showed that very different combinations of m/z bins were implicated (Appendix 17). Analyses based on E2 profile gave very similar results (Appendix 17). This suggests that a few distinct chemical signatures underlie plant discrimination by aphids of different ecotypes and that the compounds involved are different from those that explain overall acceptability. Seven top scoring m/z bins from both extracts for aphid discrimination were considered for further analysis (Table 3.1). Of these m/z bins 182 and 166 in both polar and non polar plant extracts and 183 in the polar plant extracts, were of high concentration in plants associated with high acceptance scores by MS aphids and low acceptance score by TP aphids (Figure 3.2). Compounds in these bins were generally more abundant in *Medicago* than in *Trifolium* species. In contrast, m/z bins 269, 291, 292 and 285 in the non-polar extracts were of high concentration in plants associated with acceptance by TP aphids and rejection by MS aphids (Appendix 18). Compounds in these bins were abundant in *T. pratense* and *T. nigrescens*, in particular.

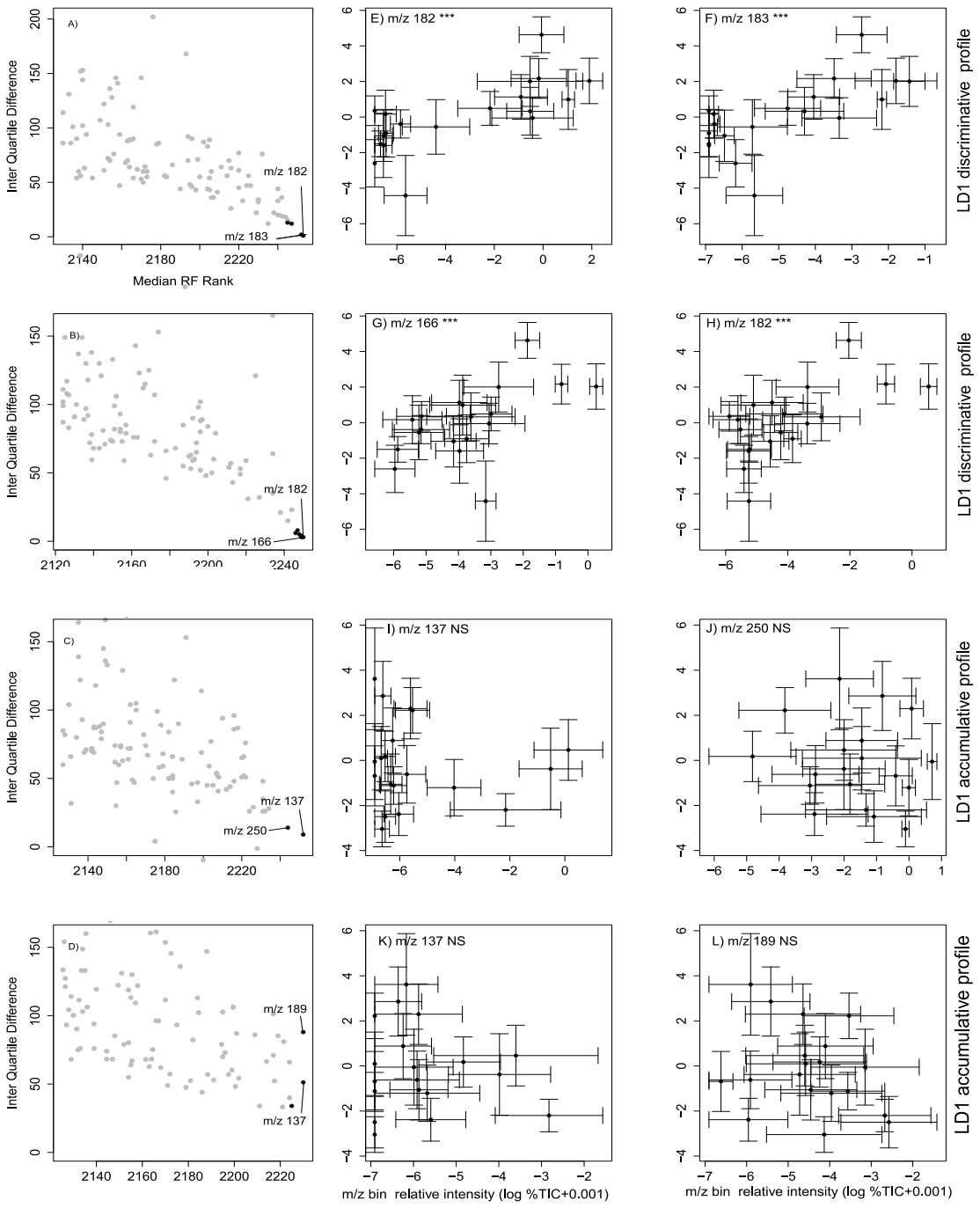


Figure 3.2: A-D; Median ranked importance (mean decrease in Gini coefficient) of the top 100 m/z bins from RF regression models against the difference between 0.25 and 0.75 quartiles of the variation in m/z bin importance between 500 individual RF runs. Rank RF importance was determined in 500 regressions RF models using LD1 discriminate acceptance profile against A) polar data and B) non polar metabolic data and LD1 overall acceptance profile against C) polar data and D) non polar metabolic data. For each regression, discrimination scores for each species were drawn at random from a normal distribution determined by the observed means and standard errors (Figure 3.1). Labelled points in black are key m/z bins used for further investigation. In total there 995 m/z bins with some level of intensity in both the non polar and polar fractions. E:L; Key m/z bin values highlighted by RF models correlated with LD1 discriminate acceptance score. Mean ± SEM, Significance tested Spearman's rank correlation coefficient with FDR correction. **P*<0.05, ***P*<0.001, ****P*<0.001.

Searches of the KEGG (www.genome.jp/kegg/pathway) and MetaCyc (www.biocyc.org) databases revealed putative identities of m/z bin 166 as L-phenylalanine and m/z bin 182 as L-tyrosine. Comparison of tandem mass spectrometry fragmentation patterns of these bins to standards, confirmed these putative compound identities (Figure 3.3). Identification of other M/z bin values has remained inconclusive as tandem MS fragmentation patterns failed to match fragmentation patterns of standard of putative compounds.

Table 3.1: Top Mass/charge (m/z) bins from polar and non-polar samples identified by RF models. The relationships between abundance (%TIC per bin) and the LD1 discrimination profile were tested with Spearman's rank correlation (S). DF = 17.

m/z bin value	ρ	*P*
Polar		
m/z 182	0.76	> **0.001**
m/z 183	0.73	>**0.001**
m/z 166	0.79	> **0.001**
Non polar		
m/z 182	0.70	**0.001**
m/z 166	0.65	**0.003**
m/z 269	-0.53	**0.018**
m/z 285	-0.53	**0.019**
m/z 291	-0.18	0. 46
m/z 292	-0.32	0.18

3.4.3 Analysis of L-phenylalanine and L-tyrosine pathways

Pathway analysis of downstream compounds associated with L-phenylalanine and L-tyrosine showed a number of compounds had a higher abundance in plants with higher MS aphid acceptance (*M. littoralis, M. sativa, M. orbicularis*) than higher TP accepted plants (*T. nigrescens, T. pratense and T. repense*) (Figure 3.4). These included higher concretions in MS accepted plants of M/z 198 (polar: F = 4.97, df = 1, 34, *P* =0.033) and m/z 154(nonpolar: F= 5.34, df =1,31 , *P* =0.028) values, putatively identified as L-DOPA and dopamine in MS accepted plants. Additionally there are a number of other putative compounds downstream from dopamine that seem to be associated with aphid acceptance (Figure 3.4 and 3.5). These include putatively identified 4-Hydroxyphenyllactate (m/z 183 in polar samples: F=33.83, df = 1, 34, p-value = 0.001, nonpolar samples: F= 4.6, df = 1,31, *P* =0.04),

Norcoclaurine (m/ 272 in polar sample: F = 3.08, df = 1,31 , *P* = 0.046) and 4-Hydroxyphenylacetate (m/ 153 in polar sample: F = 8.47 , df = 1, 34, *P* =0.006). M/z 183 was also found to be important in the RF models and significantly correlated to LD1 differential score (Figure 3.2).There were no significantly higher concentrations of the downstream metabolites examined found in TP accepted plants, with the exception of m/z 286 (F=6.1979, df= 1,31,*P*=0.018) putatively identified as Coclaurine (Figure 5). Overall this give the impression that in MS accepted plants pathways associated L-phenylalanine and L-tyrosine metabolism may be linked to divergent aphid acceptance.

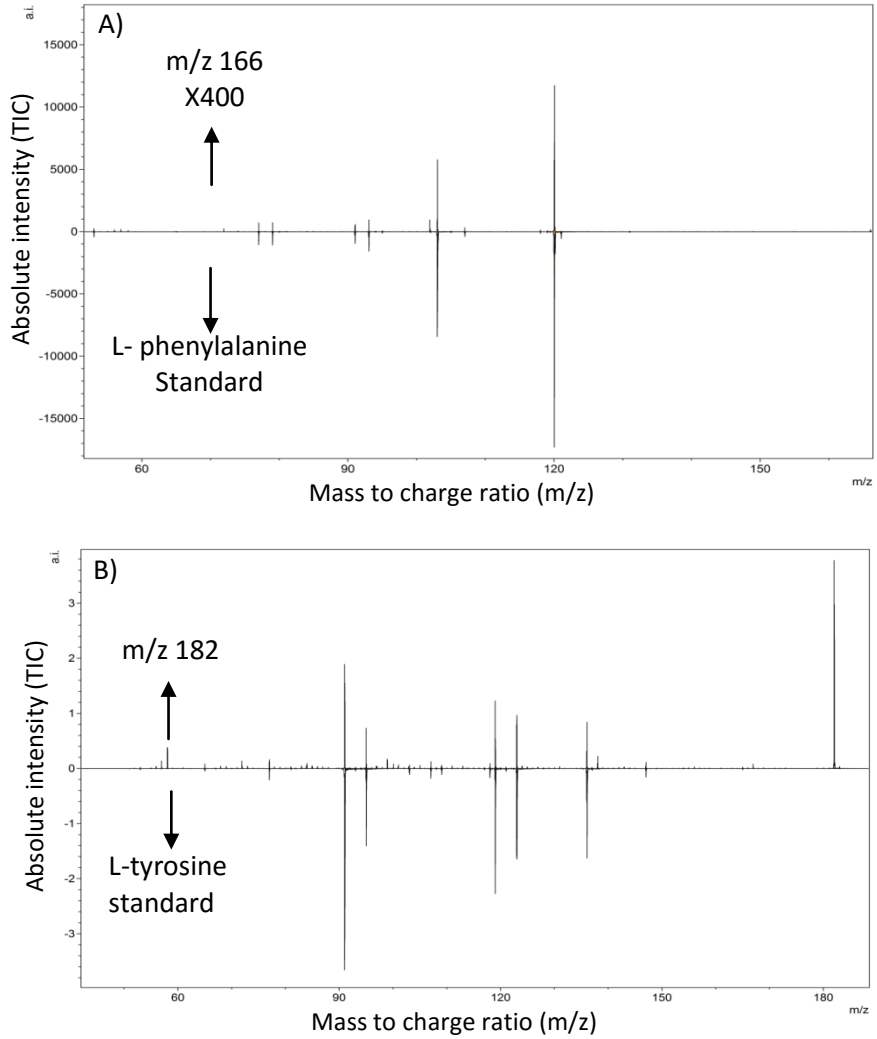


Figure 3.3: Tandem mass spec plots of fragmentation patterns. A) m/z 166 in polar samples against L-phenyalanine and B) m/z 182 in polar samples against L-tyrosine. The unmatched peak at the far right of plot B) is at m/z 182 so is likely to be the target compound which remained unfragmented within the more complex extract matrix.

3.5 Discussion

Our strategy of analysing the differential responses of aphid host races to 19 different plant species, using Random Forest regression, revealed a very small number of m/z bins capable of explaining a high proportion of the variance in aphid discrimination profiles, as measured using electrical penetration graphs. The compounds responsible for discrimination by host races appear to be different from those underlying general acceptability to aphids.

A limitation with untargeted MS techniques using MALDI is the notion of *charge theft* during ionisation known as ion suppression (or competitive ionization) (Duncan *et al.* 2008). Ion suppression occurs when MS is used to test a material that consists if a complex matrix of many a different compounds, meaning ionisation of the sample favours compounds at higher concentrations or with particular chemistries, such as amino acids, which then bias the end TIC (Duncan *et al.* 2008). This can produce noisy data in the lower concentrated compounds. As we were constrained by relatively small replicate sizes this meant the occurrence of false negatives (i.e. missing the importance of a less concentrated compound or compounds less likely to accept positive charges) in the RF model was likely. Additionally, knowing many m/z values identity for certain is not always possible as i) some compound could overlap in the spectra and can be difficult to separate in tandem MS fingerprints, ii) acquisition of pure sample of putative compound for tandem MS is always not possible, iii) techniques and machinery used in mass identification can differ across laboratories meaning the accuracy of masses used in online database may vary, and iv) our knowledge of the metabolomes of all plants is not complete, meaning masses of novel chemistry cannot be identified without extensive follow up. These are common challenges shared across the field of metabolomics (Sugimoto *et al.* 2012). However, the untargeted metabolic profiling approach used has some major advantages. Most importantly it avoids bias from any preconceived expectations about on what compounds should be looked for. This allows us to discover new compounds or compounds leads that would otherwise not be discovered using a targeted approach. The use here of many plant species also gives the statistical power to enable important signals to be detected.

In plants accepted by MS aphids and rejected by TP aphids, m/z bins 182 and 166 in the polar/non-polar fraction had consistently high values and these bins have been identified as L-phenylalanine and L-tyrosine. It is possible that plant discrimination by aphids occurs because of a direct response to these metabolites. Indeed, previous studies have shown that L-tyrosine can act as an allelochemical, deterring insect herbivory. For instance, hyper-production of L-tyrsoine in plants of the genus *Inga* (Fabaceae) has been linked to their defence against insect attack (Lokvam *et al.* 2007, Lokvam 2006). Alternatively, the metabolism of L-tyrosine may generate active compounds. The identification of both L-phenylalanine and L-tyrosine in this study is of particular interest as they are directly linked by the same plant metabolic pathway (Figure 3.4) (www.genome.jp/kegg/pathway.html). The presence of these two compounds suggests that aphid discrimination may respond to expression of this pathway, or to metabolites produced by this pathway (figures 3.4 and 3.5). Indeed, L-tyrosine metabolism is the precursor of a number of biologically active alkaloid families, including several plant-derived psychoactive compounds (www.genome.jp/kegg/pathway.html, www.biocyc.org). It could be that pehenylalanine and L-tyrosine show up simply because they are part of a pathway to other compounds that are not picked up in this analysis but are actually determining preference.

Work by Sempruch *et al.* (2013) has shown that prolonged exposure of *Pisum sativa* plants to pea aphids increases L-tyrosine decarboxylase (TyDC) expression across the whole plant. It has also been shown that production of tyramine by L-tyrosine decarboxylase in winter triticale (x *Triticosecale*) reduced plant acceptance by grain aphids (*Sitobion avenae*) (Sempruch *et al.* 2009). This suggests that aphids can trigger whole plant defence priming, via TyDC expression, to metabolise L-tyrosine. TyDC activation could be underpinning the production of defensive alkaloids (Sempruch *et al.* 2013, Premont *et al.* 2001), although TyDC has also been linked to the production of structural compounds known as hydroxycinnamoyl amides (HCAA). HCAA are used by plants to strengthen cell walls against decomposition and have been shown to improve resistance towards abiotic stress and various pathogens (Newman *et al.* 2001, Facchini *et al.* 1999, Zacarés *et al.* 2007, Macoy *et al.* 2015). It is possible that aphids specialised on *Medicago* hosts have evolved

either tolerance or suppression of a plant defence response, involving heightened L-tyrosine metabolism, while TP aphids have not and that this is reflected in behavioural responses to L-tyrosine and/or phenylalanine concentrations in leaves.

There is evidence to suggest that aphid discrimination of host plants is due to an interaction with plant chemistry early in plant penetration (Powell *et al.* 2006, Schwarzkopf *et al.* 2013). Additionally, aphids in general are known to manipulate plant metabolic pathways during early feeding (Hogenhout and Bos 2011,Tjallingii 2006, Will and van Bel 2008, Will *et al.* 2013), including the pea aphid (Will *et al.* 2007), suggesting the possibility of co-evolution between *A. pisum* races and the metabolic expression their host plants. Consequently aphids might differentially perceive or respond to biological pathways that lead to the production of harmful or deterrent compounds produced from L-tyrosine. In support of this suggestion, a number of putative metabolites downstream from L-tyrosine were in higher concentrations in plants accepted by MS aphids than in those accepted by TP aphids, such as norcolaurine, 4-hydroxyphenyllactate, 4-hydroxyphenyllactate and scopeletin, protocatechuatealdehyde (Figure 3.4 and 3.5).

The higher elevation of m/z bins with a putative identification of L-DOPA and dopamine in MS aphid accepted plants is of particular interest (Figure 3.4 and 3.5). Dopamine and L-DOPA are important secondary metabolites in insects as they are used in the production of melanin, necessary for cuticle formation and for insect cognition; including the regulation of feeding behaviours (Chapman 2013 pp37, Barron *et al.* 2010, Vavrick *et al.* 2011). In plants L-DOPA, dopamine and their metabolites are found across the plant kingdom with varied and potentially important biological functions (Rehr *et al.* 1973, Kulma and Szopa 2006). This includes members of Fabaceae that are known to contain L-DOPA (Andrews and Prideham 1965, Brain 1976, Wichers *et al.* 1993, Patil *et al.* 2015) and dopamine (Kulma and Szopa 2007). Several *Mucuna* species and *V. faba* in particular are known to contain considerable concentrations of L-DOPA (Brain 1976, Wichers *et al.* 1993, Patil *et al.* 2015). Plant derived L-DOPA and dopamine is interesting in respect to plant-insect interactions as they may possess toxic qualities (Van Alstyne *et al.* 2006,

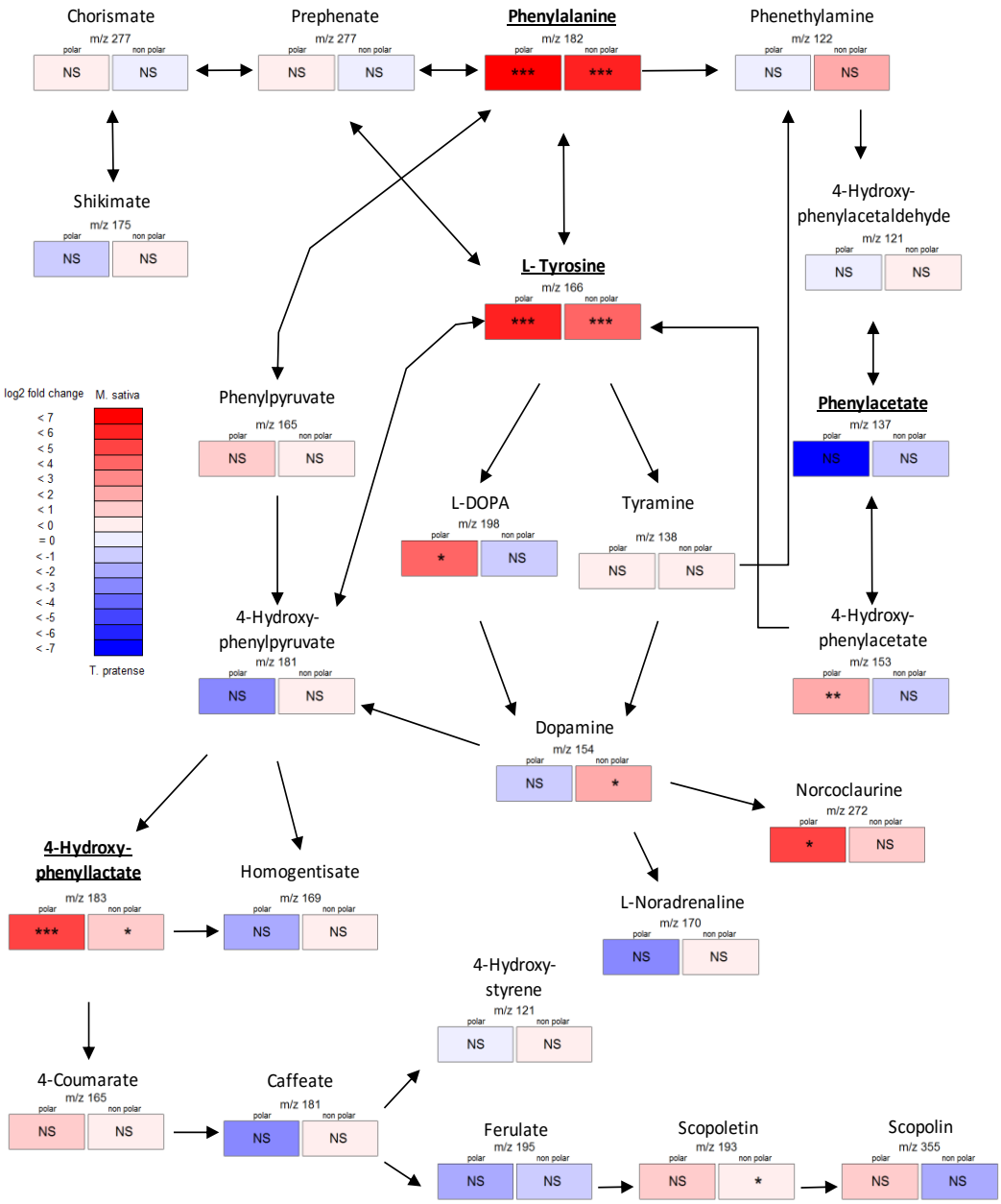


Figure 3.4: Plant metabolic pathways stemming from phenylalanine and L-tyrosine metabolism. Colour represents the Log fold change in the mean %TIC of m/z bin values between top three the most divergent *Medicago* plants (*M. sativa (*n=6), *M. littoralis* (n=6), M. orbicularis(n=6))and the top three *Trifolium* plants (*T .pratense* (n=5*), T. nigrescens (*n=6), and *T. repens* (n=7)). The pathway is based on *M. trunculata* pathway in KEGG pathway (www.genome.jp/kegg/pathway.html). Significance is tested between the two groups on the combined log %TIC values with linear regression model. Compounds underlined are those with corresponding m/z bin values identified in both discriminatory and accumulative RF models.

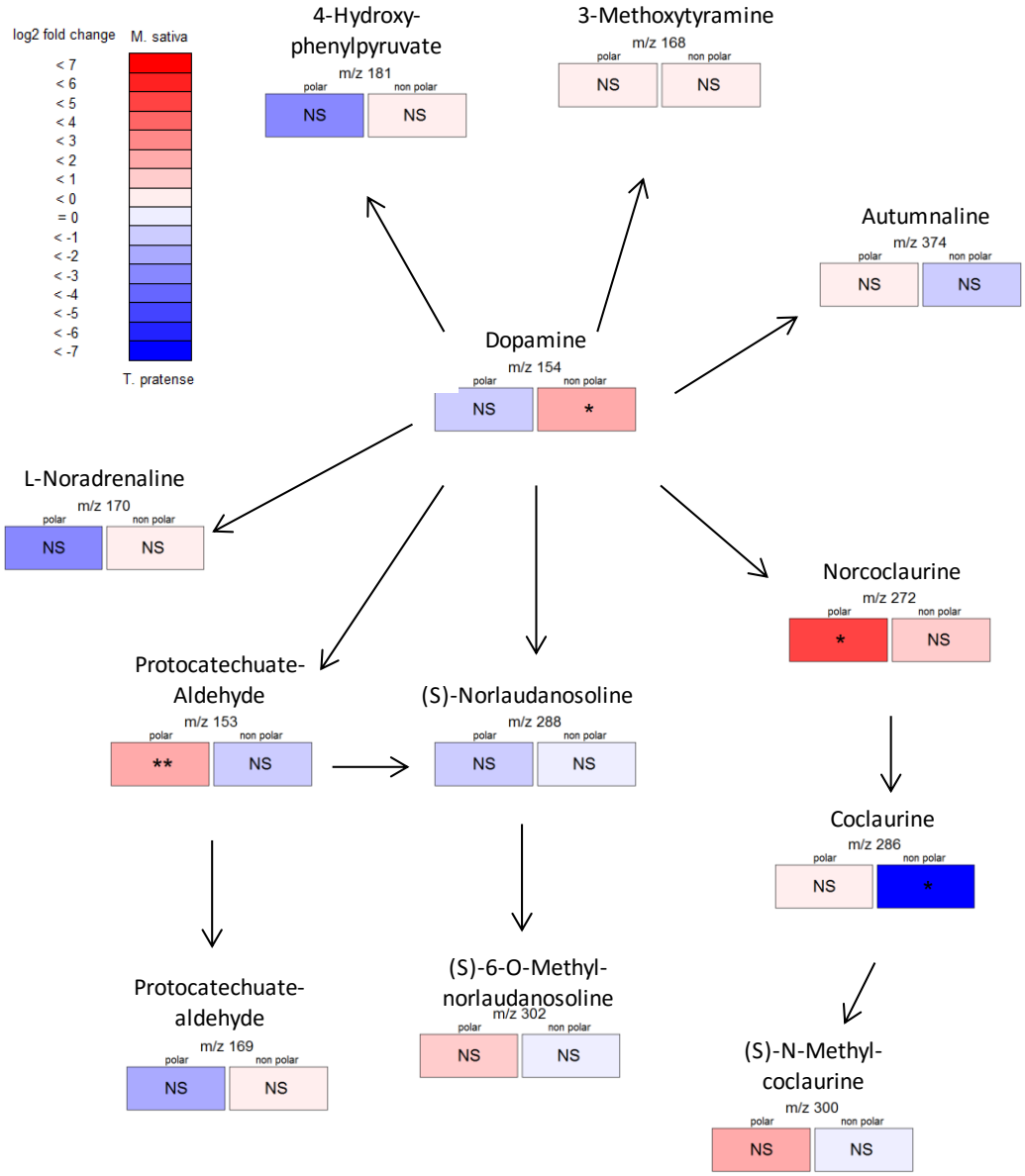


Figure 3.5: Plant metabolic pathways stemming from dopamine metabolism. Colour represent the Log fold change in the mean %TIC of m/z bin values between top three of the most divergent *Medicago* plants (*M. sativa* (n=6), *M. littoralis* (n=6), M. orbicularis(n=6))and the top three *Trifolium* plants (*T .pratense* (n=5), *T. nigrescens* (n=6), and *T. repens* (n=7)). The pathway is based on *M. trunculata* pathway in KEGG pathway (http://www.genome.jp/kegg/pathway.html). Significance is tested between the two groups on the combined log %TIC values with linear regression model. Compounds underlined are those with corresponding m/z bin values identified in both discriminatory and accumulative RF models.

Huang *et al.* 2011, Gary *et al.* 2013) and have been shown to disrupt insect cuticle development (Rehr 1973). In addition plant can metabolise dopamine into melanin to cause the release of reactive oxygen species (ROS) which are involved in cell-defence signalling (Gupta and Igamberdiev 2014) and triggers the formation of HCAAs in response to biotic and abiotic stress (Macoy et. al 2015).

The general increase of other downstream putative phenylopropanoids (plant metabolites derived from L-phenylalanine) and putative catecholamines (compounds derived from L-tyrosine that posses benzene rings with two hydroxyl side groups) in MS accepted plants suggests the L-tyrosine pathway as a whole may play a role in aphid acceptance. Across the plant kingdom the metabolim of L-tyrosine to L-DOPA and dopamine leads to a host of interesting plant alkaloids known, which could function as aphid deterrents, including the ipecac alkaloids, isoquinoline alkaloids (e.g. mescaline), glucosinolates and isoflavonoid (Piasecka *et al.* 2015), benzylisoquinolines, morphine, norepinephrine, epinephrine and phenethylamine (Herbert *et al.* 1985, Premont *et al.* 2001, Kulma & Szopa 2007, Hagel & Facchini 2013, Piasecka *et al.* 2015). Indeed both epinephrine and norepinephrine have previously been identified in bean (*Phaseolus vulgaris*) and pea (*Pisum sativum*) plant material (Kulma & Szopa 2007). Such compounds can be biologically active at very low concentrations, readily modified to other compounds and many are produced predominately within the leaf (Kulma & Szopa 2007). This suggests such metabolites are potentially functional as rapid and diverse metabolic plant defence responses to aphid attack.

This evidence suggests a potential role of other interesting metabolites not identified in this analysis but which include highly active compounds that could influence aphid-host plant choice. With the relatively small replicates and low sensitivity inevitable from using a broad spectrum approach the role in *A. pisum* host choice of broader secondary metabolites stemming from L-tyrosine is currently notional. In summary there are two possible explanations for the association of L-phenylalanine and L-tyrosine that: i) that these compounds are constitutive (non changing) in plant and act divergently towards aphids to elicit or repel aphid feeding, or ii) MS

accepted plants possess inducible metabolic responses associated with L-tyrosine metabolism to which the two aphid have divergent adaptations towards. To establish which one of these hypotheses is true requires further investigation. However, the possibility of a link between divergent *A. pisum* acceptance to M. sativa plants and highly biologically active plant compounds, not only adds weight to the pathway hypothesis of plant discrimination by *A. pisum*, but also provides promising lines of investigation.

In addition to the three m/z bins associated with acceptance by MS aphids, four other m/z bins 269, 291, 292 and 285 in the non-polar plant extracts were associated with TP aphid acceptance and MS aphid rejection. These compounds were predominantly abundant in *T. pratense* and *T. nigrescens*. Compound identification of these compounds was inconclusive as Tandem MS fragmentation patterns of putative compounds did not match fragmentation patterns of sample masses of putative compounds. It is feasible these m/z values represent the fragments of larger compounds. The use of TOF MS to investigate more mass ranges and with different ionising energies alongside in silico modelling of fragmentation patterns using software such as such as chemdraw©, may reveal the likely identity of these compounds. Unfortunately this level of investigation was beyond the scope of this project.

3.5.1 Integration of plant metabolomics with aphid genetic studies

These results make an initial link between our current knowledge of genetic divergence, underlying *A. pisum* host race formation, and the chemical basis of the divergence in host plant selection by different races. For instance it has been shown that *A. pisum* chemoreceptors (CR) could play an important evolutionary role, with many genes having undergone a significant and recent gene expansion which is thought to be driven by positive selection (Smadja *et al.* 2009). To support this claim Smadja *et al.* (2012) found that fewer members of the Cr families were more genetically divergent between host races than would be expected under neutrality. In addition to the sequence divergence, another study has identified divergent copy

number variations (CNV) between aphid races in loci for both the CRs and monooxygenase cytochrome P450 (P450) genes (Duvaux *et al.* 2015). CNV between races occurred particularly within gustatory receptor loci, suggesting that copy number evolution may be important in specialisation, perhaps through effects on gene expression (Duvaux *et al.* 2015).

The divergent evolution of a subset of odorant and gustatory receptors, points to a mechanism of host plant specialisation based on aphid perception of plant chemical constitution (Smadja and Butlin 2009). On the other hand, the divergent copy number of P450 genes suggests divergent adaptation of aphid ability to metabolise plant allelochemicals (Duvaux *et al.* 2015). In support of this idea, aphids are well known for their ability to both avoid and suppress plant chemical defences (Giordanengo *et al.* 2010, Schwartzberg and Tumlinson 2014).

3.6 Conclusion

Our study adds significantly to the evidence that aphid host race formation and speciation is driven by specialised adaptations to the chemistry of plants via the perception and/or breakdown of a few specific plant compounds. While previous work has suggested that *A. pisum* interactions with leaf chemistry in early plant probing were key to host acceptance, until now little was known of the plant chemistry responsible. Although, it has not been possible to test functions of chemoreceptors that appear to have diverged under selection, our results provide an important step forward in identifying the critical plant metabolites involved in divergent host selection, using a novel approach that can readily be applied in other systems.

Chapter 4: The influence of host-plant pre-exposure to pea aphid (*Acyrthosiphon pisum*) aphids from two races upon aphid acceptance and key plant compounds.

4.1 Chapter Summary

Phytophagous insects need to adapt to a diversity of constitutive and induced plant defences, many of which are thought to be chemical. Previous work suggests that pea aphid (*Acyrthosiphon pisum)* host race divergence has occurred mainly in response to host plant chemistry. The aim of this research is to investigate the role of plant defence responses in host acceptance by *A. pisum*.

To study this *Medicago sativa* and *Trifolium pratense* plants were exposed to a native *A. pisum* clone, a non-native *A. pisum* clone or neither (i.e. a control). Difference between exposed and non-exposed plants was tested for by measuring the acceptance of second naive inoculum of aphids using electrical penetration graph (EPG). This was to investigate if plant responses to the two *A. pisum* clones are constitutive, or if aphid attack results in the induction or suppression of plant defences. EPG results suggested that: i) during probing stages a divergence in aphid response is caused by constitutive deterrents in *M. sativa* that are not present in *T. pratense*, ii) for *M. sativa* plants, during or, just before, phloem feeding, there is a divergence in aphid induction or suppression of defences that is most marked for TP_232 aphids, and iii) on *T. pratense* there is divergent suppression of plant defences at some point between the probing and phloem stage that has a positive effect on subsequent feeding by either aphid race. In addition aphid induced changes to seven previously identified target masses in the plants were measured using time-of-flight mass spectrometry (TOFMS). It was found that the L-tyrosine concentration in *M. sativa* was slightly reduced when exposed to the non-native aphid clone. Divergent aphid acceptance appears to be in response to both constitutive feature of plant and induced plant defences, with aphids appearing to subvert some induced plant defences.

4.2 Introduction

Specialist phytophagous insects have a very close association with their host plants, which often act as their food resource, mating site and habitat for most stages of their life-cycles. This close relationship requires insects to adapt to plant structure, phenology, nutritional content and especially plant defences (Simon *et al.* 2015).

Insects need to either tolerate (e.g. through detoxification) or develop mechanisms to subvert these defences. In addition, plant compounds, including defence compounds, can be used by many insects for host plant location and identification. For example Ramsey *et al.* (2014) showed that low concentrations of nicotine can stimulate feeding by tobacco adapted clones of the peach-potato aphid (*Myzus persicae*), despite nicotine being toxic to other *M. persicae* clones.

There are two broad classes of defences deployed against insects by plants, constitutive and induced. Constitutive defences are permanently expressed regardless of the presence of insects, whilst induced plant responses are changes in plants that result from insect activity, especially feeding (Walters 2011). Defences can be induced in response to chemical elicitors such as Pathogen-Associated Molecular Patterns (PAMPS) (Jaouannet *et al.* 2014) and more recently coined herbivore-associated molecular patterns (HAMPS) (Mescher and De Moraes 2015). Moreover, inducible defences can be modulated by the phytophagous insect by the masking of physical cues or by the production of effector proteins that suppress defence activation and/or signalling cascades (Walling 2008, Bonaventure 2012). Thus, when observing insect-plant interactions over time the complexity of insect behaviour can be summarized into three types of changes: i) an unchanged insect response to constitutive features of a plant, ii) an increase in insect repulsion, post insect attack as plant defences are induced or iii) improvement of insect host acceptance because the insect possesses traits that subvert induced plant defences or improve plant quality. As plant response plays a crucial role in evolution of plant insect-interactions, studying the various roles of constitutive, induced and subverted plant chemistry on selection is a useful way to test how phytophagous insects have adapted to different host plants.

The multiple host specialised biotypes found in the pea aphid (*Achrythosiphon pisum*) species complex (Peccoud *et al.* 2009a) makes it a useful subject to study the role of plant chemistry in divergent host selection. The role of constitutive, induced and subverted plant chemical defences on host-plant choice is an especially interesting question in regards to aphids, as it has been long recognised that plant

chemistry plays an important role in their choice of hosts (Emden 1972, Pickett *et al.* 1992, Dixon 1998 pp32-38, Powell *et al.* 2006). At the same time, as stealthy (i.e. eliciting very little plant defence response), specialist sap-feeders, aphids have a close relationship with the internal chemistry of host plants (Walling 2000, Walling 2008) and it is suggested that plant chemistry may play an important role in divergent host plant selection by *A. pisum* races (Smadja *et al.* 2012, Duvaux *et al.* 2015, Simon *et al.* 2015).

In Chapter 3, a number of compounds were found to correlate significantly with host acceptance by two races of aphids on multiple *Medicago* and *Trifolium* plant species. For *A. pisum* divergent acceptance of *Medicago* species is associated with L-phenylalanine and L-tyrosine metabolic pathways and appears potentially important (Chapter 3). In addition, all *A. pisum* host plants are from the Fabaceae, a family known for an especially high diversity of secondary metabolites, many of which are thought to have a defensive function (Wink 2013). These compounds could also function as constitutive chemical attractants or deterrents to aphid acceptance. However, plants are known to exhibit a range of induced defence signalling responses to aphid attack, including protein phosphorylation, membrane depolarisation, calcium influx and release of reactive oxygen species (ROS) (Giordanengo *et al.* 2010, Garcia-Brugger *et al.* 2006). It may be expected that after aphid exposure, plants would become primed to resist further attack as observed when plants have been exposed to other insect species (Walters 2011).

Despite the evidence of aphid induction of plant defence signalling pathways, aphids are considered to elicit comparatively few defence responses from plants compared to chewing phytophagus insects. This means that aphids are likely to possess mechanisms that lead to suppression of plant responses (Giordanengo *et al.* 2010, Schwartzberg *et al.* 2011, Ali *et al.* 2014). This phenomenon was demonstrated by Ali *et al.* (2014) on milk weed (*Asclepias syriaca)* where monarch butterfly (*Danaus plexippus)* caterpillar growth rates benefited from defence attenuation by *Aphis nerii* aphids, while *A. nerii* growth rate was impaired by *D. plexippus* feeding. Schwartzberg (2011) also showed that *A. pisum* was able to

suppress the release of defensive volatiles in *V. faba*, which are normally induced by chewing armyworm caterpillars (*Spodoptera exigua)*. Indeed, aphids are known to be able to subvert induction of various plant defences (Will and van Bel 2008, Tjallingii 2006, Will *et al.* 2007, Hogenhout and Bos 2011, Will *et al.* 2013), including manipulation of plant metabolic pathways associated with insect defence (Gao *et al.* 2008, Schwartzberg and Tumlinson 2014). Consequently, it might be expected that *A. pisum* attack may facilitate subsequent feeding by other aphids.

While it is evident that aphids possess various mechanisms to distinguish between host plants, as well as tolerate and manipulate their host's chemistry, our understanding of how this influences divergent host selection is less clear. The objective of this study was to explore whether divergent host plant acceptance by two *A. pisum* clones can be related to constitutive plant compounds, induced plant defences, or plant defence suppression. We also aimed to determine whether pre-identified plant metabolites changed in response to aphid feeding. This was investigated using two approaches. The first was to record the change in acceptance behaviour of two host specific *A. pisum* clones on *M. sativa* and *T. pratense* using a cross-comparison study of host plants pre-exposed to two host-specific aphid clones (one native on *T. pratense,* the other native on *M. sativa*). This approach aimed to reveal whether differences in acceptance by two divergent aphid clones on native and non-native plants were indicative of i) constitutive plant defences or cues, which would result in fixed acceptance or rejection behaviours, ii) induction of plant defences which would result in increased aphid rejection after pre-exposure or iii) aphid abilities to subvert plant defences which would result in facilitation of other aphids to accept plants. The second approach used a targeted metabolomics assay to compare *M. sativa* and *T. pratense* control plants and plants pre-exposed to the same two aphid clones. The compounds targeted were those previously associated with divergent host plant acceptance in Chapter 3.

It was expected that the effect of pre-exposure would be divergent between aphid host races. We hypothesizes that pre-exposure to adapted aphids would increase acceptance by other aphids due to subversion of plant defences while host plant

exposure to non-adapted aphids would decrease acceptance and/or increase rejection by other aphids. In addition it was expected that any induction or facilitation effects may be dependent on which subsequent aphid clone fed next from the plant (i.e. a native aphid clone may facilitate only its own clone type while a non-native aphid may induce defences that affect only its own clone type). If patterns in induced and non-induced resistant factors are different between different plant aphid combinations, then it is likely there are divergent mechanisms underlining aphid acceptance of plants, resulting in a potential source of divergent selection. Consequently, any observed divergence between aphid clones in the effect of pre-exposure on subsequent aphid feeding responses, or host plant chemistry, would give an insight into the underlying mechanisms behind divergent host plant selection and specialisation in *A. pisum*.

4.3 Methods

4.3.1 Plant culture

In this experiment we used two plant species, *M. sativa* and *T. pratense* (details of seed sources can be found in Appendix 2). Plants were propagated using the methods described in Chapter 2, except plants used for experimentation were four weeks old and were fed only once in the third week with 40% of full strength Rorison's solution (Appendix 3), diluted with distilled water.

4.3.2 Aphid culture

Two *A. pisum* clones were used in this study, one was an *M. sativa* specialist LSR1 (International Aphid Genomics Consortium 2010) and the other was a *T. pratense* specialist L7Tp-232 (source SE France, supplied by JC Simon, INRA, Rennes). The clones are referred to as MS_LSR1 aphids and TP_232 aphids, respectively. All aphids in this study were taken from age-controlled populations created by the methods described in Chapter 2. In brief, this protocol involved inoculating *Vicia faba* plants (the universal *A. pisum* host plant) with adults for 24 hours, then using the resulting progeny for experimentation at 10 days post-larviposition. The aphids

were maintained on *V. faba* plants in a climate controlled cabinet with 16 hour day length at temperatures of 20^0C during the day and 15^0C during the night.

4.3.3 Plant exposure to aphids

Plants were exposed to aphids by placing fifteen adult aphids into small fine-meshed bags using a fine-haired paint brush. Bags were then placed over the first fully developed leaf of each plant, giving the aphids free access to preferred feeding sites on the leaf. Mesh bags were then secured around the petiole with wire and each exposed plant was sealed within a culture pot.The control plants were treated identically, but without the aphid inoculum in the mesh bag. The exposed and control plants were left for 24 hours in a climate controlled cabinet, with 16 hour day length at temperatures of 20^0C during the day and 15^0C during the night. All aphids, including any resulting offspring, were then removed from the plants. It is known that in at least some plant species herbivore-induced plant volatiles (HIPVs) released by defence primed plants can influence neighbouring plant (Kessler et al. 2006, Frost et al 2008a), though this effect may only be effective at higher concentration and over limited distances (i.e. within plant signalling) (Frost et al. 2008b). As a result plants were arranged into separate trays by their plant treatment combinations within the well ventilated climate controlled cabinet room. However, as culture pots were vented it is possible volatiles could pass between plants to influence host defences.

4.3.4 EPG profiling of plants exposed to aphid clones

Electrical penetration graph (EPG) profiling (Tjallingii 1978, Chapter 2) was used to measure whether *M. sativa* and *T. pratense* pre-exposure to MS_LSR1 and TP_232 aphids affected the level of acceptance exhibited by a second cohort of 10 day old aphids (*naive aphids*) of MS_LSR1 and TP_232 clones. There were three treatments on both *M. sativa* and *T. pratense*: i) plants exposed to 15 individual MS_LSR1 aphids, ii) plants exposed to 15 individual TP_232 aphids and iii) unexposed control plants.

EPG was performed on the naive aphids using the protocol described in Chapter 2. Each plant was tested with either a naive MS_LSR1 aphid or a naive TP_232 aphid 24 hours after pre-exposure. It was possible to test only eight aphids for each run of EPG. On every day of testing the EPG was run twice, once in the morning and once in the afternoon. Aphid treated plants were always placed next to a control plant of the same species and tested with the same aphid clone, thus, there were twice as many control plants as plants exposed to aphids. This setup allowed four of the eight possible treatment combinations (two plant species tested x two aphid exposure treatments x two naive aphid clones) to be run simultaneously in each EPG recording session and all eight combinations to be run in one day. To reduce any bias incurred between the morning and afternoon EPG recordings treatment combinations were blocked over 12 days (Appendix 19), so that each plant-aphid combination was used equally between the morning and afternoon EPG sessions. The number of successful EPG recordings for each tested plant species/aphid exposure treatment/naive aphid combination is shown in Table 4.1.

Table 4.1: Number of successful EPG recordings for each plant species-aphid treatment-aphid clone combination.

	M. sativa plants		*T. pratense* plants	
Treament	MS aphid tested	TP aphid tested	MS aphid tested	TP aphid tested
Control	21	24	24	20
MS aphid exposed	12	12	12	9
Tp aphid exposed	10	11	12	11

4.3.5 Data processing and analysis of EPG recordings

EPG recordings were processed using the protocols described in Chapter 2. In brief, EPG recordings were interpreted manually using the Stylet+a software (www.epgsystems.eu/downloads) into waveform signals. Then the waveform signals were processed using the Excel macro, developed by Sarria *et al.* (2009), to convert

them into waveform measurements which were cleaned of uninformative variables with 50% or more of values equal to zero, or missing. Remaining missing values were imputed using the RFimpute()function from the *randomForest* R package Liaw and Wiener (2002). The waveform recordings were then further cleaned by removing one variable from every pair of waveform measurements that were correlated with r > 0.80. As a result 60 remaining waveform measurements (Appendix 7) were used for PCA with the princomp() function in base R .

The first PCA component (PC1) was used to give an overall measure of aphid acceptance. Previously it was shown that PC1 scores of cleaned EPG waveform measurements are reliable indicators of aphid acceptance (Chapter 2). The PCA in this study was fitted in the reverse direction, so to make interpretation easier the values of PC1 scores had their sign changed, thus the positive values of PC1 score represented acceptance and negative values represented rejection by aphids.

In addition, the individual waveform measurement of 'total duration the aphid stylet spends feeding within the phloem' (duration of E2) was used to provide information about the process of phloem acceptance and feeding (Chapter 2). The individual waveform measurement of the 'number of times an aphid probed a plant' (number of probes) was used as a measure of early plant rejection as frequent probing correlates with rejection (Chapter 2). These waveform measurements were selected as they represented clearly separate stages of acceptance (probing earlier probing behaviours and E2 later probing behaviours) and were likely to be influenced by independent plant defence mechanisms (Tjallingii & Esch 1993, Tjallingii 2006, Schwarzkopf *et al.* 2013). The effect of aphid pre-treatment, which naive aphid was used for testing and their interaction was analysed for each of the two plant species using linear mixed models (LMM) for PC1 scores and probing data, and generalised linear mixed models (GLMM) for the E2 data, with a tweedie distribution. Mixed models were performed using lmer() and glmer() functions from the 'lme4' R package (Bates *et al.* 2005). To perform GLMM with a tweedie distribution also required the tweedie r package (Dunn 2014). Data transformation was performed before LMM and GLMM analysis to normalise data and control heteroscedastic variance by taking the natural

log of PC1 scores and square root of probing data. To control for the among-block variance, the day treatments were run was considered a random factor in the model. Also the effect of taking measurements in the morning and afternoon was treated as a fixed factor, but found to have no significant effect and excluded from the final analyses (Appendix 20).

Hypothesis testing was then carried out by performing planned contrasts between lmm and glmm coefficients using the glht () function from the 'multcomp' R package (Horthorn *et al*. 2008). In this function, Z-tests used model error (not sample error) to estimate the significance of pre-selected comparisons from the variance-covariance matrix of the lmm and glmm model parameters. Then false discovery rate (FDR) *P*-value correction was used to account for the number of contrasts made. In this study five separate contrasts were tested to answer specific questions:

i. whether acceptance differences exist between aphid clones without pre-treatment of plants (the difference in acceptance responses between MS_LSR1 or TP_232 aphids on control plants)

ii. whether aphid pre-treatment had an effect on aphid acceptance (the difference between control plants and aphid pre-treated plants)

iii. whether the aphids used in plant pre-treatment had an effect upon aphid acceptance (the difference between aphid acceptance on MS_LSR1 and TP_232 pre-treated plants)

iv. whether the effect of pre-treatment on aphid acceptance differed between testing with MS_LSR1 and TP_232 naive aphids (the interaction between contrast (ii) and naive aphid clone)

v. whether the effect of the aphids use in pre-treatment on aphid acceptance differed between testing with naive MS_LSR1 and TP_232 aphids (the interaction between contrast (iii) and naïve aphid clone).

These five questions were tested separately for *M. sativa* and *T. pratense* by the acceptance or rejection of the following five null hypotheses (H_0):

1st. H_0= there is no significant difference in aphid acceptance between naive MS_LSR1 or TP_232 .

2nd. H_0= there is no significant difference in acceptance by naive aphids between controls and pre-treated plants.

3rd. H_0= there is no significant difference in naive aphid response between plants pre-treated with MS_LSR1 aphids and plants pre-treated with TP_232 aphids.

4th. H_0= the difference between control vs. treated plants is the same for naive MS_LSR1 and TP-232 aphids.

5th. H_0= the difference in acceptance between MS_LSR1 and TP-232 treated plants is the same for naive MS_LSR1 and TP-232 aphids.

4.3.6 Metabolic changes on plants exposed to two aphid clones

An additional objective of this experiment was to test whether the metabolites that were associated with aphid acceptance in Chapter 3 changed in concentration after plants were exposed to aphids. A fresh cohort of plants was exposed to aphids with the method described above so that the metabolites in leaves of exposed and control plants could be compared. Leaf samples were cut from each plant with a scalpel, weighed and placed in an Eppendorf tube which was immersed in liquid nitrogen. Frozen leaves were stored at -80˚C before extraction. Metabolites were extracted from the leaf material using the cold methanol-water-chloroform extraction method (Field and Lake 2011), which allows both polar and non-polar metabolites to be extracted. All extracts were stored at -80˚C before the analysis.

Each sample was diluted with methanol according to the original leaf weight to account for leaf size variation. Then samples were diluted 1:10 sample to methanol and analysed using liquid chromatography/electrospray ionisation time-of-flight mass spectrometry (ESI-LC-TOF MS) (see Appendix 21 for full instrument

settings). The number of replicates for each plant- aphid exposure combination is given in Table 4.2. Metabolite profiles for each plant were created by binning the crude m/z values into 0.2-unit bins and then calculating the sum of the percentage total ion count abundance values [% total ion count (%TIC)] across each bin (Field and Lake 2011).

Table 4.2: Sample size of each plant species - aphid exposure treatment combination for metabolomic profiles

	M. sativa plants		*T. pratense* plants	
Treament	Polar	Non polar	Polar	Non polar
Control	9	11	10	10
MS aphid exposed	9	10	13	14
Tp aphid exposed	7	9	9	10

4.3.7 Targeted analysis of plant metabolic profiles

The targeted masses, defined as the 0.2 bin of their mass to charge ratio (m/z bin values), were those previously identified in Chapter 3. These target m/z bins were: 182, 166, 183, 269, 291, 292 and 285. It was hypothesised that target masses would diverge from the control plants in relative intensity, as measured by their percentage ion count (%TIC) following aphid exposure. Change in %TIC of m/z bins between exposure treatments for each plant species was tested using generalised linear models (GLM). The differences between the treatments and the controls were tested using Tukey's honest significance test.

4.4 Results

4.4.1 Number of probes

The probing data indicated markedly different patterns compared to the E2 data (Figure 2). Firstly, on *M. sativa* controls there were significantly more probing events performed by TP_232 aphids than MS_LSR1 aphids (1st H_0 for *M. sativa*

rejected: P< 0.001) (Figure 4.1). However, there was no difference between the two aphid clones on the control *T. pratense* plants (1st H_0 for *T. pratense* not rejected: *P* = 0.63) (figure 2B and Table 4.3). Secondly, on neither host plant was there a significant difference in the number of probes, either between control and aphid pre-exposed plants, or between pre-exposed plants, and this was irrespective of the naive aphid clone used for testing (see Table 4.3). These results suggest that only on *M. sativa* plants did the two naive aphid clones express any difference in the number of probes made and probing was not affected by aphid pre-exposure treatment.

4.4.2 Phloem feeding behaviour as measured by waveform E2

E2 durations on *M. sativa* control plants were significantly higher for naive MS-LSR1 aphids (1st H_0 for *M. sativa* rejected, *P* < 0.001) than for naive TP-232 aphids, with the reverse acceptance pattern observed on *T. pratense* (1st H_0 for *T. pratense* rejected: *P* < 0.001), as expected (Figure 4.2 and Table 4.3).

On *M. sativa* there was no significant difference in aphid E2 durations between control versus pre-exposed plants (2nd H_0 for *M. sativa* rejected: *P* = 0.63), but the E2 scores for between pre-exposure treatments were not the same (3rd H_0 for *M. sativa* rejected: *P* < 0.001). However, this effect was not particularly large (Figure 4.2A and Table 4.3). For *M. sativa* plants there was no significant interaction between control versus pre-treated plants and the aphid used for testing (4th H_0 for *M. sativa* not rejected: *P* = 0.51). There was a significant interaction between pre-treatments and naive aphid clones (5th H_0 for *M. sativa* rejected: *P* = 0.017). It appeared that exposure of *M. sativa* plants to MS_LSR1 aphids increased E2 durations for both naive aphid clones (Figure 4.2A and Table 4.3). In addition, TP_232 aphid clones appeared to benefit more than MS_LSR1 from plant pre-exposure to aphids (Figure 4.2A).

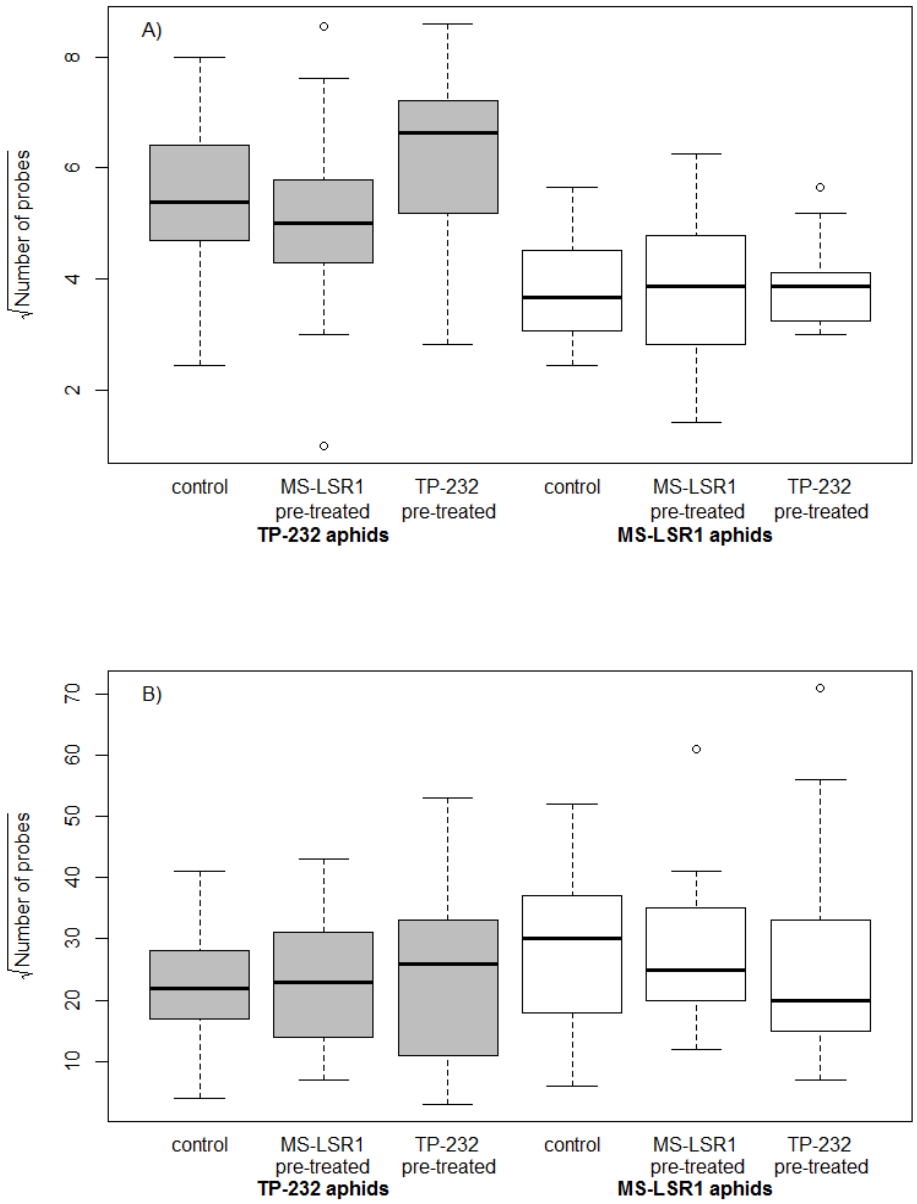


Figure 4.1: Square root of the number of probes made by aphids on control unexposed plants and plants pre-exposed to MS_LSR1 or TP_232 aphids. Comparisons are for aphid responses on A) *M. sativa* and B) *T. pratense*. Grey bars represent responses of naive TP_232 aphids and white bars represent responses of naive MS_LSR1 aphids. Median ± 0.25 and 0.75 quartiles. Outliers > ± 1.5*IQR.

Table 4.3: Plant contrasts statistics for LMM and GLM models

	Duration of E2				$\sqrt{Number\ of\ probes}$				Log PC1 scores			
Contrast tested	Estimate	Sd.	Z	P	Estimate	S.E.	Z	P	Estimate	S.E	Z	P
***M. sativa* plants**												
Difference between aphid tested on control plants	**3.42**	**0.44**	**7.70**	**>0.001**	**-1.77**	**0.40**	**-4.49**	**>0.001**	**1.98**	**0.32**	**6.14**	**>0.001**
Control vs. treatment difference	0.30	0.64	0.47	0.63	0.63	0.57	1.14	0.32	0.08	0.46	2.12	0.46
MS vs. TP treatment Difference	**1.83**	**0.47**	**3.92**	**>0.001**	-0.67	0.40	-1.73	0.21	-0.4	0.33	0.23	0.81
Control vs. treatment difference x aphid tested	-1.06	1.28	-0.83	0.51	-0.05	1.11	-0.04	0.96	-0.39	0.92	-0.43	0.81
MS vs. TP treatment difference x aphid tested	**3.42**	**0.44**	**7.706**	**0.017**	1.21	0.79	1.53	0.21	0.71	0.66	1.08	0.46
***T. pratense* plants**												
Difference between aphids tested on control plants	**-2.28**	**0.56**	**-4.048**	**>0.001**	0.59	0.39	1.53	0.63	0.32	0.34	-0.93	0.58
Control vs. treatment difference	**2.40**	**0.70**	**3.45**	**0.001**	0.33	0.5	0.60	0.89	**2.007**	**0.49**	**4.1**	**>0.001**
MS vs. TP treatment Difference	-0.23	0.43	-0.538	0.59	-0.001	0.39	-0.01	0.99	-0.74	0.347	-2.13	0.08
Control vs. treatment difference x aphid tested	2.97	1.40	2.12	0.056	0.28	0.78	0.36	0.90	-0.51	0.98	0.47	0.64
MS vs. TP treatment difference x aphid tested	0.56	0.83	0.67	0.59	0.5	0.39	1.23	0.3	0.32	0.34	-0.93	0.64

On *T. pratense* there was a highly significant difference in E2 duration between controls and pre-exposed *T. pratense* plants (2^{nd} H_0 for *T. pratense* rejected: $P < 0.001$). However, there was no such difference between the pre-exposure treatments (3^{rd} H_0 for *T. pratense* not rejected: $P = 0.59$) (Figure 4.2B and table 4.3). In addition on *T. pratense* plants there was no difference in acceptance between which naive aphid clones was tested, either in the control versus pre-treated comparison (H_0 for 4^{th} H_0 for *T. pratense* not rejected: $P = 0.056$), or between pre-treatment comparison (5^{th} H_0 for *T. pratense* not rejected: $P = 0.59$). This suggests that for *T. pratense* overall aphid exposure increased E2 durations, irrespective of which aphid clones were used for pre-exposure or testing (Figure 4.2B and Table 4.3). However, it should be noted that for naive MS_LSR1 aphids increases in E2 durations were only seen in a relatively small number of individual aphids (Figure 4.2B).

4.4.3 PC1 acceptance behaviour

Initial analysis of log-transformed PC1 scores (log(PC1)) also suggested significantly larger PC1 scores for MS_LSR1 aphid clones than for TP_232 on *M. sativa* controls (1^{st} H_0 for *M. sativa* rejected: $P < 0.001$) (Figure 4.3A and Table 4.3). However, there was no such difference between aphid clones on *T. pratense* control plants (1^{st} H_0 for *T. pratense* not rejected: $P = 0.58$) (Figure 4.3B and Table 4.3). Results suggested that only on *T. pratense* plants was there a significant difference in log(PC1) scores between the control and aphid pre-exposed plants (3^{rd} H_0 for rejected: P=0.001) (Figure 4.3B and Table 4.3).

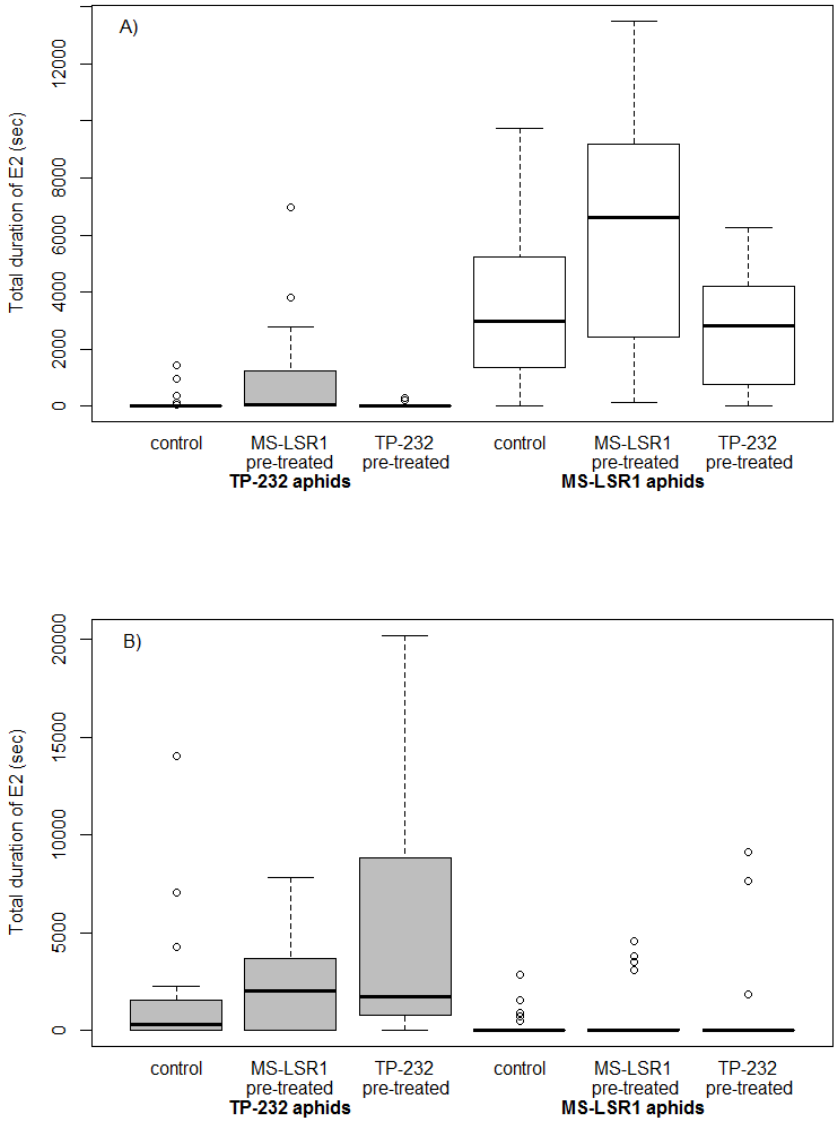


Figure 4.2: Total duration of E2 waveforms for aphids on unexposed control plants and plants pre-exposed to MS_LSR1 or TP_232 aphids. Comparisons are for aphid responses on A) *M. sativa* and B) *T. pratense*. Grey bars represent responses of naive TP_232 aphids and white bars represent responses of naive MS_LSR1 aphids. Median ± 0.25 and 0.75 quartiles. Outliers > ± 1.5*IQR.

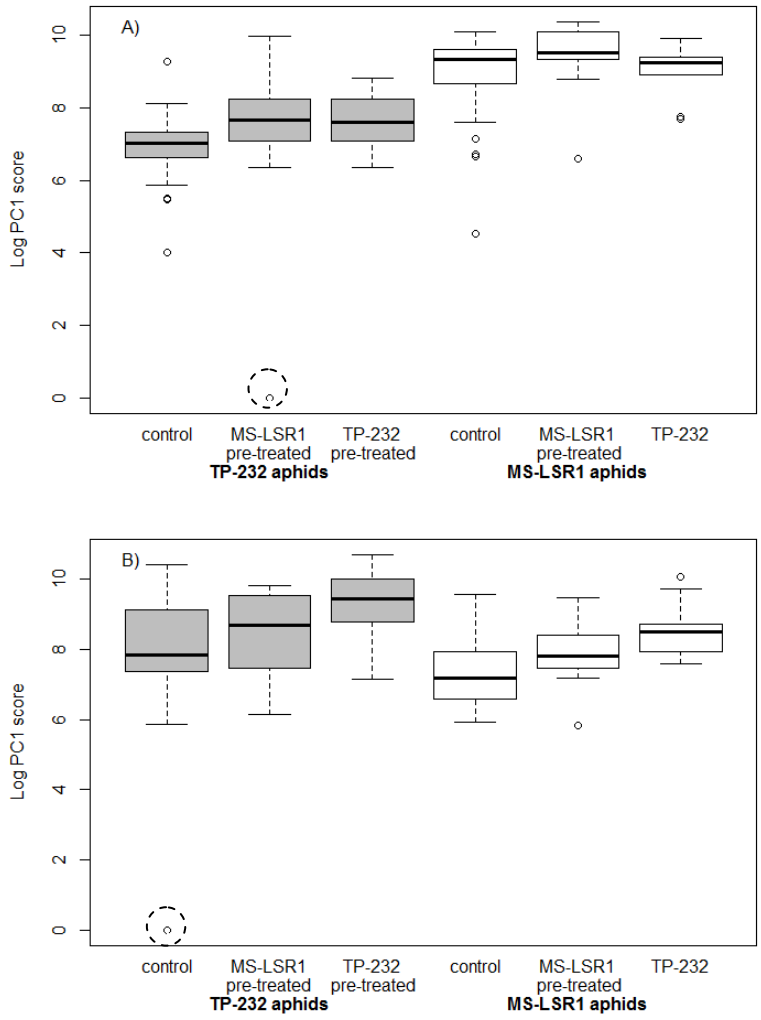


Figure 4.3: Log(PC1) scores for aphids on unexposed control plant and plants pre-exposed to MS_LSR1 and TP_232 aphids. Comparisons are for aphid responses on A) *M. sativa* and B) *T. pratense*. Grey bars represent responses of naive TP_232 aphids and white bars represent responses of naive MS_LSR1 aphids. Dotted circles represent outliers removed in later analysis. Median ± 0.25 and 0.75 quartiles. Outliers > ± 1.5*IQR.

However, inspection of the data (Figure 4.3) showed these results were likely to have been influenced by two outlying data points: one TP_232 aphid response to *T. pratense* control plants and one MS_LSR1 aphid response to *M. sativa* pre-exposed to MS_LSR1 aphid (indicated by a dotted circle in Figure 4.3). Removal of these data points (Figure 4.4) changed the significance for log(PC1) scores markedly (Table 4.4). With the outliers removed from the data there was a significant difference in log(PC1) scores between naive MS_LSR1 and TP_232 aphid

responses on the control plants of both plant species (1st H_0 for *M. sativa* rejected: $P < 0.001$, 1st H_0 for *T. pratense* rejected: $P < 0.035$).

On *M. sativa* plants, outlier removal increased the significance of the difference between control and pre-exposed aphid plants (2nd H_0 for *M. sativa* rejected: $P=0.002$). However, there was no significant difference in log(PC1) scores between pre-exposed plants (3rd H_0 for *M. sativa* not rejected*: P* = 0.27). For *T. pratense* plants there was no difference between the naive aphid clones tested either in the control versus pre-treated comparison (H_0 for 4th H_0 for *M. sativa* not rejected: P = 0.27) or between the two aphid pre-treatment (5th H_0 for *M. sativa* not rejected: P = 0.7). This suggests that aphids caused a small increase in log (PC1) score on *M. sativa* but that this was neither dependent on the aphid clone *M. sativa* was pre-exposed to, nor the naive aphid clone tested.

Outlier removal also changed test significance values for *T. pratense* plants (Figure 4.4 Table 4.4). There was a significance in the difference in log(PC1) scores between control verses aphid pre-exposed plants (2rd H_0 for *T. pratense* rejected: $P < 0.001$) and between TP_232 and MS_LSR1 pre-exposed plants (3rd H_0 for *T. pratense* rejected: $P=0.027$) (Table 4.4). However, these changes in response to pre-exposure were still the same for both naive aphid clones on controls verses pre-exposure treatment comparison (4th H_0 for *T. pratense* not rejected: $P= 0.9$) and between pre-exposure treatment comparison (5th H_0 for *T. pratense* not rejected: P = 0.7). This suggests that aphid pre-exposure of *T. pratense* plants increased aphid overall acceptance and was slightly stronger when exposing with TP_232 aphids, but the increase in overall acceptance was the same for both naive TP_232 and MS-LSR1 aphids (Figure 4.4).

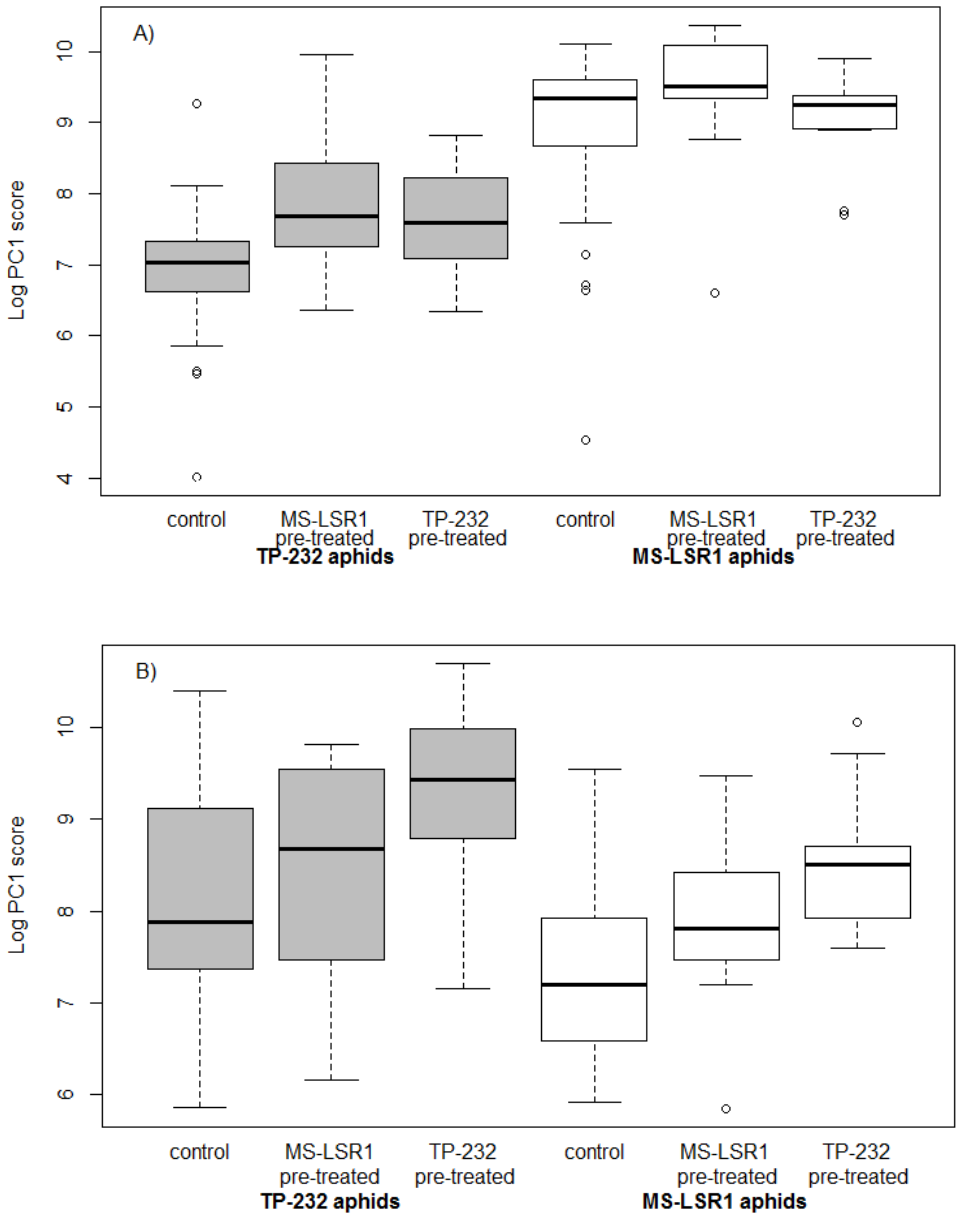


Figure 4.4: Log PC1 scores with outliers removed for aphids on unexposed control plant and plants pre-exposed to MS_LSR1 or TP_232 aphids. Comparisons are for aphid responses on A) *M. sativa* and B) *T. pratense*. Grey bars represent responses of naive TP_232 aphids and white bars represent responses of naive MS_LSR1 aphids. Median ± 0.25 and 0.75 quartiles. Outliers > ± 1.5*IQR.

Table 4.4: Test statistic for LMM and GLM model contrasts for log PC1 scores with outliers removed

Contrast tested	Log PC1 scores (outlier removed)			
M. sativa	Estimate	S.E	Z	*P*
Difference between aphids tested on control plants	**-1.99**	**0.26**	**-7.59**	**<0.001**
Control vs. treatment difference	**1.2420**	**0.3754**	**3.308**	**0.002**
MS vs. TP treatment difference	0.9242	0.7509	1.231	0.27
Control vs. treatment difference x aphid tested	0.3408	0.2693	1.266	0.27
MS vs. TP treatment difference x aphid tested	-0.1907	0.5386	-0.354	0.72
T. pratense Difference between aphids tested on control plants	**0.63**	**0.27**	**2.301**	**0.035**
Control vs. treatment difference	**1.69**	**0.39**	**4.338**	**<0.001**
MS vs. TP treatment difference	**-0.71**	**0.27749**	**-2.567**	**0.027**
Control vs. treatment difference x aphid tested	-0.01	0.78189	-0.127	0.9
MS vs. TP treatment difference x aphid tested	-0.32	0.55581	- 0.584	0.7

4.4. 4 Effect of exposure on targeted metabolites

For five of the seven target m/z bin values identified in Chapter 3 there was no significant change in %TIC, for either polar or non-polar samples. M/z 183 was not detected at all. A small change in m/z bin 182 %TIC was detected in aphid exposed *M. sativa* plants (Figure 4.5). This effect was significant in the uncorrected analysis ($F=3.82$, df =2, $P = 0.038$), however, FDR correction for the seven comparisons removed significance ($P = 0.263$). Although there is a risk that this result may be a false positive, it is likely some of the m/z bins investigated here, such m/z bin 182 and m/z bin 166, are non-independent and highly correlated; thus, FDR correction may be too conservative. In addition, Chapter 3 showed that m/z bin 182 was also

the most important compound in polar samples for prediction of host discrimination, suggesting that this result is worth investigating further.

Post-hoc analyses of m/z182 using Tukey's honest significance test showed that the main difference was between MS_LSR1 and TP_232 treated *M. sativa* plants (Estimate = -0.32, Estimate error=0.12, z= 2.53, *P*=0.034). There was a relatively small difference between TP_232 treated and control *M. sativa* plants (Estimate = -0.26, Estimate error=0.13, z= 2.04, *P*=0.062) and no difference between MS_LSR1 exposed and control plants (Estimate = 0.05, Estimate error=0.85, z= -0.63, *P*=0.53). This suggests that TP_232 aphids may be causing a reduction in the concentrations of m/z 182 in *M. sativa*, while at the same time MS_LSR1 aphids may be causing a very small increase in m/z 182 concentrations in *M. sativa* .

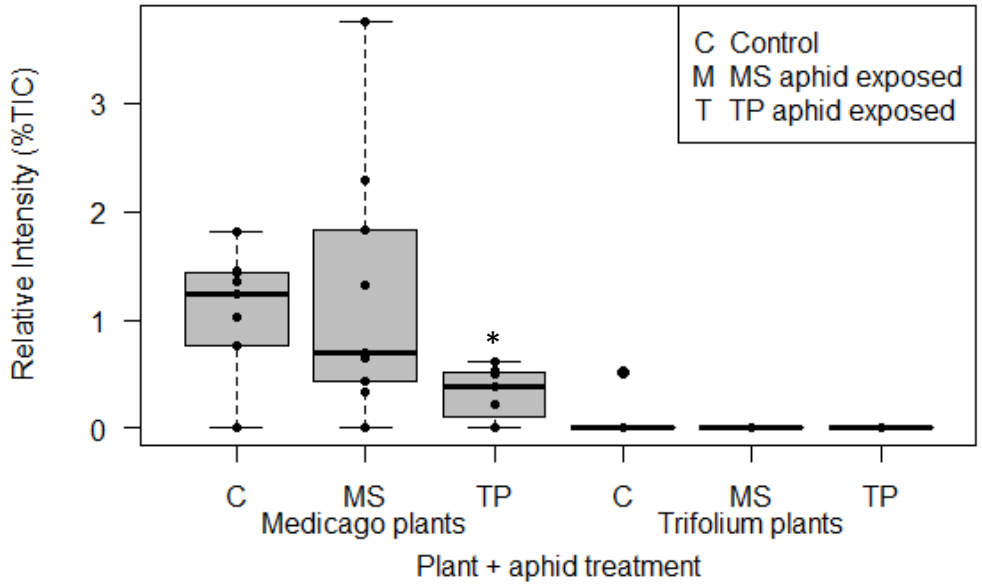


Figure 4.5: Box and whisker plot of the change in relative intensity [%total ion count (%TIC)] in polar samples of m/z bin 182 between control and treated* M. sativa *and T. pratense plants. Significance was tested between control plants and the individual aphid treated plant using one sided Student's t-test. *>0.01 . ± 0.25 and 0.75 inter quartile range. Outliers = ± 1.5*IQR.

4.5 Discussion

In Chapter 3 a chemical signature for divergent *A. pisum* acceptance was identified of which two compounds, phenylalanine and L-tyrosine appeared to be linked in the same plant metabolic pathway. In this study two divergent clones were used as bioassays to test if plant responses to aphids are fixed, and so due to constitutive deterrents or attractants, or change and so are the result of induced or suppressed plant defences. In addition change in specific plant compounds were also studied.

4.5.1 Divergent acceptance towards constitutive, induced and subverted plant features.

This study used aphid responses, as measured by EPG, to plants pre-exposed and unexposed to aphids, as a bioassay to test plant responses towards aphids attack. Probing is a very early stage of the acceptance process and, as the majority of probing events last only a few seconds, the main interaction is likely to be with cells at, or near, the leaf surface (Tjallingii & Esch 1993, Schwarzkopf *et al.* 2013). E2 waveforms measure phloem feeding so represent the last stage of acceptance decisions, where the main interaction is between aphid and phloem (Tjallingii & Esch 1993, Tjallingii 2006). PC1 is a holistic measure of acceptance and shown in Chapter 2 to be a good indicator of aphid more general behavioural response to plants.

The aim of this study was to ask if difference in aphid plant responses to plant are fixed (constitutive) between plants or changed by plant exposure to aphid. In addition to ask if any pre-exposure effect is specific to which clone is used for exposure (i.e. native vs. non-native aphid) and also if aphid exposure effects on to acceptance aphid specific to a which naive aphid subsequently fed.

Analysis of EPG data reveals three significant aspects of divergent host plant responses to pre-infestation by the two aphids used in this study. These are that: i) constitutive features of *M. sativa* not present in *T. pratense* caused divergent

acceptance by aphids, ii) on *M. sativa* plants there was a divergent effect of aphid clones to either increase or decrease plant deterrence toward aphids during, or just before, phloem feeding and iii) on *T. pratense* aphids had divergent ability to decrease plant deterrence towards aphids at some point between the probing and phloem stage. These results point to the occurrence of both constitutive and induced/repressed plant responses, occurring at different stages in the plant that differ between plant aphid combinations.

The occurrence of a divergent aphid-host interactions toward differing multilayered host plant defence systems could be a source of the divergent selection in *A. pisum* necessary for specialism and host divergences. In this study there is evidence for both divergent i) constitutive difference between plants and ii) differences in induced or suppressed plant responses to aphids:

i) Constitutive differences between control plants

From previous research (Chapter 2) it was expected that MS_LSR1 and TP_232 aphid clones would accept more their native control host plant than the alternative control non-host plant in this experiment. On *M. sativa,* this was found to be the case across all acceptance measurements (duration of E2, number of probes and PC1), which supports the expectation that the two aphid clones have divergent acceptance at early and late stages of stylet probing on this species. As there was also no influence of aphid pre-exposure on the number of probes naive aphids subsequently made on either of the host plants, it is likely that the early barriers to TP_232 aphids accepting *M. sativa* plants are constitutive.

In contrast, on *T. pratense* plants, there was only a significant divergence between the two aphids in terms of E2 scores, with weaker separation in the PC1 scores and no separation at all for number of probing events. This suggests that only in the later stages in probing did the two aphid clones make divergent choices to accept or reject *T. pratense* and that, during the earlier stages of plant probing, aphids respond similarly to constitutive characteristics of *T. pratense*. This difference in *A. pisum* probing between *M. sativa* and *T. pratense* suggests *M. sativa* plants possess

constitutive plant factors encountered in early probing that elicit a positive response from MS_LSR1 and deter TP_232 aphid clones, which are not found in *T. pratense* plants. This evidence suggest the importance of the chemical cue or chemical deterrents that influence aphid probing is different between the host plant used here, with selection by constitutive attractant or deterrents being much stronger *M. sativa* plants than *T. pratense* plants.

ii) Differences in induced or suppressed plant responses to aphids

Evidence of induction or subversion of plant responses by aphid in order to improve or inhabit acceptance or performance aphids have been previously reported in the literature (e.g. Takemoto *et al.* 2013, discussed below in more detail), mean it might be expected that similar patterns would be found in this study (e.g. by subverting chemical defences or increasing attractant chemical cues).Because of divergent ecological selection between the aphid clones, it might be expected that clones would better subvert defences on their own host plant and would be more likely to invoke a defence response from their non-host. If neither of these two pre-exposure effects were observed, it would have suggested that differences in acceptance between MS_LSR1 and TP_232 aphid clones are the result of constitutive features of plants only.

Changes to aphid acceptance were observed in response to pre-exposure of host plants to aphids. For non-adapted TP_232 clones on *M. sativa*, the effect of pre-exposure by native MS_LSR1 clones appeared to increase phloem feeding durations, while pre-exposure by non-native TP_232 clone had no effect or may have even reduced feeding durations. It appears that on *M. sativa,* MS_SLR1 clones subverted plant defences and so improve the feeding success of other aphids while TP_232 clones induced defences with the reverse or no effect. This pre-exposure facilitation effect was small and defence induction is not evident when naive MS_LSR1 clone behaviour was recorded. This could be explained if MS_LSR1 plant defence suppression was rapid, i.e. it was active within the duration of the EPG recording. In this case, plant defence subversion by naive MS-LSR1 aphids would negate the

effects of pre-exposure by other aphids. However, these results could also be interpreted as the non-adapted TP_232 clone inducing the expression of specific plant deterrents on *M. sativa,* that are only effective against its own clone. Further experiments with different exposure times could establish if this is the case. On *T. pratense,* a less specific facilitation effect on phloem feeding duration was apparent as pre-exposure to either aphid clone caused facilitation of feeding for both naive aphid clones. In addition on *T. pratense* the relatively sporadic facilitation effect that pre-exposure had for naive MS_LSR1 phloem feeding durations, suggests that constitutive defences also play a role in divergent host discrimination at the phloem stage.

PC1 scores for *M. sativa* suggest both aphid species possess traits to improve acceptance by other aphids. PC1 scores for feeding on *T. pratense* also indicated pre-exposure by either clone tended to improve acceptance by both clones, but TP_232 aphid clones produced a higher level of facilitation. Both clones seem to posses adaptations to subvert *T. pratense* plant defences, although TP_232 aphids appear more effective.

Interestingly, on *T. pratense,* there was no effect of pre-exposure on probing and no divergent effect of pre-exposure on phloem feeding. This suggests that the site of influence of TP_232 on divergent facilitation shown by PC1, is occurring at some point between the leaf epidermis and the phloem. However, caution is required because if a non-adapted aphid also rejects a host because of constitutive plant factors, it would be expected they would spend less time interacting with a plant. Consequently, a non-native aphid might not suppress a defence during pre-exposure simply because they spend less time on the plant, due to other plant deterrents. It is therefore not possible to rule out that constitutive plant factors could have caused some of the difference in log PC1 score between TP_232 and MS_LSR1 clones. However, the result here was supported by observations in Chapter2, that for *Trifolium* plant species the EPG recorded pre-phloem ‘repetitive potential drops’ behaviour was more divergent between TP and MS aphid, than other waveform

measurements. This suggests the two aphid clones have divergently adapted to suppress defences that precede phloem feeding in *T. pratense*.

Comparison of E2 and PC1 results suggest that facilitation by aphid to increase acceptance is occurring on both host plants, but that this facilitation is occurring are different sites within the leaf, and tends to be greater when a plant is exposed to an adapted aphid. However, which aphid subsequent naive aphid feeding make little difference. Taken together it is reasonable to speculate that the divergent aphid-host interactions observed here represent a divergence in aphid co-evolution to specific host multilayered plant defence systems. Such close co-evolution could provided the divergent selection between *A. pisum* races necessary for specialism and host divergences.

4.5.2 Mechanism of divergence in *A. pisum* clone responses to plant chemistry

EPG results of this study suggest three divergent effects of plant chemistry upon aphid acceptance: i) during probing stages a divergence in aphid response is caused by *M. sativa* constitutive compounds that are not present in *T. pratense*, ii) for *M. sativa* plants, during or just before phloem feeding, there is a divergence in aphid induction or suppression of defence chemicals that is most clearly revealed by its influence on TP_232 aphids, iii) on *T. pratense* there is divergent suppression of plant defences, at some point between the probing and phloem stage, that has a positive effect on subsequent feeding by either aphid race. These patterns suggest divergent evolution of both plant defences and aphid-specific adaptive traits to subvert them, i.e. the two plants vary in their defensive mechanisms and/or aphid attractants. Aphid clones, it seems, have divergently evolved in their abilities to respond to plant defences, thus creating asymmetric behavioural responses.

4.5.3 Constitutive plant compounds and *A. pisum* acceptance

In this study evidence of constitutive plant features, most likely chemical, influencing aphid behaviour, suggests aphids possess host plant-specific adaptations

to plant chemical deterrents or attractants. Results here support the proposal that chemosensory (Cr) genes, which form part of the chemical sensory system, and cytochrome P450s (P450) genes, which have a role in detoxification (Snyder and Glendinning 1996), are important in the evolution of divergent *A. pisum* biotypes in response to plant chemistry (Jaquiéry *2012 et al.,* Smadja *et al. 2012,* Duvaux *et al.* 2015). For instance, some Cr gene families have undergone a recent rapid expansion under positive selection, reflecting the relatively recent ecological divergence in *A. pisum* biotypes (Smadja *et al.* 2009). A candidate targeted screening approach, comparing three *A. pisum* biotypes adapted to *Lotus pedunculatus, M. sativa,* and *T. pratense,* showed that several targeted Cr genes had higher levels of differentiation between *A. pisum* biotypes, than would be expected by chance (Jaquiéry *et al.* 2012, Smadja *et al.* 2012). Divergence in the number of complete or partial copies of some chemosensory genes between eight *A. pisum* biotypes also suggests chemosensory proteins are expressed differently between *A. pisum* races (Duvaux *et al.* 2015). Both the genetic divergence and gene copy number variation are, in particular, evident in the odorant (Or) and gustatory receptor (Gr) gene families (Smadja 2012, Duvaux *et al.* 2015) which in this experiment could be associated with the acceptance of plants at or before the early probing stage.

There is evidence that *M. sativa* and *T. pratense* contain chemicals that stimulate early stages of aphid feeding. Del Campo *et al.* (2003) used reciprocal surface treatments of *M. sativa* and *T. pratense* leaves with extracts of the two plants and artificial diets containing *M. sativa* and *T. pratense* extracts and found divergent positive response by two aphid biotypes. However, as Del Campo *et al.* (2003) boiled leaf material to prepare the leaf extracts, it is probable that many compounds were altered, so potentially destroying any deterrent compounds. Also, it is possible that deterrent compounds became ineffectual once removed from their original background environments or are only expressed in response to aphid attack. Consequently, the role of both constitutive and induced plant deterrents in *M. sativa* and *T. pratense* cannot be ruled out by the Del. Campo *et al.* (2003) study.

When the *A. pisum* genome was compared to that of the host generalist peach-potato aphid (*Myzus persicae)*, it was found *A. pisum* had fewer P450 genes than *M. persicae* (by about 40%) and so is likely have a reduced ability to metabolise toxic compounds (Ramsey *et al.* 2010). Ramsey *et al.* (2010) suggested reduced p450 diversity in *A. pisum* is a consequence of evolving narrower host ranges, so exposing themselves to a lower diversity of plant toxins (Ramsey 2010). However, low P450 diversity within a single *A. pisum* clone is complemented by an observed divergence in copy number variation (CNV) of 34 P450 genes (of 60 P450 genes looked at) observed between eight *A. pisum* races (Duvaux *et al.* 2015). This P450 CNV is thought to relate to a gene dosage effect on the expression of specific P450 genes (Duvaux *et al.* 2015). Combined, these two studies suggest that P450 proteins of *A. pisum* races have, restricted but highly targeted, abilities to neutralise specific toxins. Divergent P450 expression could explain the divergence in acceptance of constitutive plant features seen in this experiment.

The potential function of plant allelochemicals (any compound with detrimental physiological effect on animals) is interesting as the host plants that *A. pisum* races use are all from the Fabaceae family. The Fabaceae contain a wide diversity of metabolites (Wink 2003, Wink 2013), many of which could function as constitutive defences for plants and/or as cues for *A. pisum* to discriminate between hosts. In addition, as a group of nitrogen fixing plants, the Fabaceae family contain a high number of biologically active and sometimes highly toxic nitrogen-based compounds (Wink 2003). The role of nitrogen-containing compounds in defence is supported in Chapter 3, where acceptance of *Medicago* plants was linked to L- phenylalanine and L-tyrosine pathways, which could potentially lead to many nitrogen containing plant defensive alkaloids.

4.5.4 Evidence of induction and suppression by *A. pisum*

Aphids produce many potential elicitors of plant defence pathways, principally via the egestion of gelling and watery saliva. Aphid gelling saliva contains the enzymes cellulase and pectinase and their enzyme products, such as oligogalacturonides, are known to act as elicitors of plant defences (Will and van Bel 2008). In addition, as aphid stylets pass through the apoplast to the phloem they regularly penetrate cells

and use their watery saliva to sample host cell chemistry (Tjallingii and Esch 1993). Watery saliva is a mixture of a large number of proteins and metabolites, some of which are likely to act as elicitors of plant defences (Carolan *et al.* 2011, Giordanengo *et al.* 2010). Plant wound signalling pathways, such as the JA and ET pathways, are also activated by aphid stylets penetrating cells in search of the phloem (de Ilarduya 2003, Thompson and Goggin 2006). All of the above mechanisms could explain the low level defence induction observed in this study.

There were stronger effects of aphid facilitation than plant defence induction in this study. This is not surprising as aphids are often considered to be stealthy feeders (Giordanengo *et al.* 2010, Schwartzberg *et al.*2011, Ali and Agrawal 2014). Evidence here of aphid facilitation, most likely due to defence subversion, complements the work of Takemoto *et al.* (2013) which showed performance of a single clone of *A. pisum* on *V. faba* was improved on leaves already being fed upon by *A. pisum*. Facilitation of acceptance and/or performance by other aphid individuals has also been observed in studies of other aphid species (Gonzales *et al.* 2002, Prado and Tjallingii 1997, Dugravot *et al.* 2007, Brunissen *et al.* 2009) and even between-species facilitation has been reported (Kidd *et al.* 1985, Dugravot *et al*. 2007). Facilitation may be a common phenomenon in aphids, which is not surprising considering that in natural groups aphids are likely to feed alongside individuals of an identical genotype.

Our understanding of how aphids subvert plant defences is currently poor although there has been considerable progress in recent years. For example, whilst aphid watery saliva may expose an aphid to plant defences, it can also act as a medium to allow aphids to introduce proteins that manipulate and suppress plant responses (Carolan *et al.* 2011, Sharma *et al.* 2014). In particular, effector proteins, known to modulate host plant-defence chemistry, such as M1-zinc metalloprotease and CLIP-domain serine protease, have been identified in *A. pisum* saliva (Sharma *et al.* 2014).

One of the best-documented examples of plant defence suppression by aphids is in the prevention of coagulation of phloem proteins by watery saliva within (and possibly just before entering) the phloem (Tjallinii 2006). This protein coagulation process is known as phloem occlusion and acts as a way to seal phloem vessels in the event that they become compromised. Phloem occlusion is triggered by a calcium ion (Ca^{2+}) flux across the membrane of the phloem, mediated by calcium channels in the cell membrane (Will and van Bel 2008). Fabaceae species in particular have a unique mechanism for phloem occlusion, involving the coagulation proteins known as forisomes (Will and van Bell 2008). As the aphid stylet enters the phloem it is likely to cause sufficient mechanical and chemical disturbance to trigger a Ca^{2+} flux and phloem occlusion must be counteracted if aphids are to feed. Proteins within the watery saliva are thought to act as Ca^{2+} binding proteins, such as C002 protein (Sharma *et al.* 2014). It has been shown that silencing the C002 gene in *A. pisum* causes detrimental effects on the ability of aphids to colonize *V. faba* plants (Mutti *et al.* 2008).

It is likely that some of the facilitation to feed observed in this study was due to Ca^{2+} ion channel inhibition. However, whether this ability to suppress ion channels was specific to given aphid biotype-plant species interactions, or were a general trait across all aphids is unclear. Will *et al.* (2009) found two clones of *A. pisum* and six other aphid species on *V. faba* expressed conserved behavioural responses to phloem occlusion suppression, meaning aphids to some extent display similar adaptations to phloem occlusion. However, in the same study it was also found that the saliva proteome varied greatly among the aphid types (with over 50% difference in salivary protein content between clones), including the two *A. pisum* clones (Will *et al.* 2009). Considering our current lack of understanding of the function of many aphid salivary proteins (Carolan *et al.* 2011), there is opportunity for divergent adaptation of aphids to suppress phloem occlusion. For example, some sort of lock and key mechanism between plant Ca^{+2} channels and aphid suppression proteins might be one mechanism to cause aphid-plant specific suppression of phloem occlusion. This may be something to consider for future research into aphid phloem occlusion and calcium ion inhibition.

The aphid subversion of host-plant defences observed in this study may also involve the manipulation of plant signalling pathways, specifically the jasmonic acid (JA) and salicylic acid (SA) pathways (Giordanengo *et al.* 2010). It is thought that some aphid species elicit ‘‘decoy’’ defences to divert JA-regulated defences, that reduce insect performance, into less harmful SA defences (Thompson and Goggin, 2006, Will and van Bel 2008, Gao *et al.* 2008). JA-regulated defences have been linked to increased plant deterrence and reduced population growth of aphids (Thompson and Goggin 2006, De Vos *et al.* 2007, Gao *et al.* 2007), while SA pathways are generally linked to plant defences against pathogen attack (Tanaka *et al.* 2015). Takemoto *et al.* (2013) found that in *V. faba* plants, pre-exposure to *A. pisum* aphids not only improved performance of a new *A. pisum* inoculum but also decreased JA concentrations in exposed plants. This aphid facilitation was found to be negated when plants were pre-treated with JA (Takemoto *et al.* 2013). *A. pisum* appears to be able to avoid triggering the JA pathway, and instead induce the SA pathway (Gao *et al.* 2008, Schwartzberg and Tumlinson 2014). Gao *et al.* (2008) also found that *A. pisum* infestation of aphid resistant *M. truncatula* plants did not induce large changes in the expression of genes of the octadecanoid pathway, which leads to the production of JA. This was in contrast to the response to blue-green aphid (*Acyrthosiphon kondoi)* infestation, where genes involved in the JA pathway were exclusively or predominantly induced (Gao *et al.* 2008). This shows that, at least at a species level, different aphids possess divergent abilities to divert JA metabolism and express the SA pathways instead; a possible mechanism for divergent facilitation caused by the *A. pisum* clones seen here. Interestingly, the JA associated resistance to *A. kondoi* appeared to also be associated with acceptance when the stylet is within the phloem, so JA pathway regulation could explain the divergent facilitation effect seen in E2 scores in this study (Klingler *et al.* 2005). However, it must be noted there are cases of increased SA pathway expression being linked to a decrease in aphid performance on plants (Mohase and van der Westhuizen 2002, Li *et al.*2006, Donovan *et al.*2013, Schweiger *et al.* 2014), suggesting that the roles of the JA/SA pathways are more complex and possibly specific to individual aphid-host plant interactions. Indeed, Schweiger *et al.* (2014) reported that both JA and SA pathway expression negatively affect *M. persicae* performance on *Plantago lanceolata*, and

that, only when both JA and SA pathways were over-expressed did aphid performance improve (Schweiger *et al.* 2014).

Currently, the specific mechanisms behind plant pathway manipulations are not fully understood. It could be that aphid suppression is mediated by aphid saliva. However, a study by Schwartzberg and Tumlinson (2014) shows that the honeydew of *A. pisum* was able to suppress JA in *V. faba* plants. A possible mechanism suggested by Schwartzberg and Tumlinson (2014) is that sugars contained in aphid honeydew secretion, interact with plants. However, another proposed theory is that bacteria in the honeydew elicit increased SA pathway expression in plants (Schwartzberg and Tumlinson 2014). Such bacterial elicitors are also recorded in the watery saliva and have been shown to induce plant defences (Chaudhary *et al.* 2014). This could explain the findings by Ferrari *et al.* (2007) where the composition of bacterial symbiont *Regiella insecticola*, in five *T. pratense* adapted *A. pisum* clones, caused aphid clone-host plant specific response in aphid performance. In this study *R. insecticola* reduced both performance and acceptance on *V. faba* plants but improved the performance of on *A. pisum* clone on *T. pratense* plants (Ferrari *et al.* 2007).

4.5.5 Metabolic response of L-tyrosine to TP aphids

Of the key metabolites previously identified by RF modelling in Chapter 3, only m/z bin 182 was found to be potentially different in polar samples in response to aphid treatment of plants. M/z bin 182 in *M. sativa* has previously been identified as L-tyrosine (Chapter 3). Here, there was a reduction of *M. sativa* L-tyrosine concentrations observed in plants exposed to TP_232 aphids and slight increase in plants exposed to MS_LSR1 aphids.

Previous research suggests L-tyrosine could act as an allelochemical to insects (Lokvam *et al.*2006, Lokvam *et al.*2007) and so might be able to also deter aphids. Hyper-production of L-tyrsoine in plants of the *Inga umbellifera* (Fabaceae) has

been linked to defence against insect attack (Lokvam *et al.* 2006). Lokvam *et al.* (2006) identified a dose-dependent response in larval growth by *Heliothis virescens* (Lepidoptera: Noctuidae), from both *I. umbellifera* extracts with high L-tyrosine content, as a well a purified L-tyrosine. However, if L-tyrosine in *M. sativa* was acting directly as a defence compound, it would be expected L-tyrosine would be regulated as a result of MS_LSR1 aphid defence suppression and up regulated in response to TP_232 aphid defence induction. Instead the opposite trend was observed, suggesting that changes in L-tyrosine concentration may be the result of aphid elicitors or effectors acting upon L-tyrosine metabolic pathways.

There are two alternative mechanisms that could explain the reduced L-tyrosine in TP_232 exposed plants. Firstly, *M. sativa* plants maintain a concentrated pool of L-tyrosine, which is metabolised when downstream plant pathways are activated by aphids. Secondly, pathways that produce L-tyrosine are being switched off in response to aphids. In support of the first mechanism a study by Facchini *et al.* (1999) on opium poppy (*Papaver somniferum*) plants with an induced increase in expression of L-tyrosine decarboxylase (TyDc), caused a 30% reduction in cellular pools of L-tyrosine compared to wild type plants. This L-tyrosine reduction was complemented by a two-fold increase in cell wall bound tyramine (Facchini *et al.* 1999). This is in line with the results in this study from the tandem MS experiments showing that tyramine was present in some of the plant tissues probed by the aphids (Appendix 22). TyDc metabolism of L-tyrosine to tyramine in cell walls is linked to the production in the cell wall of hydroxycinnamic acid amides, which are known to improve plant resistance to both fungal and bacterial infestations (Newman *et al.* 2001, Zacrés *et al.* 2007, Macoy *et al.* 2015). It is thought that hydroxycinnamic acid amides form a physical barrier to pathogens and improve cell wall resistance to degradation by pathogen enzymes (Macoy *et al.* 2015). Intriguingly, these defences seem to be located only in the plant cell wall (Kulma and Szopa 2007), exactly at the site at which aphids are in contact with plants during host acceptance.

Production of tyramine by L-tyrosine decarboxylase in winter triticale (x *Triticosecale*) has been shown to lessen plant acceptance by grain aphids (*Sitobion*

avenae) (Sempruch *et al.* 2009). Further work by Sempruch *et al.* (2013) showed increased TyDc expression in pea (*Pisum sativa)* plants after 24 hours and seven days of plant exposure to *A. pisum*. This indicates that, in at least some host plant species, *A. pisum* can trigger whole plant defence priming, via TyDc expression, to cause L-tyrosine consumption. Sempruch *et al.* (2013) suggest this aids the production of defensive alkaloids. However, it is possible that TyDc expression is actually causing metabolic responses that increase the resistance of cell walls to aphid penetration.

As previously known plant responses to pathogens a response to aphid attack, such as TyDc expression and the production of cell wall structures, is not necessarily a surprising. Firstly, it has been shown that aphid saliva contains at least one bacterial protein, chaperonin GroEL that originates from their obligate symbiont *Buchnera aphidicola* (Chaudhary *et al.* 2014). GroEL not only elicits pathogen associated plant defences but, interestingly, such plant responses reduced the fecundity of several aphid species (Chaudhary *et al.* 2014). In addition, aphids are often described as inducing '*pathogen like'* metabolic responses in plants (Sharma *et al.* 2014) because their damage to plants is much more akin to fungal infection than to chewing insects. This includes the syringe like damage to plant cells (Tjallingii & Esch 1993), similar to the way fungal hyphae penetrate plant cell walls. Also aphid salivary enzymes, such as pectinase and polygalacturonase, dissolve the cell wall and middle lamella to aid stylet penetration in a similar way to fungi (Reese and Black 1990, Cherqui & Tjallingii 2000, Sharma *et al.* 2014). Consequently, it is reasonable to speculate that aphids themselves may trigger pathogen-like defences such as cell wall strengthening.

It is also possible that host plants metabolise L-tyrosine in order to generate active defence compounds. In Chapter 3, it was found that many important downstream metabolites from L-tyrosine were present, at slightly higher concentrations in plants accepted by *M. sativa* adapted aphids than in plants accepted by *T. pratense* adapted aphids. The metabolism downstream of secondary metabolites, such as L-DOPA and dopamine (Chapter 3; Figure 3.4 and 3.5), could explain the observed L-tyrosine

depletion. It would be interesting to observe the changes in concentration of L-tyrosine in plants at various time points and be able to relate the point of change of TP_232 aphid acceptance on *M. sativa* plants to chemical change, including monitoring downstream metabolite concentrations and RNA expression, to see if associated enzymes such as L-tyrosine decarboxylase are being activated.

A further possibility is that pathways that produce L-tyrosine are being blocked or diverted, possibly in response to the upregulation of the JA pathways as a defence against the non-native aphids. Such defence responses by *M. sativa* could result in the upregulation of the JA pathway and so divert resources from pathways that lead to L-tyrosine production. Upregulation of the JA pathway has not been shown to have a direct effect on L-tyrosine pathways, but the L-tyrosine pathway is linked to the SA pathway via the shikimate pathway (Tzina and Galilia 2010, Lin *et al.* 2014). Speculatively, the change in L-tyrosine observed here due to TP_232 aphid exposure could be affected by the same plant metabolic pathway diversion. As discussed above, the SA pathway has been found to downregulate in response to JA upregulation, so if in this study JA was upregulated in response to the non-native TP_232 aphid, this might explain reduced concentrations of L-tyrosine. Future research into plant pathway manipulation by *A. pisum* may benefit from investigating other L-tyrosine associated pathways.

If such a diversion from the L-tyrosine pathways is occurring it would be expected that changes in the other pathways, such as upregulation of the JA signalling pathway, would also be observed. Again a targeted approach to both RNA expression and changes in metabolic pathways may be useful in further research. This would give a much clearer indication of the role of other, potentially interesting, plant metabolic responses play in aphid host specificity.

4.6 Conclusion

This study demonstrated a difference in the mechanisms two divergent clones of *A. pisum* use to induce or subvert defences at different sites in the plant. In addition results here point to constitutive defences that affect aphid acceptance differently, with a likely constitutive barrier to aphids in early probing of *M. sativa* not found in *T. pratense*. These results suggest acceptance occurs at different points during aphid probing, with divergent aphid-plant interactions occurring at each site. It is likely that these differences are driven by different plant chemical mechanisms and in *M. sativa* one of these mechanisms may be linked to the expression of L-tyrosine.

The two divergent races used here form part of a speciation continuum where host plant selection plays an important part in both divergent selection and reproductive isolation (Peccoud and Simon 2010). While only two clones were used in this experiment, in Chapter 2 these clones were shown to have strongly correlated acceptance behaviours and fitness responses to a clone of the same race. Assuming the two clones used here are typical for their race, these results here suggest that if we wish to study the influence of host plant chemistry on *A. pisum* divergence fully, is it necessary to investigate both constitutive and induced chemical responses in plants.

This study makes a first link between divergent host selection behaviour and the multiple underlying aphid-plant chemical interactions. However, much is yet to be understood about the precise mechanisms of these interactions. Indeed, a recent salivary secretome study by Carolan *et al.* (2011) found that, of the 300 identified proteins found in *A. pisum* saliva, around 40% were unique and had unknown functions. Consequently, there is considerable potential for further mechanisms of plant-aphid interaction to be discovered. Targeting future studies on the fine-scale changes of inner leaf chemistry and pathway expression might give some clarity to the underlying processes that select for divergent aphid-plant interactions.

Chapter 5: Testing the direct effect of L-phenylalanine and L-tyrosine on *A. pisum* food preference using choice tests with artificial diets.

5 .1 Chapter summary

In previous Chapters phenylalanine and L-tyrosine were identified as playing potential key roles in the divergent acceptance of plants by two pea aphid (*Acyrthosiphon pisum)* host races. Phenylalanine and L-tyrosine could be functioning as direct attractants or repellents towards *A. pisum*. To test this theory, one *Medicago sativa* adapted clone and one *Trifolium pratense* adapted clone, were tested on artificial sucrose diets with varying chemical composition. Three experiments were run in which the two aphid clones were given a choice of i) diets lacking in L-phenylalanine against diets with three different concentrations of L-phenylalanine, ii) diets lacking in L-tyrosine against diets with three different concentrations of L-tyrosine and iii) diets with and without chlorogenic acid. The chlorogenic acid diet functioned as a negative control to test the method. In all three diets there was no significant difference in aphid preference between aphid clones or between compound concentrations. In the chlorogenic acid test aphids significantly preferred diets lacking chlorogenic acid. However, aphids possessed no preference for or against the L-phenylalanine diets or the L-tyrosine diets. These results suggest L-phenylalanine and L-tyrosine may not have a direct effect on aphids, although the method used does not allow this be concluded decisively.

5.2 Introduction

A major challenge in any metabolic profiling study is determining whether individual compounds, found to correlate with a trait of an organism, are causal to the biological phenomenon of interest. Comprehensive metabolic libraries (e.g. KEGG, Metacyc) have become increasingly powerful tools for finding the putative functions of compounds. However, a significant limitation is that the data available tends to be based on a small number of model organisms, which may not be representative. In addition, biological systems are inherently complex and specific compounds may have multiple functions in an organism, some of which may not yet be known. Thus, whenever possible, researchers should aim to understand the roles of individual compounds in their specific study system and this requires experimental manipulation.

In Chapter 3, the concentrations of phenylalanine and L-tyrosine in leaf extracts were found to correlate with greater acceptance by *Medicago* adapted aphids (MS aphids) relative to *Trifolium* adapted aphid (TP aphids). In addition, evidence presented in Chapter 4 further suggests that L-tyrosine may be involved in divergent *A. pisum* aphid-plant interactions, as L-tyrosine concentrations in *M. sativa* were found to be altered by pre-exposure to a TP aphid clone. These two compounds may be part of a plant metabolic pathway that plays a role in divergent aphid acceptance. However, in Chapter 4, electrical penetration graph (EPG) recordings indicated a mix of both constitutive and induced elements being involved in the acceptance or rejection of *M. sativa*. Thus it is would be informative to test whether phenylalanine or L-tyrosine themselves act as attractants or deterrents to aphids.

There is evidence in the literature that amino acids act as attractants to aphids. By measuring the uptake of ^{14}C labelled diets, Prosser *et al.* (1992) demonstrated that *A. pisum* aphids prefer artificial diets with higher amino acid concentrations. Vogel and Moran (2011) found differences among *A. pisum* clones in their requirement for essential amino acids (EAAs) and noted that one of six *A. pisum* clones, dependent upon dietary EAAs, had reduced fitness when arginine was not present. This clone-specific dietary requirement for arginine was linked to a mutation in the facultative aphid symbiont *Buchnera aphidicola* (Voegel and Moran 2001). Conversely, it has been shown that L-tyrosine in feeding solutions inhibited growth rates of the tobacco budworm *Heliothis virescens,* suggesting that L-tyrosine may have allelochemical effects on some insects (Lokvam *et al.* 2006).

The use of artificial diets is a long established method for testing the biological roles of specific compounds in aphid biology. Aphids respond well to artificial diets and, after many decades of development, diets are now able to support aphids for extended periods of time. Indeed, a single strain of *Myzus persicae* was maintained on artificial diets for over 30 years (Van Emden 2009). Consequently, using

artificial diets under controlled conditions it is possible to ask if specific compounds can affect aphid preference and acceptance.

There are multiple approaches to measuring the effects on aphids of compounds in artificial diets. The most straightforward approach is to use measures of performance, including fecundity (Douglas *et al.* 2006, Del Campo *et al.* 2003, Ramsey *et al.* 2014), mortality (Sadeghi *et al.* 2009, Mittler and Dadd 1963) and the rate of development (Mittler and Dadd 1963, Douglas *et al.* 2006). However, as discussed in Chapter 2, relying exclusively on measurements of aphid performance risks overlooking aphid traits that specifally influence acceptance and host choice.

Measurements of change in aphid mass, as they feed, have been used as a proxy for the rate of diet uptake (Vogel and Moran 2001). However, rate of sap uptake can be more precisely measured by inoculating diets with radio isotopes such as P^{32} or C^{14} and measuring their uptake by aphids (Mittler and Dadd 1963, Prosser *et al.* 1992). EPG has even been used to measure the stylet activity of aphids on varied diets (Sauvion *et al.*2004, Goławska and Łukasik 2012).

Quantification of the effects that specific chemicals have on aphid preferences can be most simply achieved by presenting aphids with diets of varying chemical composition. The researcher can observe which diets aphids tend to settle on more. Such a direct choice method has been developed by Sauvion *et al.* (2004) to test aphid responses to lectins, using choice chambers and supplemented artificial diets. Consequently, an adaptation of the Sauvion *et al.* (2004) choice chamber has been used here to test the hypothesis that aphids respond directly to host plant phenylalanine or L-tyrosine content, either as attractants or repellents, and to determine whether responses to the two compounds are divergent between a *Medicago sativa* specialist and a *Trifolium pratense* specialist *A. pisum* clone.

5.3 Methods

5.3.1 Aphid culture

Two *A. pisum* clones were used, a *M. sativa* specialist LSR1 (International Aphid Genomics Consortium 2010) and a *T. pratense* specialist L7Tp_232 (source SE France, supplied by JC Simon, INRA, Rennes), referred to below as MS_LSR1 aphids and TP_232 aphids, respectively. Aphids were taken from age controlled populations created using the method previously described (Chapter 2). In brief this involved inoculating *V. faba* plants with adults for 24 hours then using the resulting progeny for experimentation. The aphids were maintained on *V. faba* plants in a climate controlled cabinet with 16 hour day length at temperatures of 20^0C during the day and 15^0C during the night. Aphids were used for experimental purposes at 10 days old.

5.3.2 Artificial diets

Artificial diets were prepared, based on diets developed by Prosser and Douglas (1992) (Appendix 23) except that L-tyrosine and L-phenylalanine concentrations were varied in experimental diets (Table 5.1). Amino acid, mineral and vitamin stock solutions (Appendix 23) were prepared in advance and stored at -20^0C. Stock solutions were then thawed and combined on the day of experimentation. The sucrose solution (Appendix 23) was then prepared and added. The sucrose concentration of artificial diets was 0.25 $mol.l^{-1}$, which was the optimal concentration for ingestion rate determined by Douglas *et al.* (2006).

Following the experimental design discussed below L-tyrosine and L-phenylalanine were added to the solution at varying concentrations (Table 4.1). These concentrations were at 1x, 5x and 10x the original diet concentrations of L-tyrosine and L-phenylalanine used by Douglas and Prosser (1992). Above 10x both amino acids were close to saturation, thus representing an extreme concentration. In addition to the test diets, a diet was also formulated by adding chlorogenic acid at 0.0001 $mg.ml^{-1}$ to the 'full' diet formulation. Chlorogenic acid is a known aphid repellent with little toxic effect on aphids (Sauvion *et al.* 2004, Leiss *et al.* 2009). The chlorogenic acid diet functioned as a negative control in this experiment, with

aphid rejection of chlorogenic acid diets validating the methods used. Potassium phosphate was then added to the solution to bring the solution acidity to between pH 7.0 and pH 7.5 (Appendix 23).

Table 5.1: Test diet formulations

Diet	L- L-tyrosine concentration (mg/ml^{-1})	L- phenylalanine concentration (mg/ml^{-1})
Full diet	0.085	0.34
Chlorogenic acid †	0.085	0.34
L- L-tyrosine lacking	0	0.34
L- L-tyrosine enhanced 1	0.425	0.34
L- L-tyrosine enhanced 2	0.85	0.34
L- phenylalanine lacking	0.085	0
L- phenylalanine enhanced 1	0.085	1.7
L- phenylalanine enhanced 2	0.085	3.4

† the *Chlorogenic acid* diet had chlorogenic acid added at 0.0001 mg ml^{-1}

5.3.3 Choice tests

Binary choice experiments were carried out using a modification of the methods of Sauvion *et al.* (2004). Cylindrical chambers were cut into blocks of Perspex (Figure 5.1). The mouths of Eppendorf tubes were sealed with a 0.15 ml. sachet of diet solution, encapsulated in a double layer of parafilm. The tubes were mounted in pairs at either end of the cylindrical boxes, to create a chamber offering aphids two diet choices. The base of each Eppendorf tube was removed to allow inspection of aphids on the diet sachet (Figure 5.1).

Individual adult aphids were placed in the choice chambers. Both aphid clones were tested at the same time, with clones being placed in chambers in an alternating sequence (Figure 5.2). For the experiments with three diet concentrations, diets were arranged in a balanced block design, spread across seven rows of choice chambers (Appendix 24). Other than during inspection, the aphids were kept in the dark to avoid any influence of directional light. Each aphid was offered a choice between an artificial diet deficient in the test compound and a diet containing the test compound (Table 5.2). Each chamber was scored hourly for 11 hours, noting if an aphid was on the control diets sachet (score = +1), the treatments diets sachet (score = -1) or

neither (score = 0). Three independent experiments were performed as described in Table 5.1. The first five hours of recording were not used in the analysis, in order to exclude the time it took for aphids to recover from the disturbance of being placed in the chamber and to give time for aphids to sample each side of the chamber before recording. Each compound was tested over two separate days. For L-tyrosine on the second test day to increase the number of replicates aphid where tested only on the first and third concentration, with the two diets being tested in an alternating sequence.

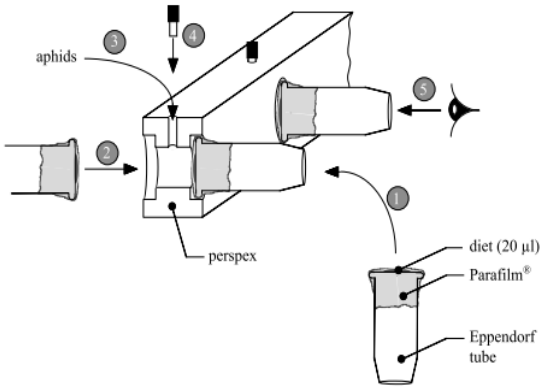


Figure 5.1: Choice test as designed by Sauvion et al. (2004) to test aphid reactions to diets of different compositions. 1) diet preparation, 2) mounting diets on choice cage, 3) introduction of aphids, 4) chamber closed off, 5) visual scoring. Diagram taken from Sauvion *et al.* (2004).

Table 5.2: Diet comparisons made between control and treated diet options.

Choice tests	**Control option**	**Treatment option**
Exp 1: Negative control (Chlorogenic acid)	Full	Chlorogenic acid
Exp2: L-tyrosine		
Tyr test 1	L-tyrosine lacking	Full
Tyr test 2	L-tyrosine lacking	L- L-tyrosine enhanced 1
Tyr test 3	L-tyrosine lacking	L- L-tyrosine enhanced 2
Exp3: L-phenylalanine		
Phe test 1	L-phenyalanine lacking	Full
Phe test 2	L-phenyalanine lacking	L-phenylalanine enhanced 1
Phe test 3	L-phenyalanine lacking	L-phenylalanine enhanced 2

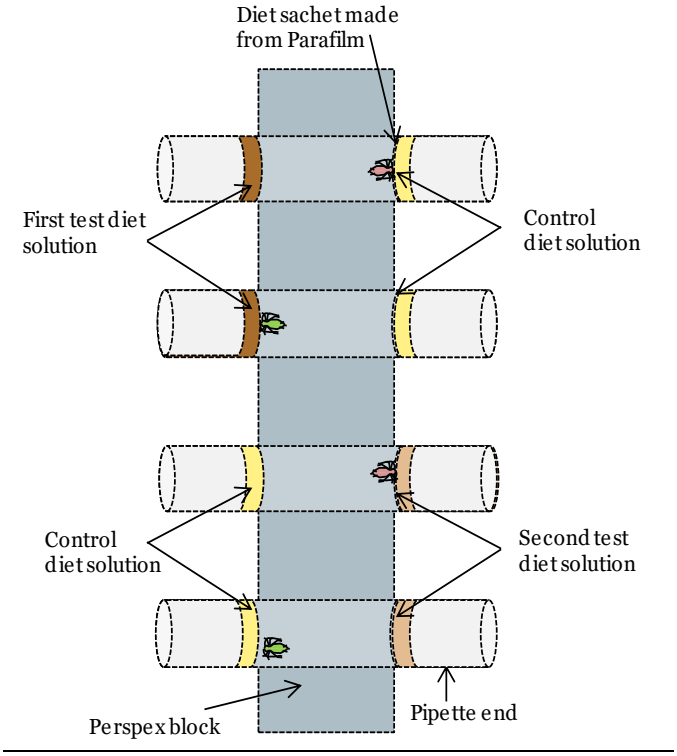


Figure 5.2: Example section of a row of diet choice tests showing the comparison of two test diets with control diets. Red aphids are MS_LSR1 and green aphids are TP_232.

5.3.4 Statistical analysis

Sides of chambers were given a positive value if they contained the control diet and a negative value if they contained the treatment sample. When an aphid made no choice during the whole test period the replicate was removed from the analysis. Values from the 6th to the 11th hour of observations were then averaged to give the *mean preference*. Positive mean preference scores meant aphids were more often observed on the control diet, while negative mean preference scores meant aphids were more often observed on the treatment diet. A score close to zero means that aphids were equally likely to be on the control or treatment diets. For each of the three experiments, a linear regression model was used to test if there was a significant difference in choice scores between the two aphid clones and, for L-phenylalanine and L-tyrosine experiments, if there was a significant difference among the three test concentrations and if there was an interaction between the aphid

clone and test concentration. A one-sample Student t-test was then performed to observe if there was any significant deviation from a mean of 0 across all aphids within an experiment.

5.4 Results

For the choice between the control and the chlorogenic acid diet there was no significant difference between the mean preference scores for the two clones (F= 0.224, df= 1, 149, *P* = 0.6362), meaning that the effect of the negative control was the same for both aphid clones (Figure 5.3). Aphids significantly preferred the control diet over the chlorogenic acid diet (t = 3.34, df = 150, *P* = 0.001).

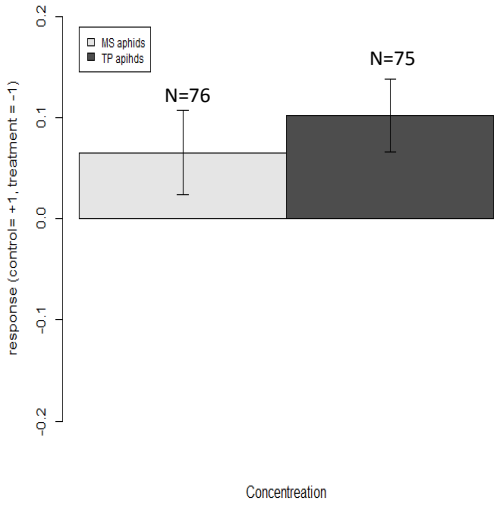


Figure 5.3: Mean aphid choice between artificial diets and negative control diets. Negative control diets contain chlorogenic acid at a concentration of 0.0001 mg ml^{-1}. . Mean ± SEM

For L- phenylalanine diets there was no significant difference in mean preference scores either between the two aphid clones (F= 0.20, df= 1, 240, *P* = 0.65) or the three L-phenylalanine concentrations (F= 0.28, df= 2, 240, *P* = 0.76), neither was there a significant interaction between aphids and L-phenylalanine concentration (F= 1.96, df= 2, 240, *P* = 0.14) (Figure 5.4A). For L-tyrosine diets there was also no significant difference in the mean preference scores between the two aphid clones (F= 0.21, df= 1, 250, *P* = 0.64) or the three L-tyrosine concentrations (F= 0.32, df= 1, 250, *P* = 0.72). Neither was there a significant interaction between aphids and the L-tyrosine concentration (F= 0.33, df= 1, 250, *P* = 0.71)(Figure 4B).

Overall mean preference scores were not significantly different from zero for the L-phenylalanine experiment (t = -1.41, df = 245, p-value = 0.159) or the L-tyrosine experiment (t = -0.2, df = 255, p-value = 0.84). This can be interpreted as there being no evidence for aphid preference for or against diets containing either compound relative to control diets. These results also indicate that there is no difference in the response of aphids from the *M. sativa* and *T. pratense* host races to L-tyrosine or L-phenylalanine in the diet, regardless of the concentration used.

5.5 Discussion

The comparison of aphid behaviour in response to negative controls shows that the method used here can distinguish preference between the full diet and a diet containing a strongly repellent compound. However, tests for both aphid clones on L-tyrosine and L- phenylalanine containing diets showed no strong preference to either diet. This suggests that neither aphid has a particularly strong behavioural response to these compounds. Assuming the use of artificial diets reflects aphid responses in nature, this would suggest aphid divergent adaptation to plants containing L-tyrosine and phenylalanine, seen in Chapter 3, is not due to the direct perception of, or the allelochemical effects of, these compounds.

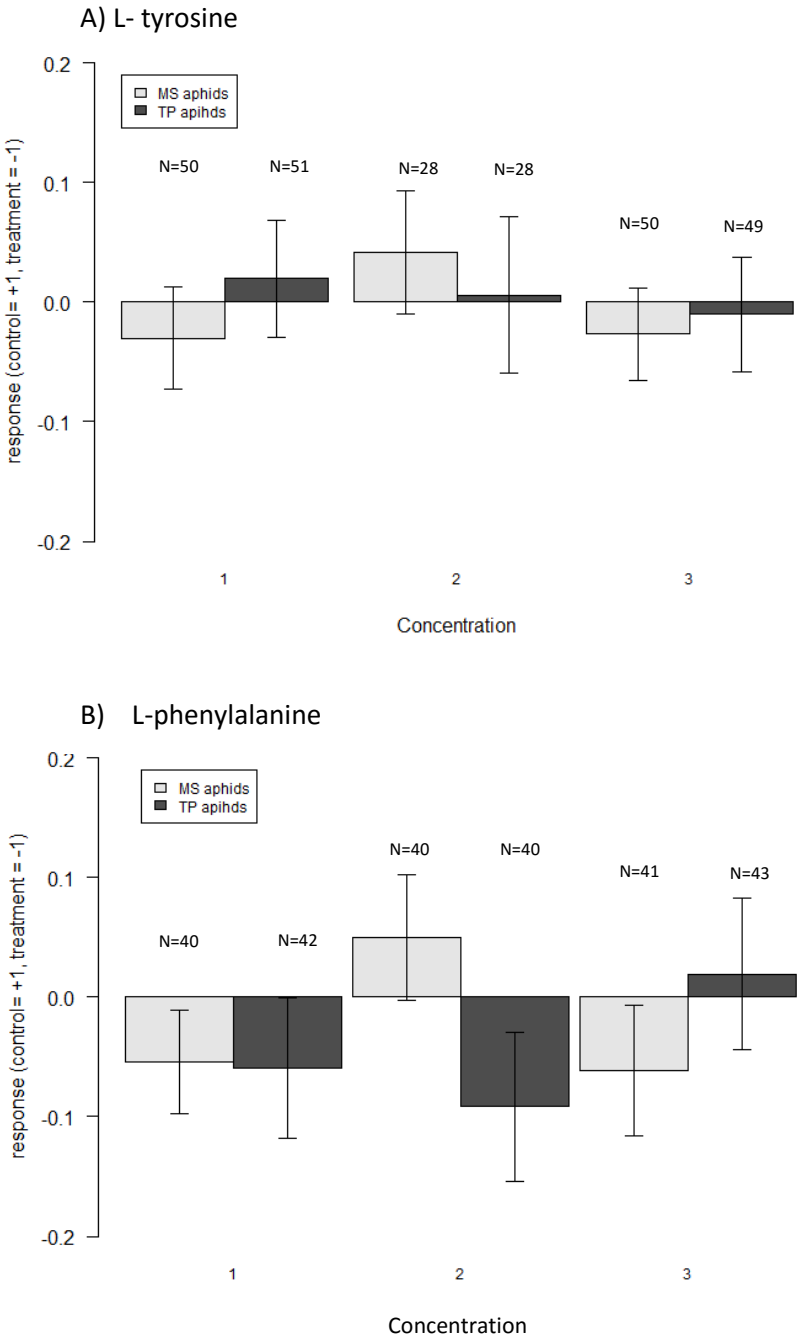


Figure 5.4: Mean aphid choice between artificial diets and test diets. Tets diet contain A) L- tyrosine at concentrations 1) 0 mg ml^{-1}, 2) 0.425 mg ml^{-1}, 3) 0.85 mg ml^{-1}, and B) L- phenylalanine at concentrations 1) 0 mg ml^{-1}, 2) 0.7 mg ml^{-1}, 3) 3.4 mg ml^{-1}. Mean ± SEM

Considering the very strong differences in acceptance seen between the MS_LSR1 and TP_232 clones in response to *M. sativa* and *T. pratense* (Chapter2), to observe no discrimination between the MS_LSR1 and TP_232 aphid clones in this experiment suggests that these compounds in isolation do not directly determine plant choice. As discussed in Chapters 3 and 4, L-tyrosine and L- phenylalanine concentrations could be correlated with aphid acceptance because they are part of a relevant metabolic pathway, in which case aphids would not directly respond to these compounds, but instead would react to downstream plant metabolites.

However, due to both biological issues and features of the experimental design used here, it is not possible to be certain if the negative result reliably reflects the role of these compounds during plant acceptance. There are several biological reasons why the results of this experiment may not be representative of what occurs in nature. Firstly, in a natural context, the compounds analysed here occur alongside a wide diversity of other plant chemicals. It could be that phenylalanine and L-tyrosine have a direct impact on aphid behaviour only in specific blends of plant compounds. Indeed, previous research has found more examples of the specific ratios of chemical compounds in blends acting as host identifiers, than examples of specific individual plant compounds (Bruce 2005). For example *V. faba* volatiles, which in isolation are repellent to the black bean aphid (*Aphis faba*), function as attractants when mixed in blends (Webster *et al.* 2008, Webster *et al.* 2010). Plant volatiles in blends have been shown to stimulate the antennal sensillae of many herbivorous insects, indicating that they are used in host location and acceptance; with compounds in blends commonly eliciting stronger responses than compounds in isolation (Bruce and Picket 2011). Compounds may also need to be presented to aphids in their plant contexts to elicit a response, such as within the cell cytoplasm or phloem sap. Finally, concentrations used in this experiment may have been either too strong, or too weak, to elicit a response from aphids. Experimentation at a wider range of concentrations, and mixes of compounds, might yield different results.

There are also several reasons why the experimental design used here might have created false negative results. Firstly adult aphids were used in each of the choice

chambers to remain consistent with previous research (Chapters 2-4), but because only one adult could fit into each chamber this restricted the number of individuals tested per replicate. In addition many aphids tended to stay away from the food sachets and this resulted in a high number of zero values recorded. This created a problem, as statistical means were drawn towards zero, possibly obscuring patterns of preference. Sauvion *et al.* (2004) avoided these problems by using 6 neonate juveniles (1-24 hours old) per chamber and taking a percentage score of preference, thus increasing greatly the chance of a positive or negative choice being recorded. The use of juveniles is also advantageous, as they have a greater tendency to stay in their preferred feeding positions than adults (Gish *et al.* 2012). Repeating this experiment with earlier instars, or with larger chambers and more adults, may have produced clearer results. Secondly, it is important to acknowledge that a negative result may be due to a lack of statistical power. Consequently, L-tyrosine and/or L-phenylalanine might have had a direct effect on aphid behaviour, but with an effect size too small to distinguish from the variance. Sauvion *et al.* (2004) used only 18-24 replicates per treatment, but this was enough to see an effect. However, this study data had much greater variance among replicates and so much larger sample sizes may have been necessary to see a significant effect. This was further exacerbated by the fact that there were half as many replicates for each treatment concentration, as there were for negative controls.

Finally, a challenge in this study was that L- phenylalanine and L-tyrosine are EAAs included in the original diet formulation by Prosser and Douglas (1992). It was decided to test L-phenylalanine with L-tyrosine in the diet and vice versa. However, any interaction effect between L-tyrosine and L-phenylalanine may have affected aphid responses to diets. If this was the case, any individual compound effects might be observed only if the other EAA is missing. It might be worth testing the effect of the test compound on aphid choice with a stripped down sucrose diet, with no other EAAs. Another approach would be to observe whether other individual EAAs, and blends of EAAs, had an effect on aphid choice of L-phenylalanine and L-tyrosine.

5.6 Conclusion

The results of the choice experiment reported here suggest neither L- phenylalanine nor L-tyrosine was either a strong repellent or attractant to *A. pisum*. However, it was not possible to determine definitively whether L-phenylalanine or L-tyrosine have a direct effect on aphid choice. Re-testing with multiple individuals per choice chamber, with a greater number of replicates and using different concentrations of L-phenylalanine and L-tyrosine, may reveal different results. However, it is possible that the influence of L-phenylalanine and L-tyrosine on aphids is via associated metabolic pathways, in which case downstream compounds may be directly causing divergent choices by *A. pisum* host races. Further testing of associated metabolites may be fruitful.

Chapter 6: Discussion and Conclusions

6.1 Study summary and implications

This study examined the link between host chemical ecology and the mechanisms behind the divergence of two *A. pisum* races. To do this, the aphid-host plant system has been viewed from three differing perspectives: A) how aphids respond to plant chemistry, B) how plants chemically respond to aphids and C) how these processes select divergently on aphid traits (Figure 6.1). It is only when these perspectives are considered holistically that we can fully appreciate the ways in which host chemical ecology acts as an important selection force upon *A. pisum* host races.

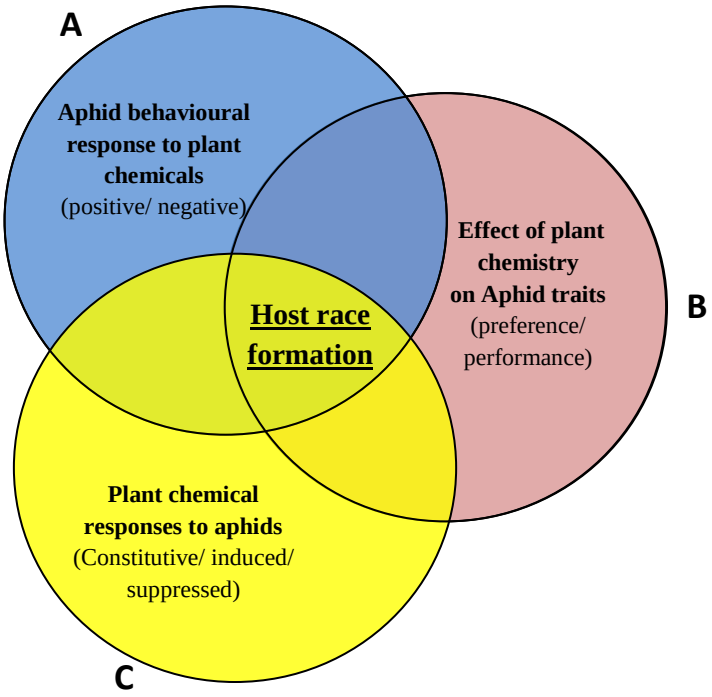


Figure 6.1: Venn diagram of the different perspectives taken in studying the role of host plant chemistry in the evolution of host plant races. For each perspective there are different organism responses to be observed, (shown in brackets).

Figure 6.2 shows how key questions and research activities of this study relate to each other. These can be interpreted below in terms of the three different perspectives (Figure 6.1). In Chapter 2 (Figure 6.2; blue boxes) a multivariate method to profile aphid host acceptance was developed and compared to performance measures (fecundity and aphid condition) across 16 species from the genera *Medicago* and *Trifolium* to illustrate the close relationship between the two (perspective "A" in Figure 6.1). A close correlation between acceptance and performance was found, suggesting these traits are closely linked and likely to be influenced by the same selection mechanisms on different hosts. This is in line with previous findings of linked preference and performance traits observed by Hawthorne and Via (2001) and Ferrari *et al.* (2006, 2008). The fact that the preference and performance trait association in *A. pisum* is maintained across other non-host plants in this study, suggests linked traits in *A. pisum* aphids are likely to be selected by continuously varying features of plants and are not simply linked to species specific host plant traits.

The E2 and LD1 acceptance profiles were then compared in Chapter 3 (Figure 6.2; red boxes) to the metabolomic content of 19 plant species to establish which plant compounds might be responsible for divergent aphid host choice (perspective "B" in Figure 6.1). Two closely related metabolites, phenylalanine and L-tyrosine, were identified and found to be in high concentrations in *Medicago* aphid accepted plants but not *Trifolium* aphid accepted plants. This lead to the suggestion that a plant metabolomic pathway may be important in *A. pisum* host race divergence.

In Chapter 4 (Figure 6.2; green boxes) host responses to aphid attack were explored further by observing the responses of two host plants, *M. sativa* and *T. pratense* (perspective "C" in Figure 6.1) when exposed to two clones with differing host specialisations. It was found that aphid behaviour could be interpreted in terms of both different responses to constitutive features of host plants and to other induced and suppressed plant responses. These different aphid-plant interactions seen on *M. sativa* and *T. pratense* seem to have occurred at different stages of aphid leaf probing

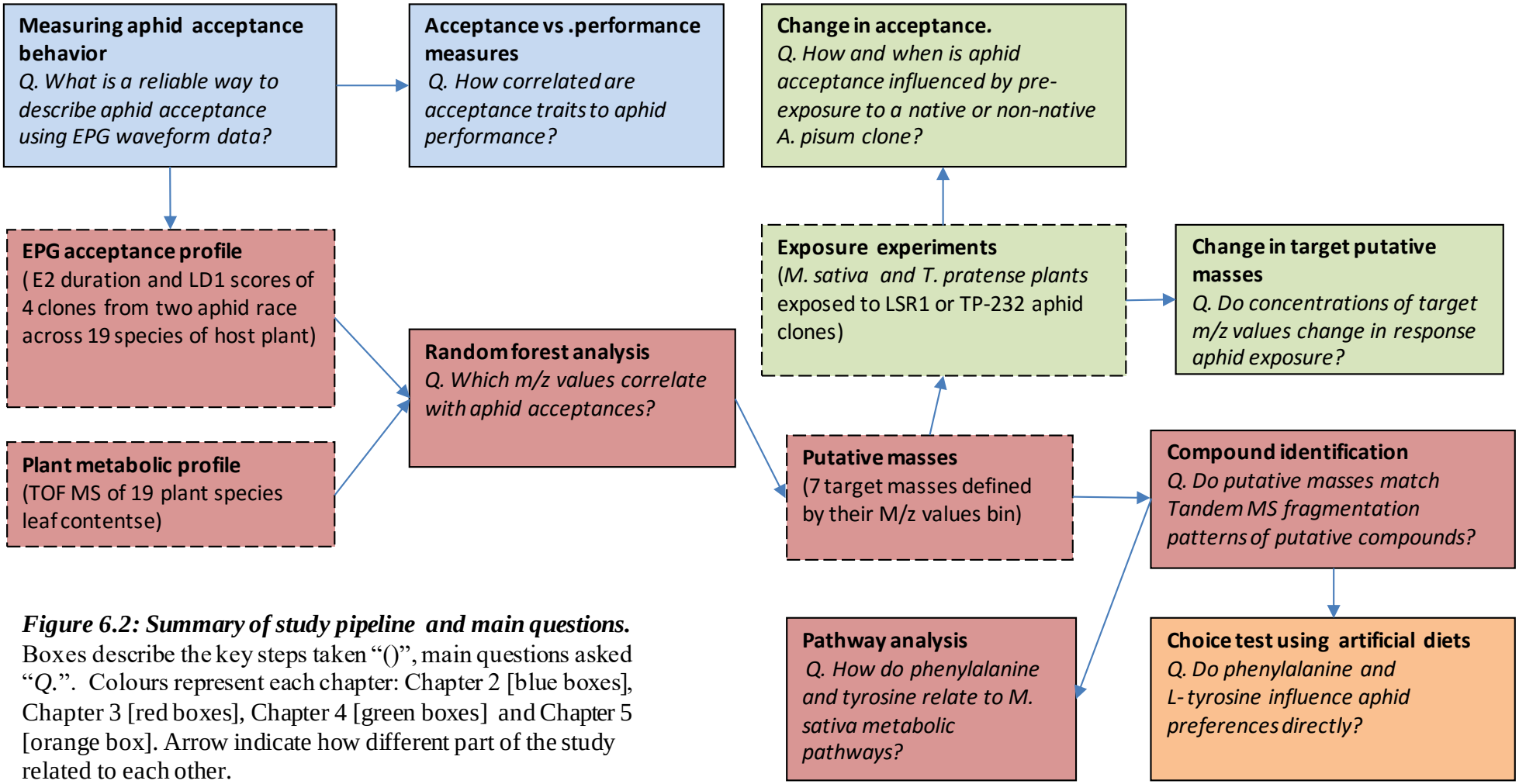


Figure 6.2: Summary of study pipeline and main questions. Boxes describe the key steps taken "()", main questions asked "*Q.*". Colours represent each chapter: Chapter 2 [blue boxes], Chapter 3 [red boxes], Chapter 4 [green boxes] and Chapter 5 [orange box]. Arrow indicate how different part of the study related to each other.

(Figure 6.3). In addition, in Chapter 4 it was found that exposure of *M. sativa* to an *A. pisum* clone native to *T. pratense* caused an apparent change in the concentration of L-tyrosine, further suggesting a role of plant metabolic pathways in the divergence of *A. pisum* races. Finally, in Chapter 5 (Figure 6.2; orange box) whether the two divergent aphid clones responded positively or negatively, to diets containing varied concentrations of L-phenylalanine or L-tyrosine, was tested. This was to determine if these chemicals play a direct role in host choice (perspective "B" in Figure 6.1). No conclusive evidence of their role as repellents or attractants was found.

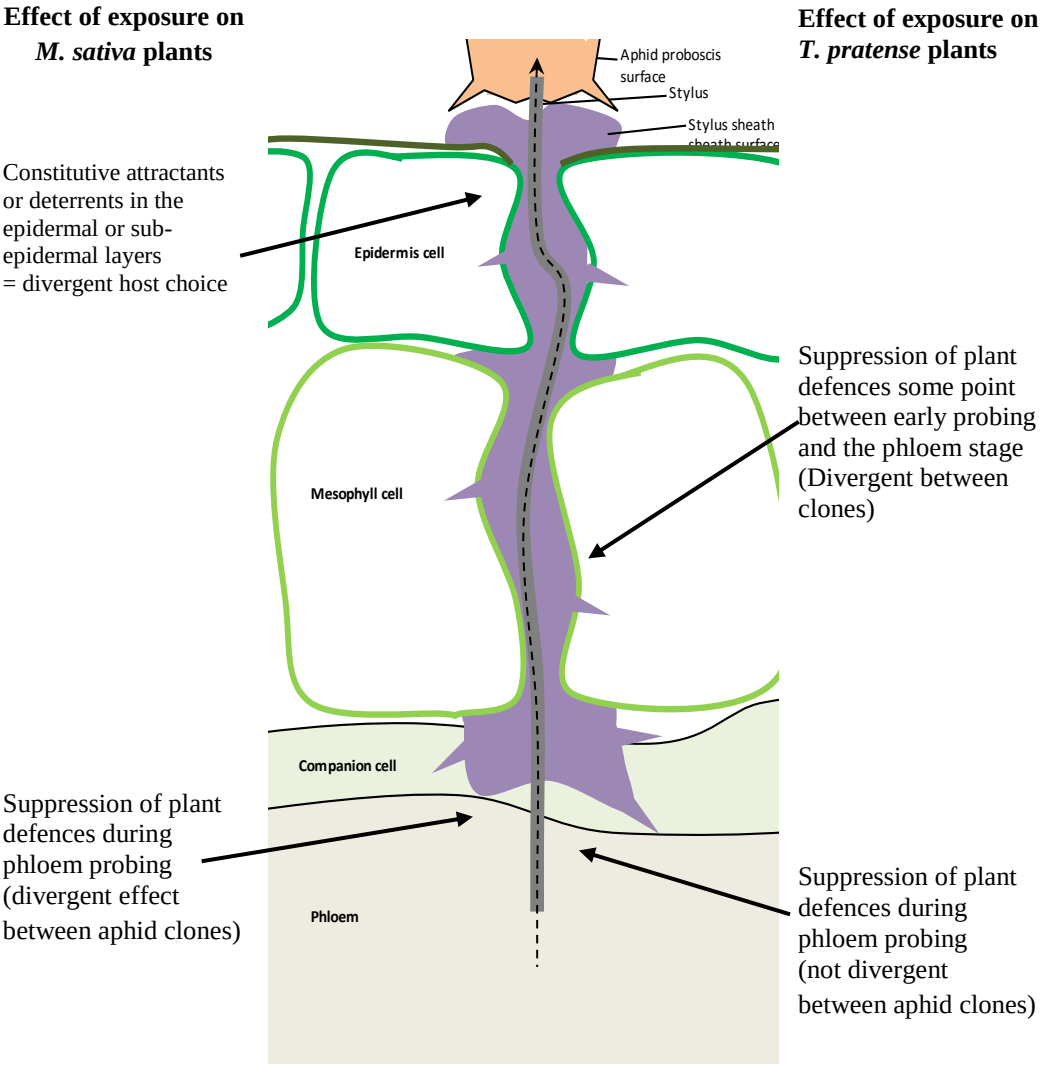


Figure 6.3: Summary of divergent aphid-plant interactions inferred from Chapter 4 EPG results.

6.2 Methodological considerations

6.2.1 EPG profiling

This study uses both PCA and LDA multivariate analysis to summarise a complex set of EPG measurements into a small number of acceptance profile scores. These multivariate profiles were found to be reliable overall summaries of aphid host acceptance (Chapter 2). However, using multivariate techniques results in a loss of specific detail in the data. While individual EPG measurements cannot definitively describe plant acceptance, the information about the individual behaviours each EPG measurement provides is valuable for interpretation of the process of acceptance, which is lost when converting them into multivariate scores.

The decision on whether to use multivariate scores or individual waveform measurements is dependent upon the research question being asked. Overall multivariate profiles are most effective when aphid host races are to be compared to other variables, such as host plant chemistry, as shown in Chapter 3. The individual waveform measurements, on the other hand, provide valuable information for testing specific hypotheses about the mechanisms underlying host plant acceptance, as in Chapter 4. Mining the full EPG data set in the hope of finding correlations followed by *post hoc* interpretation is clearly unsatisfactory. Although comparison of waveforms that represent different behaviours was found to be useful, in this and other studies (e.g. Wilkinson and Douglas 1998, Caillaud and Via 2000, Schwarzkopf *et al.* 2013), pre-selected waveform measurements should also be non-correlated. However, a degree of correlation between many EPG waveforms is inevitable as aphid decisions early in acceptance are likely to influence the occurrence and durations of later behaviours. In summary, in the design of experiments and interpretation of EPG data, research need to take into consideration whether i) it is intended to test specific mechanisms of behaviour or profile overall aphid acceptance behaviour and ii) waveforms selected for analysis are independent of each other.

6.2.2 Metabolic profiling

In this study both a targeted and a non-targeted approaches to metabolic profiling were used. Fenselau (2013) argues the role of metabolomic profiling is to identify as many metabolites as possible and be explorative (i.e. to ask "what is there") rather than seeking confirmation of what is expected (i.e. "is it there"). However, in exploring complex systems with metabolic data, targeted and non-targeted approaches have disadvantages and advantages. Targeted metabolic profiling is best used under experimental conditions to identify patterns of change that are meaningful. More pragmatically, targeted research allows scientists to utilise limited resources to study the metabolites that are most likely to be useful for interpretation. In addition, targeting a select range of metabolites and therefore fewer masses, allows for greater machine sensitivity, giving more confidence in the results.

In contrast non-targeted metabolic analysis is fraught with challenges. As non-targeted analysis requires a broader spectrum of masses, this means machine sensitivity is sacrificed. Lower sensitivity means an increased chance of false negative results, with a bias towards detecting the most concentrated or easily ionised compounds (so-called ion suppression described in Chapter 3). At the same time, when profiling across 100s to 1000s of compounds, the chances of detecting false positive results increases. Consequently when interpreting non-targeted analysis it can be hard to ascertain the reliability of results, and often requires supplementing findings with other evidence (Sugimoto *et al.* 2012).

An additional challenge with non-targeted metabolomics is that any masses discovered need to be chemically identified. Libraries of mass spectrum data are powerful tools to give putative identification to masses (Tohge and Fernie 2009). However, as many compounds or compound fragments share the same mass to charge ratios (m/z), it is not always possible to establish the true identity of some masses. Also, the fact that exact spectral results of many mass spectrometry techniques are instrument specific, adds complication to mass identification via

libraries (Bino *et al.* 2004). It is only by directly comparing m/z value fragmentation patterns to standards, or by purifying and performing extensive analytical research on putative masses, that compounds can be identified for certain. A complication with this is that not all chemical standards are easily available, while chemical analysis can be resource intensive. Because of these constraints it is common for putative compounds to remain identified on uncertain grounds or to remain unidentified. Despite these reservations, the critical advantage of broad-spectrum analyses is that, unlike targeted analysis, it allows us to discover novel lines of investigation unbiased by preconceived expectations. In the pursuit of the complete understanding of the role chemical ecology of plants play in the evolution of *A. pisum* aphid, a mix of approaches could prove valuable, with broad spectrum non-targeted analysis used to discover interesting chemistry for study and targeted analysis used to test if these masses change in response to experimental treatments.

6.3 Future perspectives

6.3.1 The role plant chemical ecology in *A. pisum* host race formations

Both previous work on divergent *A. pisum* genes associated with aphid adaptation to plant chemistry (Smadja *et al.* 2009, Jaquiéry *et al.* 2012, Duvaux *et al.* 2015) and results from this study point to plant host chemical ecology as an important component of *A. pisum* host race formation. Given this, future research should continue to explore the role of host plant chemistry. A number of studies have shown individual chemicals can play a role in the attraction and rejection by *A. pisum* (Sauvion *et al.* 2004, Golawska *et al.* 2014) and other divergent phytophagus insects (Linn *et al.* 2003, Matzkin *et al.* 2006). However, this study also suggests a link between the expression of plant metabolic pathways and divergent host selection. This is particularly interesting considering *A. pisum* is known to manipulate plant metabolomic pathways (Gao *et al.* 2008, Takemoto *et al.* 2013, Schwartzberg and Tumlinson 2014). Future work needs to consider, not just individual plant compounds, but also the wider metabolomic context in which they occur. An efficient approach would be to target particular plant metabolic pathways to see how they respond to different *A. pisum* races. This could be achieved by observing changes in expression of key host metabolites, regulating proteins, or

genes. An efficient way to do this would be to focus upon key pathways and their signalling molecules. The L-tyrosine associated pathways as well as the JA and SA defence signalling pathways appear to be of particular interest, although related pathways such as the Shikimate pathway may also be revealing. In turn, understanding which host metabolic pathways result in divergent host discrimination, may also allow research to be targeted on specific plant compounds associated with these pathways.

Incorporating host chemical ecology approaches into the study of *A. pisum* speciation requires researchers to also consider the adaptive landscapes that select upon host plants and their chemistries. It may be informative to view host plants in the context of their wider ecology, as plants often need to trade-off the expression of metabolic pathways and individual compounds in response to multiple biotic and abiotic stresses. This could have implications for how different *A. pisum* races select and live upon host plants in nature, including revealing a possible mechanism to explain how different *A. pisum* races co-occur at low frequency on plants, interbreed and so maintain gene flow.

6.3.2 Metabolomics in a multi-omic approach to study of the *A. pisum* model

The increasing availability of different large –omic scale data at every level of biology (including genomic, transcriptomic, proteomic and metabolomic), gives new opportunities to explore the *A. pisum* model for testable mechanisms of divergent selection. The value of metabolomics in the study of gene function is increasingly being recognised (Hall *et al.* 2002, Bino *et al.* 2004, Khlisch and Pohnert 2015). Indeed the comparison of genomic and metabolomic data is starting to become routine in fields such as biotechnology, drug discovery and cellular biology (Joyce and Palsson 2006, Nguyen *et al.* 2013, Doroghazi *et al.* 2014).

However, the integration of very different complex data types requires the development of innovative ways to compare data by using novel statistical

techniques. In this study orthogonal variables were used as summaries of a complex multi-factorial data set of behaviour which was then compared to metabolomic data using random forest modelling, to identify meaningful masses for further study. Other "multi-omic" approaches have been developed in other fields of biology including medicine (Kaiser *et al.* 2013, MacNeil *et al.* 2015), cell biology (Joyce and Palsson 2006), and even ecology (Hultman *et al.*2015) and may have relevance to *A. pisum* research.

6.3.3 Utilising a wide selection of plant species to study the *A. pisum* model.

An obvious next step is to expand metabolomic investigation across other plants currently characterised as *A. pisum* hosts. However, in this study plants with unknown suitability as hosts were also used, to provide a wider range of host plant environment. This created a continuum of variance in acceptance, performance and metabolic profiles, to allow for more powerful analysis. The resulting wider variance between species is preferable to using pairs of hosts with binary features (i.e. plants that are either rejected entirely or always accepted by a given race), as binary comparisons are much more likely to report differences between plants that are irrelevant to the studied phenomenon. An additional advantage of using non-typical host plant species is that the use of related and unrelated plant species allows us to test if divergent host association of *A. pisum* races is driven by the selection of host plant environments only, or is also a consequence of the phylogenetic relationships of plants.

6.4 Final conclusions

In this study a continuous orthogonal variable of a complex set of behavioural responses of one organism was successfully compared to the metabolic profile of another. This resulted in the identification of compounds with interpretable biological functions. Follow up work provided evidence that suggested the existance of multi-layered plant chemical defence system that acted differentially under exposure to clones of differing host races, suggesting aphid adaptations to host plants may require multiple co-evolved traits. In addition the difference

between host plant species could be attributed to the expression of a plant L-tyrosine, associated metabolic pathways. It is the first time when plant metabolic pathways have been shown both as a possible mechanism in the aphid-plant interaction and to result in the formation of divergent *A. pisum* host races. Consequently, the concept of plant chemical ecology as a divergent selection force should to be integrated into the study of divergent *A. pisum* host races.

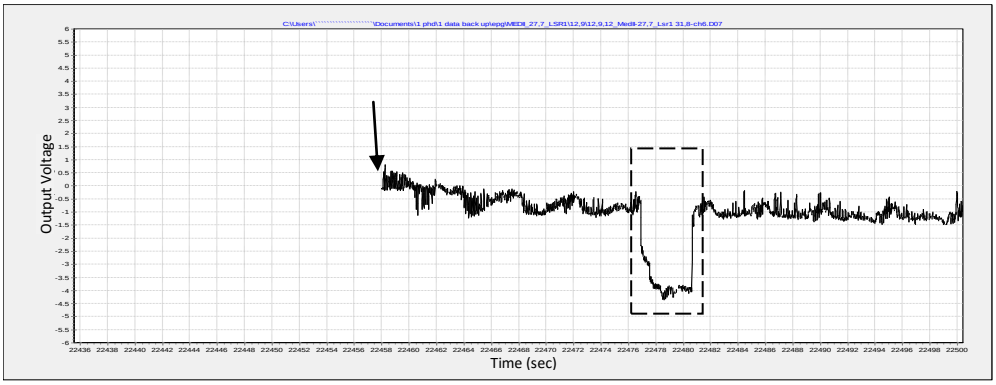
Appendices

Appendix 1: Interpretation of elelctronic penetration graph (EPG) waveforms

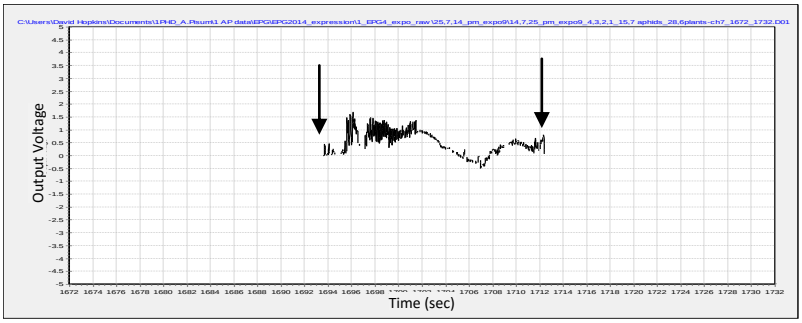
Raw EPG data consists of output potential over time which is interpreted as different stereotyped patterns known as "waveforms" using the "EPG Stylet+a" software downloaded from the EPG Systems© website (www.epgsystems.eu/downloads.php). Waveform are characterised into 8 different forms as described and interpreted by Tjallingii (1976), Tjallingii and Esch (1993), Tjallingii1 and Gabrys (1999):

1. Non-probing (np) = aphid stylet outside of the plant
2. Pathway (C) = aphid stylet within the plant apoplast
3. Potential drops (pd) = aphid stylus briefly passes through a cell wall and out again
4. Repetitive potential drops (rpd) = aphid stylet repeatedly enters cell membrane of phloem or phloem companion cell for longer periods than pd. Normally occurs just before entering the phloem, though sometimes aborted.
5. Phloem egestion (E1) = aphid salivates into the ploem
6. Phloem ingestion (E2) = aphid injesting phloem sap
7. Xylem feeding (G) = aphid ingesting xylem sap
8. Stylet derailment (F) = associated with difficulties in penetrating plant tissues

Each waveform type is illustrated below:



Firgure A1.1: Point of change from non pathway (waveform NP) to pathway (wavemform C) indicated by an arrow. The box shows a "potential drop" (wave form pd), indicating cell penetrations that are common during the pathway period.



Firgure A1.2: An example of a stylet "probe" where the aphid stlyet enters a plant leaf (first arrow) then leaves it again (second arrow). A probe is defined as any instance when the aphid sylet enters the the plant to induce a signal.

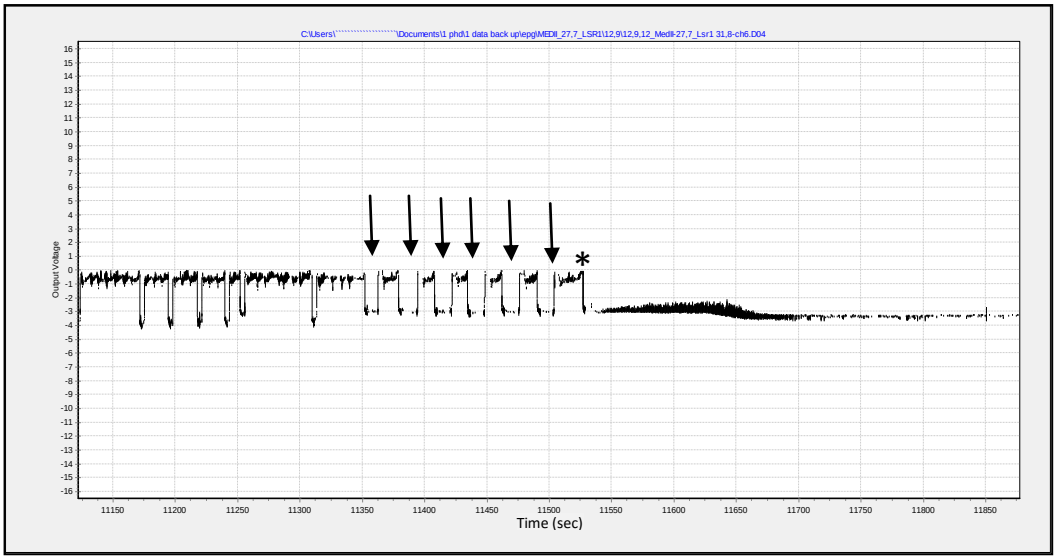


Figure A1.3: Point of change from pathway (C) to phloem stage (waveform E), showing 6 repetitive potential drops (waveform rpd, shown by arrows). Rpd is a common but poorly understood characteristic of *A. pisum* feeding. Repetitive potential drops are characterised as being different from normal potential drops as they are 11 or more seconds long, always in a sequence of three or more and typically (though not always) precede an E1 signal (*).

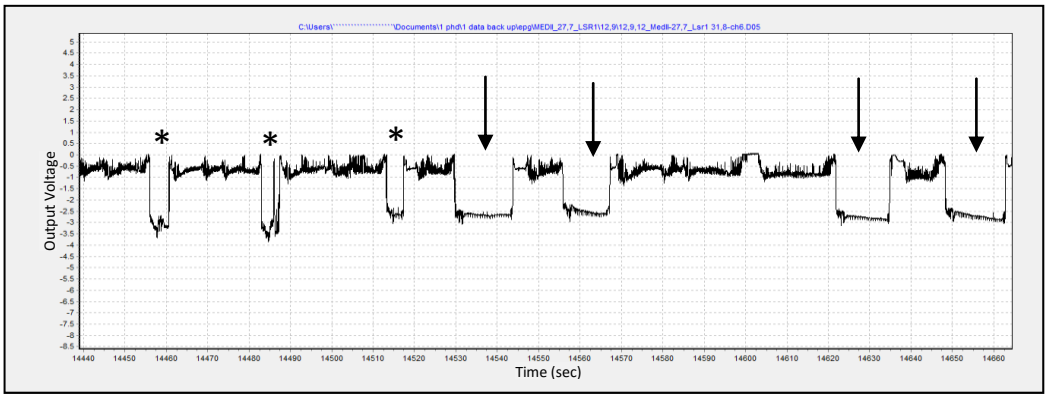


Figure A1.4: Close up showing three normal potential drops (*) next to a series of four repetitive potential drops (arrows) during the pathway phase (C).

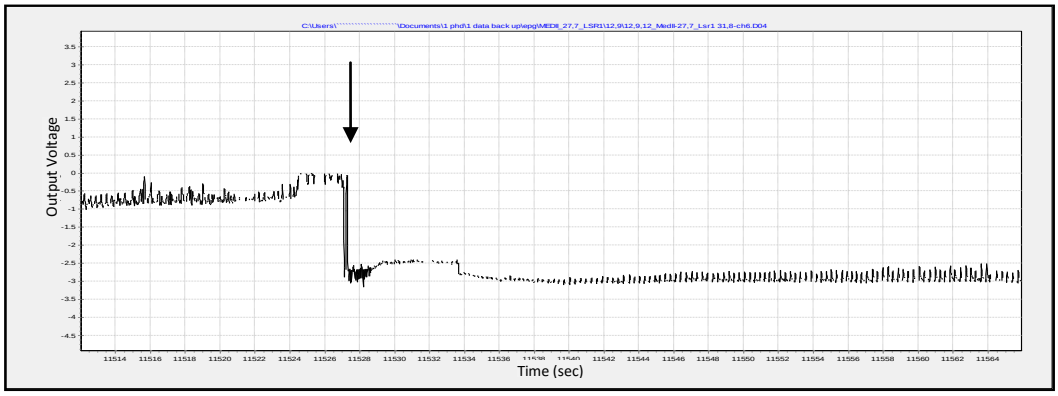


Figure A1.5: Point of change from pathway (C) to the first phloem stage (waveform E1) when aphid egests saliva into the phloem. Point of change is indicated by the arrow.

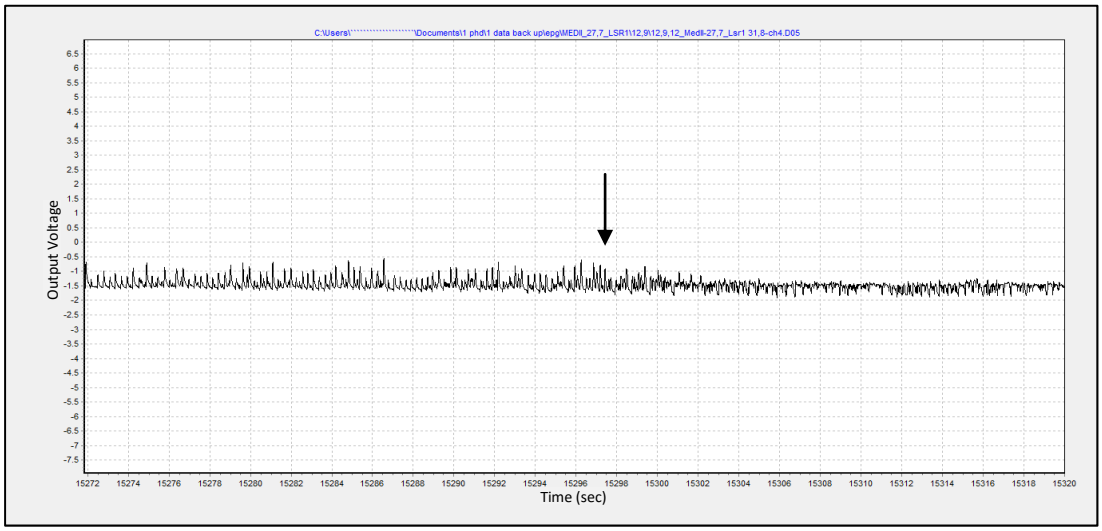
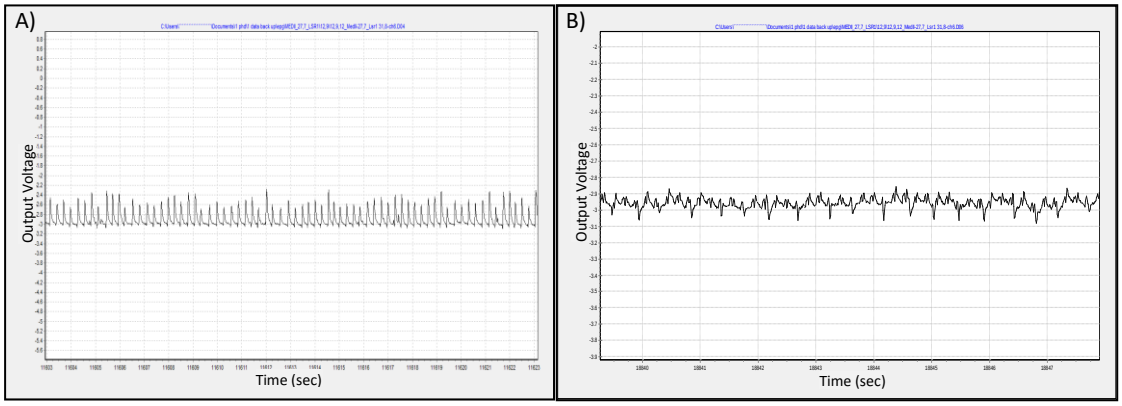


Figure A1.6: The transition between E1 phase and E2 phase when aphid ingestion of the phloem sap begins. The E1 to E2 transition occurs at the point indicated by the arrow.



Firgure A1.7: A) Close up of waveform E1 and B) close up of phloem waveform E2.

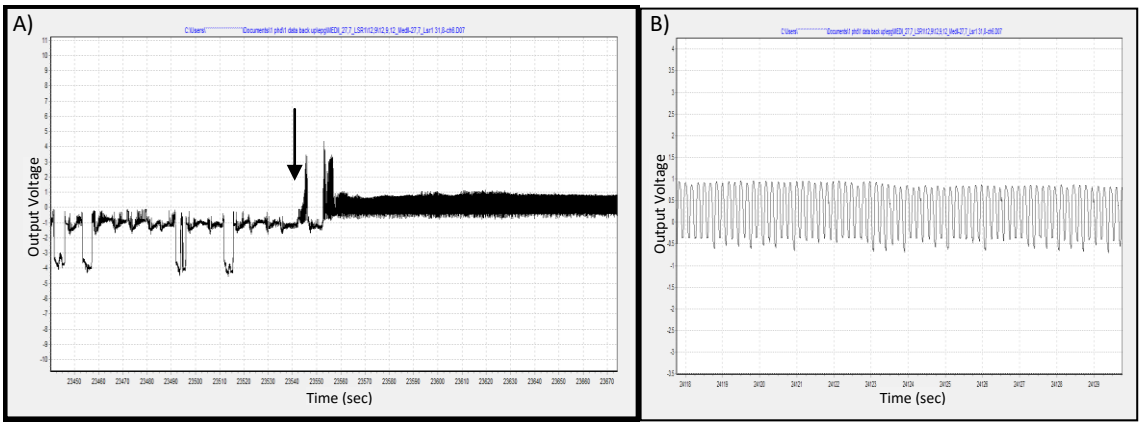
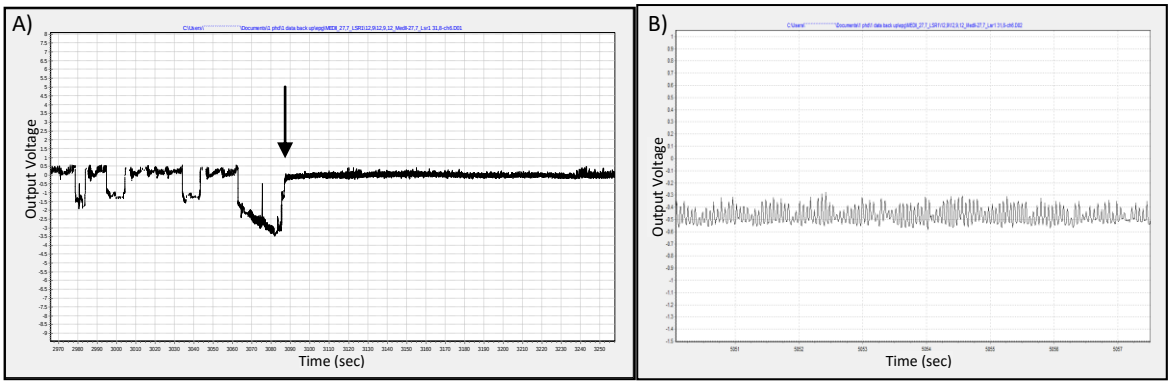


Figure A1.8: A) Point of change from pathway (waveform C) to xylem feeding (waveform G) and B) close up of the G waveform. Point of change is indicated by the arrow.



Firgure A1.10: A) Point of change from pathway (waveform C) to stylet derailment (waveform F) and B) close up of F waveform. Point of change is indicated by the arrow

Appendix 2: Seed sources

Table A2.1: List of plant species used and their seed sources

Species	Source	Details of original source
M. arabica	IBERS, Aberystwyth University, Wales	Donated 1843, Denmark, University of Copenhagen Botanical Garden
T. rubens	IBERS, Aberystwyth University, Wales	Donated sample, 886466, ex England, Ardingly, International Plant Genetic Resources Institute, Seed Handling Unit. Collected Bovec, Tolmin, Slovenia.
T. semipilosum	IBERS, Aberystwyth University, Wales	Cv. Safari. Donated, 69, Ex Wales, Bangor University
T. dubium	IBERS, Aberystwyth University, Wales	Coll. Czechoslovakia 1992
M. orbicularis	IBERS, Aberystwyth University, Wales	Donated 1861, University of Copenhagen Botanical Garden
M. littoralis	IBERS, Aberystwyth University, Wales	Donated, 1119, ex Portugal, Coimbra, Jardim Botânico da Universidade Coimbra
M. tornata	IBERS, Aberystwyth University, Wales	Donated sample, 1869, University of Copenhagen Botanical Garden
M. turbinata	IBERS, Aberystwyth University, Wales	Dinated sample, 1871, ex Denmark, University of Copenhagen Botanical Garden
T. striatum	IBERS, Aberystwyth University, Wales	Donated, ex. France, Guyancout, INRA
T. nigrescens	IBERS, Aberystwyth University, Wales	Quinequeli. Donated ex Chile, Temuco, INIA, Estacion Experimentales Garillanca per Fernando Ortega
M. laciniata	University of Sheffield	Original plants taken from Kew Gardens, UK
M. lupulina	Emorsgate Gate	Origin: Norfolk . UK
T. pratense ssp. pratense	IBERS, Aberystwyth University, Wales	Cv. AberChianti. Ex Aa 4494. Diploid red clover bred for enhanced persistence under cutting and grazing.
T. ochroleucum	Chiltern Seeds©	Origin: Oxfordshire, UK
M. truncatula	IBERS, Aberystwyth University, Wales	Breeder line, 2005. Drought selection ex Af 1734 produced in compartment 3 of Venlo. (ABY-Af 1738-2005)
T. repens *Var. small leaved*	Emorsgate Gate	Origin: Amenity
T. ambiguum	IBERS, Aberystwyth University, Wales	Cv. Summit. Ex Australia, Canberra, CSIRO, Division of Plant Industry (ABY-Ah 1475-)
M. sativa ssp. sativa	IBERS, Aberystwyth University, Wales	Cv. Sabilit. Ex AF7 (AF1)

Appendix 3: Rorison's solution

Rorison's_solution is formulated at *full strength* using the formulation in Table A3.1 and then diluted down to the concentrations used in experiments.

Table A3.1: Preparation for 1 litre of 'full stock' Rorison's solutions

Element	Stock solution	g/ L^{-1}
Macro		
Ca/N	$Ca(NO_3)_2.4H_20$	476.1
Mg	$MgSO_4$. $7H_20$	248.0
K/P	K_2HPO4.	230.7
Fe	Fe EDTA	25.0
Micro		
Mn	$MnSO_4.4H_20$	2.028
B	H_3B0	2.863
Mo	$(NH_4)_6MoO_{24}.H_20$	0.184
Zn	$ZnSO_4.7H_20$	0.44
Cu	$CuSO_4.5H_2O$	0.393

Appendix 4: Culture pot design

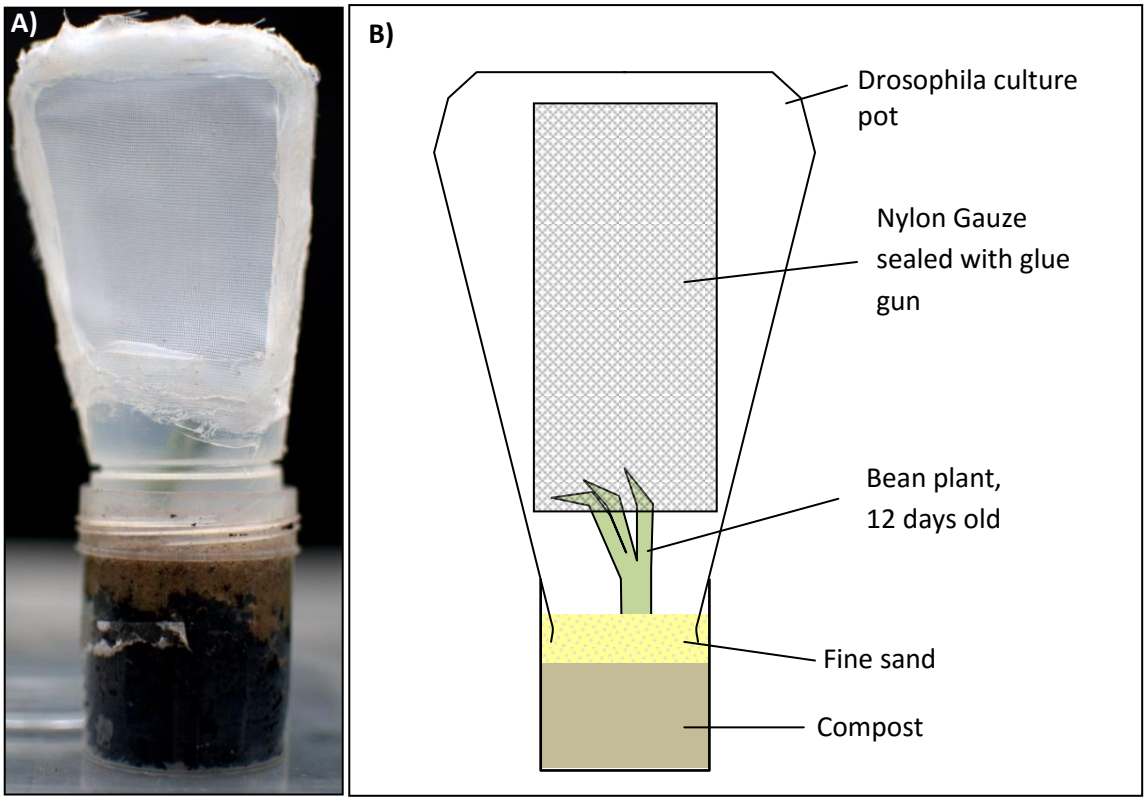


Figure A4.1: A) Photograph of "culture pot" with a single bean plant, B) diagram of "culture pot" design. Each bean plant was grown in a seed tray and transferred into a 70ml round tub. Over this a round *Drosophila* culture pot with added ventilation was inserted to give a tightly sealed container. Ventilation was made using a piece fine nylon gauze sealed with glue gun.

Appendix 5: Set-up of EPG cage

Figure A5.1: Side view of full EPG set-up contained within Faraday cage.

Figure A5.2: Front view of full EPG set-up contained within Faraday cage.

Appendix 6: Samples sizes of EPG data used in Chapter 2 and Chapter 3

Table A6.1: Sample sizes for each plant species-clone combination EPG. *(Total number of samples 580)*

	Medicago aphid race		*Trifolium* aphid race	
Plant species	LSR1	MS052	TP232	YR2
M. arabica	6	5	4	10
M. laciniata	11	7	8	7
M. littoralis	7	5	3	6
M. lupulina	10	5	8	5
M. orbicularis	9	6	6	5
M. sativa	15	11	16	14
M. tornata	11	5	9	12
M. truncatula	10	6	9	6
M. turbinata	5	6	7	8
T. ambiguum	8	6	7	8
T. dubium	6	5	11	12
T. nigrescens	5	7	12	10
T. ochroleucum	6	4	3	5
T. pallidum	5	4	9	0
T. pratense	12	7	9	12
T. repens	11	7	6	6
T. rubens	10	5	9	6
T. semipilosum	9	6	8	12
T. striatum	7	5	7	10

Appendix 7: Cleaned waveforms use in multivariate analysis

Table A7.1: List of waveform measurements <0.80 correlated used in the PCA and LDA analysis

No.	Waveform
1	Time to 1st probe from start of EPG
2	Number of probes to the 1st E1
3	Number of F
4	Duration of 1st probe
5	Duration of 2nd probe
6	Duration of the shortest C wave before E1
7	Duration of the second nonprobe period
8	Total duration of F
9	Mean duration of F
10	Number of G
11	Duration of G
12	Number of probes after 1st E
13	Number of E1
14	Number of E1 longer than 10 minutes followed by E2
15	Number of single E1
16	Duration of 1st E
17	Duration the E1 followed by first sustained E2 10 min
18	Potential E2 index
19	Total duration of E
20	Total duration of E1
21	Total duration of single E1
22	Number of probes
23	Number of C
24	Number of E1e
25	Total duration of C
26	Total duration of E1e
27	Total probing time
28	Mean duration of np
29	Mean duration of C
30	Time to from start of EPG 1st sustained E2 longer than 10 minutes
31	Time from the beginning of that probe to 1st sustained E2 longer than 10 minutes
32	Time from the beginning of that probe to 1st E2y
33	Total duration of np during the 1st hour
34	Total duration of np during the 2nd hour
35	Total duration of np during the 3rd hour
36	Total duration of np during the 4th hour
37	Total duration of np during the 5th hour

38 Total duration of np during the 6th hour
39 Number of F during the 1st hour
40 Number of F during the 2nd hour
41 Number of F during the 3rd hour
42 Number of F during the 4th hour
43 Number of F during the 5th hour
44 Number of F during the 6th hour
45 Total duration of F during the 1st hour
46 Total duration of F during the 2nd hour
47 Total duration of F during the 3rd hour
48 Total duration of F during the 4th hour
49 Total duration of F during the 5th hour
50 Total duration of F during the 6th hour
51 Number of probes during the 1st hour
52 Number of probes during the 2nd hour
53 Number of probes during the 3rd hour
54 Number of probes during the 4th hour
55 Number of probes during the 5th hour
56 Number of probes during the 6th hour
57 Time from the beginning of E1 to the end of the EPG record
58 Time from the beginning of E2 to the end of the EPG record
59 Duration of np just after the probe of the first sustained E2
60 X of probing spent in C

Appendix 8: Variance explained by PCA components

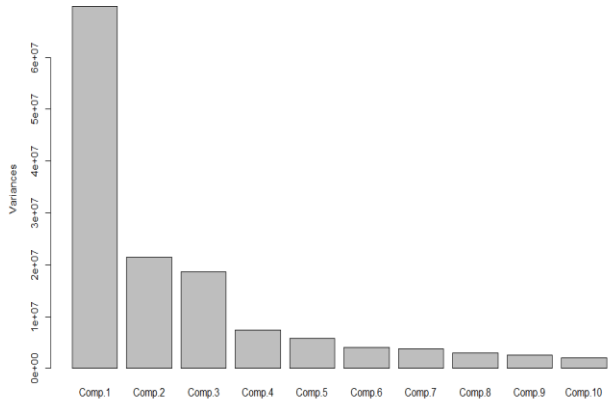


Figure A8.1: Screen plot for PCA of 60waveforms. Results show that nearly all the variance is contained within the first three PCA components, and more than half within PC1.

Appendix 9: PCA loading scores

Table A9.1: Top 5 and bottom 5 PC1 and PC2 loading scores

EPG measurement	loadings	Interpretation
PC1		
Total duration of E	0.455	Duration in phloem across all hours
Total probing time	0.411	Duration inside the plant across all hours
Duration of 1st E	0.243	Duration of 1st phloem probe
Time from the beginning of E2 to the end of the EPG	0.195	Time searching that contains feeding behaviour
Duration of 2nd probe	0.048	Duration of the second time entering a plant
...		
Total duration of np during the 4th hour	-0.049	Time spend outside of the plant in 4th hour of recoding
Total duration of np during the 6th hour	-0.050	Time spend outside of the plant in 6th hour of recoding
Total duration of np during the 3rd hour	-0.057	Time spend outside of the plant in 3rd hour of recoding
Total duration of C	-0.073	Time spent in the apoplast
Time to from start of EPG to first E2 >10 minutes	-0.709	Time spend searching the plant until a successful feeding is reached
PC2		
Total probing time	0.709	Total time inside the plant
Total duration of C	0.353	Total time inside the apoplast
Total duration of F	0.334	Total time in F waveform
Time to from start of EPG 1st sustained E2 > 10 minutes	0.187	Time to the first sustain E2 >10min
Duration of 2nd probe	0.119	Duration of the 2nd probe and aphid make into a plant
...		
Total duration of np during the 2nd hour	-0.133	Total duration stylus outside of the plant between in the 2nd hour
Time from the beginning of E2 to the end of the EPG record Z	-0.145	Duration from first E2 tot eh end of the EPG?
Total duration of np during the 1st hour	-0.154	Total duration stylus outside of the plant between 1st and 2nd hour
Total duration of E	-0.186	Duration in phloem across all hours
Duration of 1st E	-0.241	Duration of 1st phloem penetration

Appendix 10: Variance explained by LDA axes

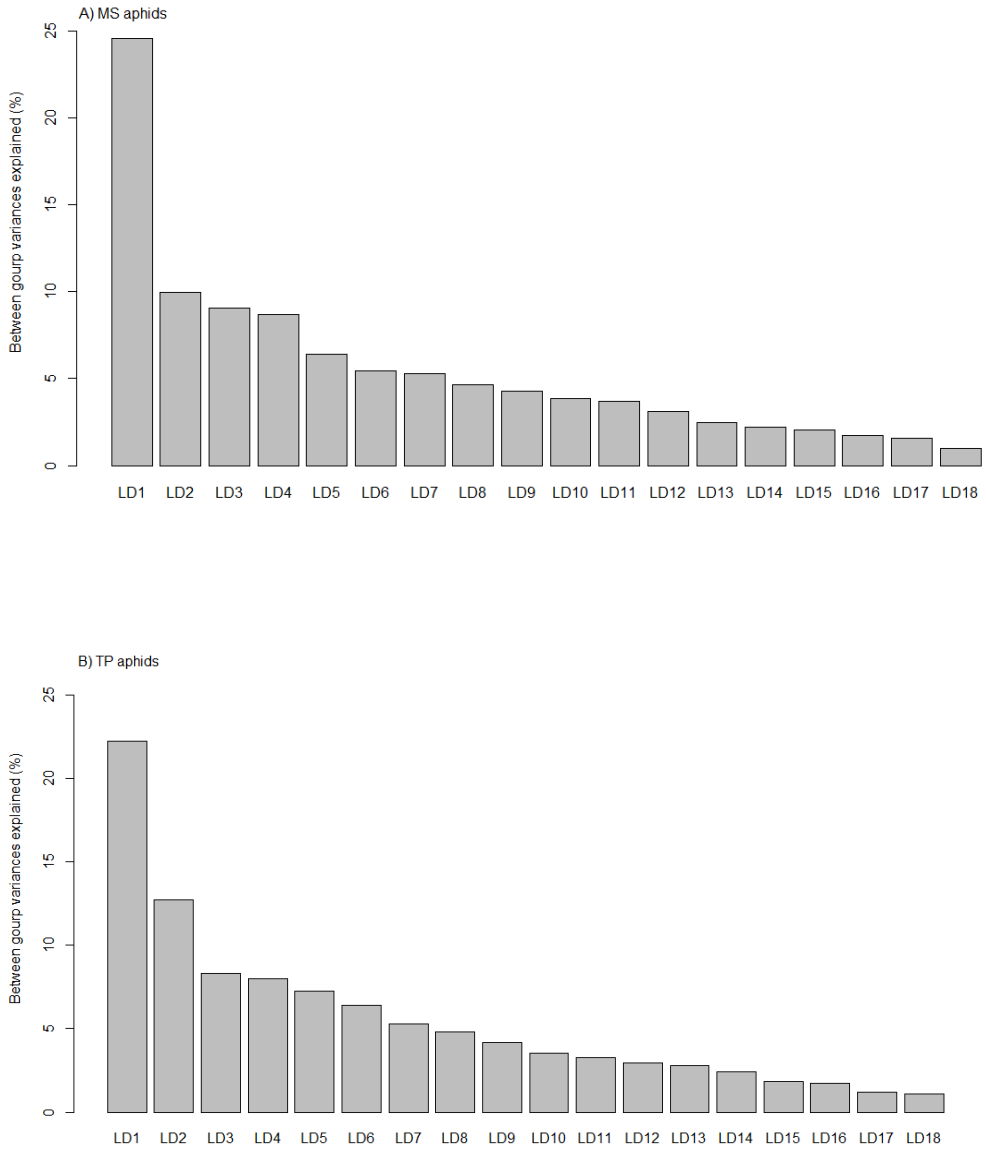


Figure 10.1: Proportion of between group variance explained by the different LD axes of LDA for A) MS and B) TP aphids. Overall LD1 explained approximately double the variance of LD2 with subsequent LD axes becoming less important.

Appendix 11: LDA loading scores

Table A11.1:Top 5 and bottom 5 LD1 and LD2 loading scores

EPG measurement	loadings	Interpretation
LD1		
Number of E1 longer than 10 minutes followed by E2	0.604925	Number of long salivation events followed by a feeding event
Number of C	0.108168	Number of times stylet enters the apoplast (from within or outside the plant)
Number of probes during the 3rd hour	0.102659	Number of probes in the 3rd hour
Number of probes during the 5th hour	0.102571	Number of probes in the 5th hour
Number of probes during the 1st hour	0.086018	Number of probes in the 1st hour
...		
Number of F during the 3rd hour	-0.22362	Number of waveform F in the 3rd hour
Number of probes	-0.22739	Total number of probes
Number of F during the 2nd hour	-0.29128	Number of waveform F in the 2nd hour
Number of G	-0.34162	Total number of G
Number of F during the 4th hour	-0.58348	Number of waveform F in the 4th hour
LD2		
Number of F	0.758004	Total number of F
Number of probes during the 1st hour	0.19946	Number of probes in the 1st hour
Number of probes during the 6th hour	0.19623	Number of probes in the 6th hour
Number of probes during the 4th hour	0.192077	Number of probes in the 4th hour
Number of probes during the 5th hour	0.180692	Number of probes in the 5th hour
...		
Number of F during the 2nd hour	-0.44876	Number of waveform F in the 2nd hour
Number of F during the 3rd hour	-0.54616	Number of waveform F in the 3rd hour
Number of F during the 5th hour	-0.57835	Number of waveform F in the 5th hour
Number of F during the 4th hour	-0.90451	Number of waveform F in the 4th hour

Appendix 12: Correlation between LD1 and PC1 scores for clones of the same race

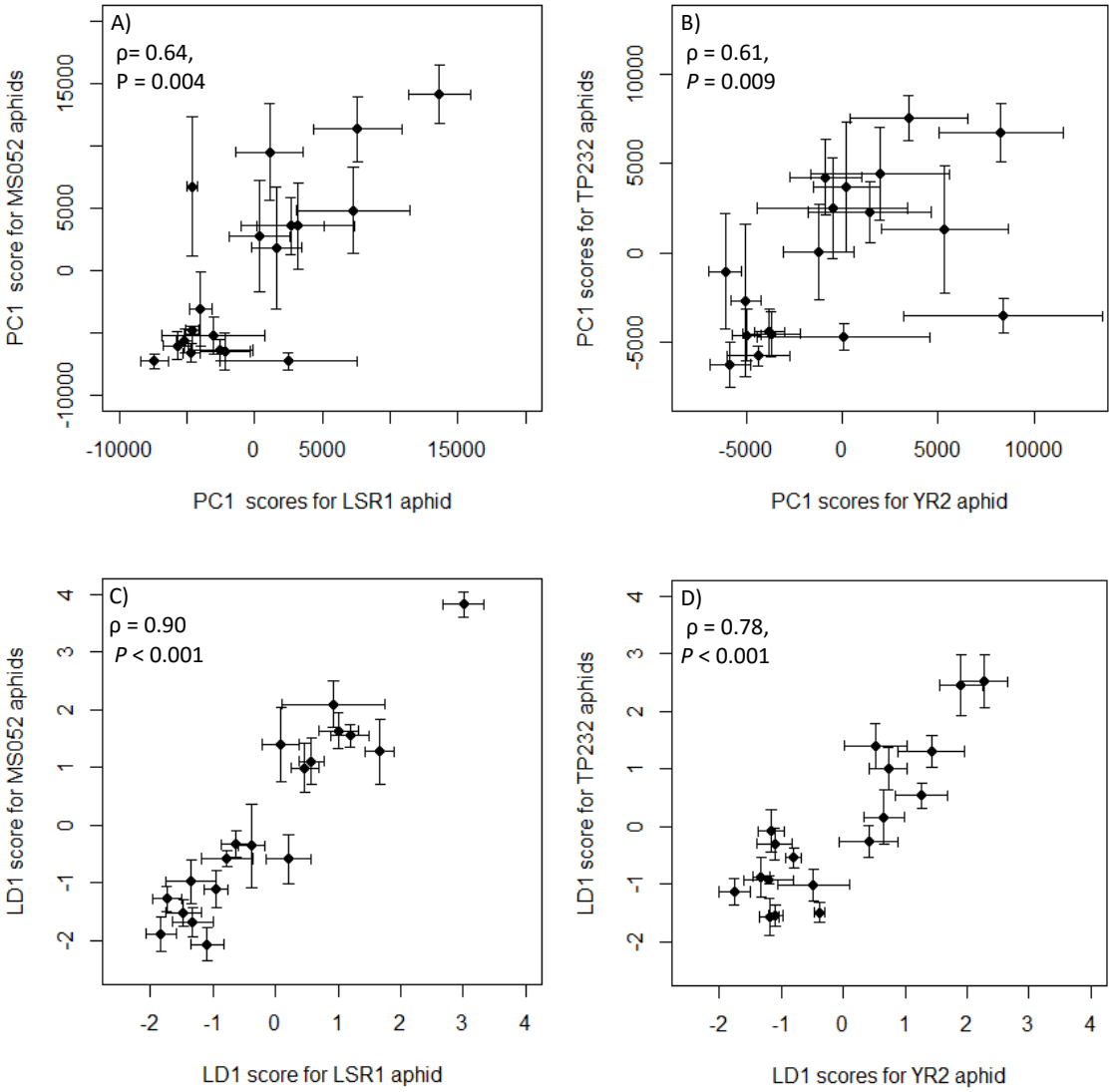


Figure A.12.1: Correlation between clones YR2 and TP232 from the TP aphid race. Each individual waveform measurement was tested: A) Total duration of E1, B) Total duration of E2, C) Number of rpd, D) Number of probes, E) Time to first E2>10 min. Significance tested with Spearman's rank correlation coefficient. Mean ± SEM . From 3 to 12 individuals per clone.

Appendix 13: Block design for species EPG in Chapter 3

Table A13.1: Number code given to each species

Number code	Plant Species	Number code	Plant Species
1	M. arabica	11	T. striatum
2	T. rubens	12	T. nigrescens
3	T. semipilosum	13	M. laciniata
4	T. dubium	14	M. lupulina
5	M. orbicularis	15	*T. pratense*
6	M. minima	16	T.ochroleucum
7	T. palladium	17	M. truncatula
8	M. littoralis	18	T. repens
9	M. tornata	19	M. sativa
10	M. turbinata		

Table 13.2: Five week block design used for EPG experiment.

		block i	block ii	block iii	block iv	block v	block vi
		ia	**iia**	**iiia**	**Iva**	**va**	**via**
Week block a	Monday	1	17	1	9	14	5
		2	15	3	8	12	8
		3	13	5	7	19	7
		19	11	7	6	4	13
	Tuesday	4	9	19	5	16	3
		5	7	9	4	15	2
		6	1	11	3	18	17
		7	19	13	2	11	1
	Wednesday	8	3	15	1	10	6
		9	5	17	19	9	19
		ib	**iib**	**iiib**	**Ivb**	**vb**	**vib**
Week block b	Wednesday	10	18	2	18	19	9
		11	16	4	17	6	10
	Thursday	12	14	6	16	1	11
		19	12	8	15	17	18
		13	10	19	14	2	15
		14	8	10	13	3	16
	Friday	15	6	12	12	13	4
		16	19	14	11	7	19
		17	4	16	10	8	12
		18	2	18	19	5	14

Appendix 14: Mass spectrometry settings for Chapter 4

MALDI TOF

Machine name:
Waters Synapt G2

(Optimisation tests done using CHCA showed target intensity of 200 and step rate of 30 worked best)

Scan conditions
Polarity = positive
Mode = Set to sensitivity mode
Scan rate = 1 scan per second
Scan duration = 120 seconds
Step rate = 50
Laser energy = 200

Voltage settings
Sample plate = 0
Extraction -= +10
Hexapole = 11
Aperture = 7

Tandem MS

Machine name:

ABI Sciex Qstar Elite

Scan settings
Scan type = product ion
Polarity = positive
Mass range = m/z 50 to m/z190
Scan length = 5 minutes
Cycles = 300
Accumulation time = 1 scan per second

Syringe pump method
Diameter = 2.3mm
Flow rate = 10.0 µL/min

Compound

Declustering Potential= 45.0
FP = 265.0
DP2 = 15.0
CE =30.0
CAD = 4
IRD = 6.0
IRW = 5.0

Source/ Gas
Ion source Gas 1 (Gs1) = 27.0
Ion source Gas 1 (Gs1) = 0.0
Curtain Gas (CUR) = 20.0
Ion spray voltage (IS) = 3500.0
Temperature (TEM) =0.0
Accumulation time = 0.999942 (sec)

Resolution
Ion energy (IE1) = 1.0
Focusing lens (IQ2) = 8.5
Collision cell rod offset (RO2) = 8.5
DC Quad lens horizontal focus (GR) = 7.8
DC Quad lens vertical focus = (TFO) = 9.8
DC Quad lens steering (TST) = -0.5

Detection
Detector (CEM) = 2500.0

Appendix 15: E2 profile accumulative and discrimination profiles

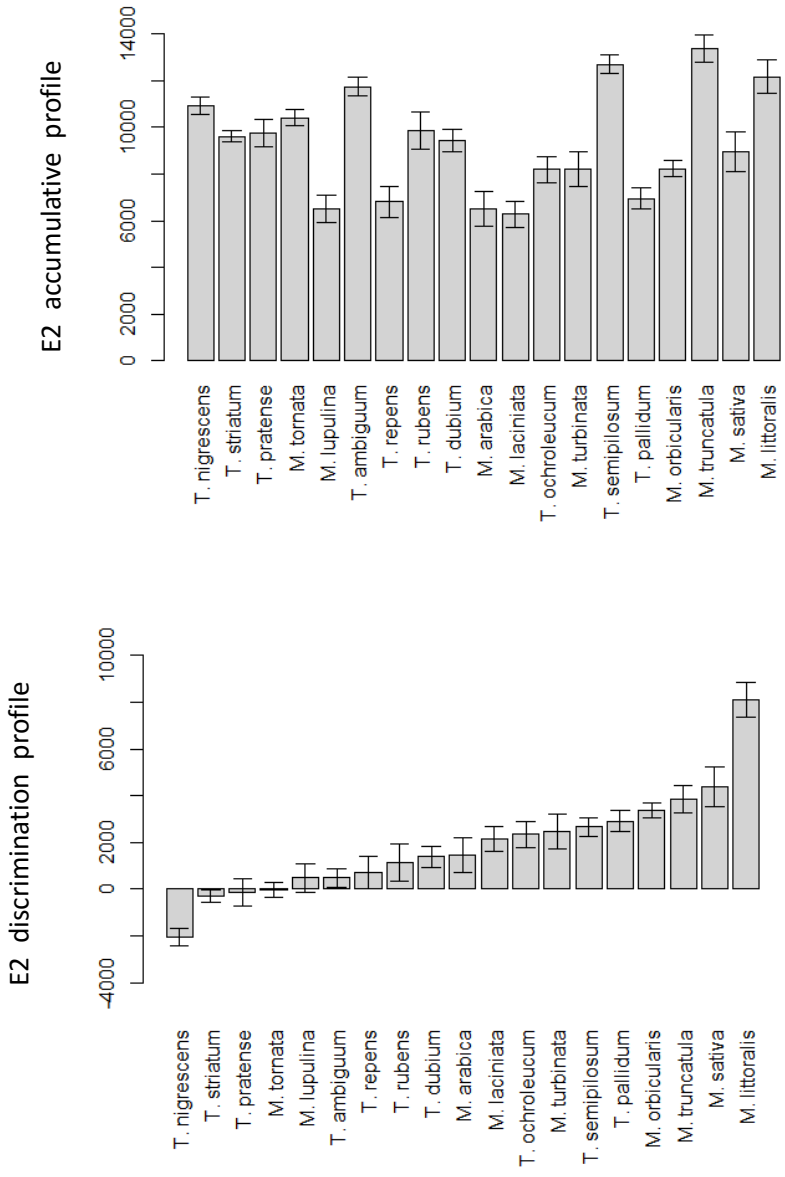


Figure A15.1: EPG E2 profiles of MS and TP aphid races for each plant species. A) Accumulative acceptance profile as measured as total duration of E2 waveform. B) Discrimination profile with positive values indicating overall MS aphid preference, and negative to overall TP aphid acceptance, as measured as total duration of E2 waveform. There were 2 clones per race, 5-11 replicates per clone and plant species.

Appendix 16: PCA of metabolomics profiles of each plant species

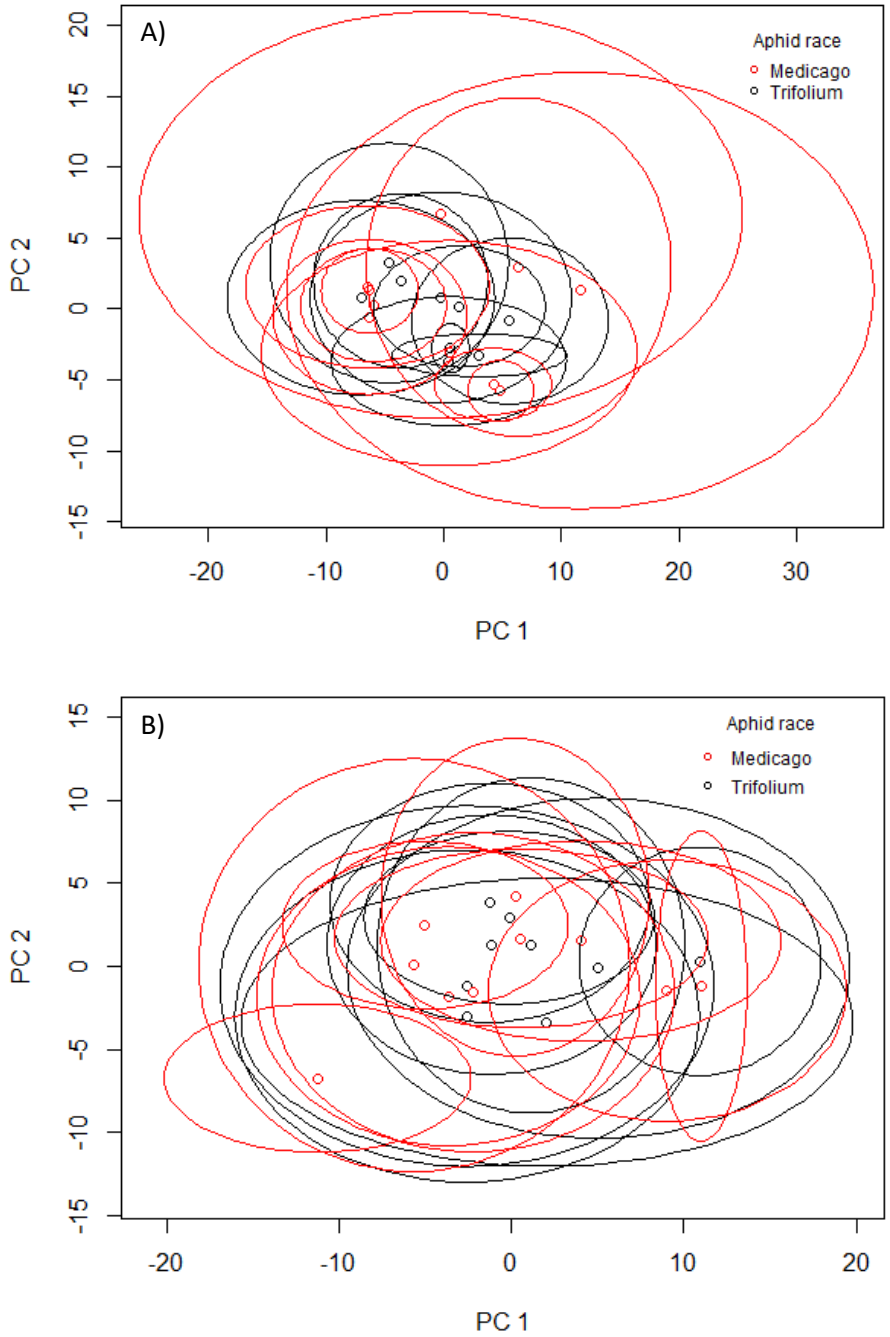


Figure 16.1. PCA of A) Polar metabolomic data and B) Non-Polar metabolic data for each of the plant species. Points represent the mean PCA PC1 and PC2 score and ellipses the respective standard deviations. Plots show significant overlap of the metabolomes of each plant species.

Appendix 17: Top 40 M/z values identified using Random Forest (RF) regression

Table A17.1. Top 40 M/z values based on their median rank value of RF importance (measured as MeanDecreaseGini)

Nonpolar, E2 profile MS aphid – TP aphid		**Polar, E2 profile MS aphid – TP aphid**		**Nonpolar, E2 profile MS aphid + TP aphid**		**Polar, E2 profile t MS aphid + TP aphid**		**Non polar, lda profile MS aphid – TP aphid**		**Polar, lda profile MS aphid - TP aphid**		**Nonpolar, lda profile MS aphid + TP aphid**		**Polar, lda profile MS aphid + TP aphid**	
m/z value	median rank	**m/z value**	median rank	**m/z value**	median rank	**m/z value**	median rank	**m/z value**	median rank	**m/z value**	median rank	**m/z value**	median rank	**m/z value**	median rank
166	2245	**182**	2252	**166**	2236	**250**	2240	**182**	2250	**182**	2253	**137**	2230	**137**	2252
182	2245	**183**	2249	**182**	2232	**176**	2237.5	**166**	2249	**183**	2252	**189**	2230	**250**	2244
292	2229	**331**	2247	**250**	2225	**331**	2234.5	**269**	2249	**309**	2247	**250**	2225	**417**	2234
269	2226	**484**	2247	**331.2**	2210	**182**	2228	**292**	2248	**166**	2245	**166**	2224	**184**	2232
250	2221	**309**	2246	**234**	2208	**137**	2226	**291**	2247	**345**	2245	**341.2**	2224	**176**	2231
291	2217	**137**	2245	**266**	2205	**138**	2222	**182.6**	2246	**363**	2244	**297.2**	2221.5	**345**	2231
138	2209	**363**	2242	**471.2**	2205	**229.2**	2222	**285**	2244	**361**	2242	**234**	2221	**477**	2231
212	2202	**385**	2241	**292**	2201	**301**	2220	**184.6**	2242	**385**	2242	**445**	2219	**435**	2226
331.2	2201.5	**250**	2236	**138**	2198	**184**	2218	**177**	2238	**145.2**	2240	**191**	2217.5	**266**	2224
471.2	2199.5	**196**	2235	**212**	2198	**487**	2218	**138.6**	2234	**331**	2240	**204.6**	2217	**325**	2222
234	2197	**357.2**	2235	**445**	2192	**417**	2217	**307**	2234	**341**	2240	**352**	2217	**379**	2222
160	2196	**176**	2234	**160**	2191	**435**	2217	**339.2**	2234	**325**	2238	**195**	2214	**196**	2221
339.2	2194.5	**325**	2234	**269**	2190	**266**	2215	**167**	2227	**347**	2235	**395**	2211	**229.2**	2221
395	2194	**166**	2232	**373**	2188	**345**	2211.5	**308**	2225	**484**	2232	**386.2**	2210	**189.2**	2220
184.6	2192.5	**344**	2232	**477**	2188	**347**	2211	**250**	2221	**196**	2230	**361**	2209	**331**	2220
167	2191	**136**	2231	**220.2**	2187	**183**	2210.5	**240.2**	2220	**357.2**	2230	**192**	2202	**363**	2218
477	2191	**348**	2231	**337.2**	2185	**287**	2210.5	**144.6**	2217	**176**	2229	**425.2**	2201	**119**	2216

373	2190	**292**	2229	**395**	2180	**189.2**	2205	**206**	2217	**271**	2226	**266**	2200.5	**418**	2216
266	2189	**347**	2222	**232**	2179	**222**	2201.5	**212**	2213	**401**	2225	**410.2**	2199.5	**487**	2216
337.2	2184.5	**417**	2219	**192**	2178	**353**	2200.5	**180**	2212	**324**	2224	**454.2**	2199.5	**202**	2214
138.6	2183	**229.2**	2218	**167**	2177	**363**	2200	**477**	2211	**202**	2222	**337.2**	2199	**206**	2212
144	2179	**301**	2217	**168**	2177	**206**	2199	**204.6**	2204	**206**	2220	**196**	2197.5	**434**	2210
220.2	2177	**231.2**	2215	**196**	2177	**477**	2196	**395**	2203	**310**	2220	**186**	2196	**182**	2208
192	2175.5	**184**	2214	**146**	2174	**351**	2195	**232**	2202	**362**	2220	**147**	2195	**234**	2207
351	2175	**202**	2214	**178.2**	2168	**325**	2194.5	**160**	2200	**184**	2216	**226**	2195	**347**	2207
180	2174	**319.2**	2214	**184**	2168	**268**	2190.5	**373**	2199	**418**	2216	**182**	2194.5	**351**	2205
184	2171	**435**	2209.5	**222**	2167	**234**	2189	**198**	2197	**346**	2215	**141**	2193	**271**	2204
232	2171	**345**	2208.5	**418**	2167	**271**	2189	**158**	2196	**287**	2214	**206**	2191	**313**	2203
146	2170	**382.2**	2208	**439**	2167	**418**	2188	**196**	2196	**301**	2211	**218**	2188.5	**138**	2202
176	2170	**287**	2205	**411**	2165	**145.2**	2187.5	**247.2**	2196	**250**	2209	**199**	2188	**231.2**	2199
445	2169.5	**206**	2204	**218**	2164	**343**	2186	**322.2**	2196	**229.2**	2207	**202.6**	2186.5	**294**	2199
218	2169	**266**	2203	**176**	2163	**196**	2185	**270**	2195	**266**	2206	**212**	2186	**222**	2197
196	2168.5	**268**	2202	**195**	2163	**237**	2185	**174.6**	2194	**344**	2206	**331.2**	2184.5	**493**	2197
177	2167.5	**418**	2200	**144**	2162	**166**	2184.5	**176**	2194	**212.8**	2205	**220.2**	2184	**367**	2194
418	2167	**477**	2200	**204.6**	2162	**212.8**	2182	**234**	2193	**222**	2205	**270**	2182.5	**382.2**	2191
352	2166.5	**311**	2199	**191**	2161	**367**	2182	**213.6**	2192	**323**	2205	**197**	2182	**178.2**	2189
178.2	2165	**351**	2197	**144.6**	2160	**269**	2181.5	**337.2**	2192	**138**	2204	**146**	2181	**303**	2187
206	2165	**349**	2195	**190**	2160	**293**	2179.5	**144**	2191	**348**	2204	**184**	2179	**387.2**	2185
411	2165	**138**	2194	**294**	2160	**309**	2179	**435**	2191	**364**	2202	**322.2**	2176.5	**185**	2184
435	2165	**222**	2192	**352**	2160	**373**	2179	**138**	2190	**373**	2200	**236**	2175	**268**	2184

Appendix 18: Correlation of M/z TIC and aphid LD1 discriminative acceptance profiles of key masses identified by random forest models.

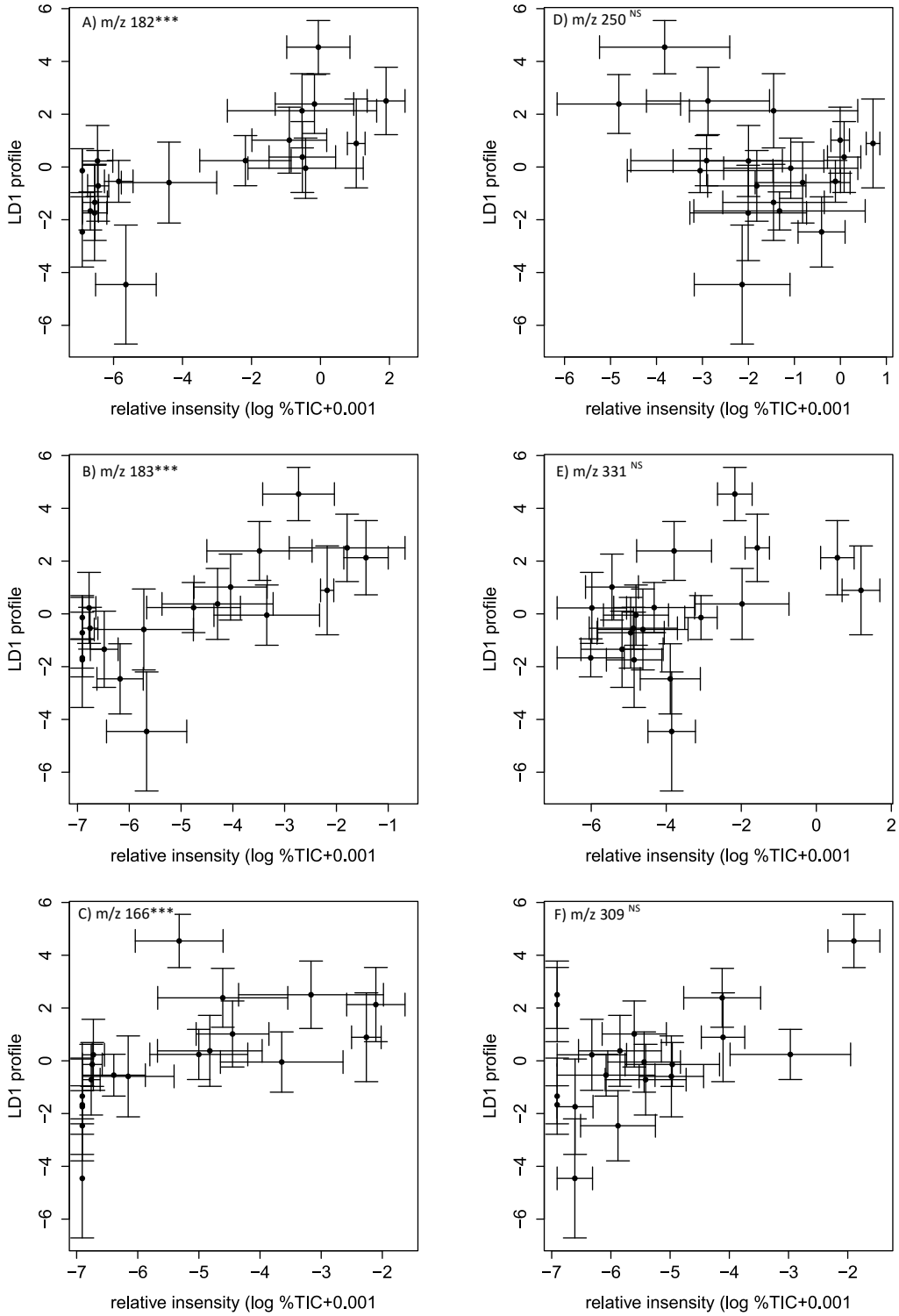


Figure A18.1. Scatter plot of intensity of m/z bin values from plant polar mass spectrometry profile against LD1 score of EPG values. Mean ± SEM. Significance tested with Spearman's rank sum test with FDR correction. * P< 0.05 . ** P< 0.01, *** P< 0.001.

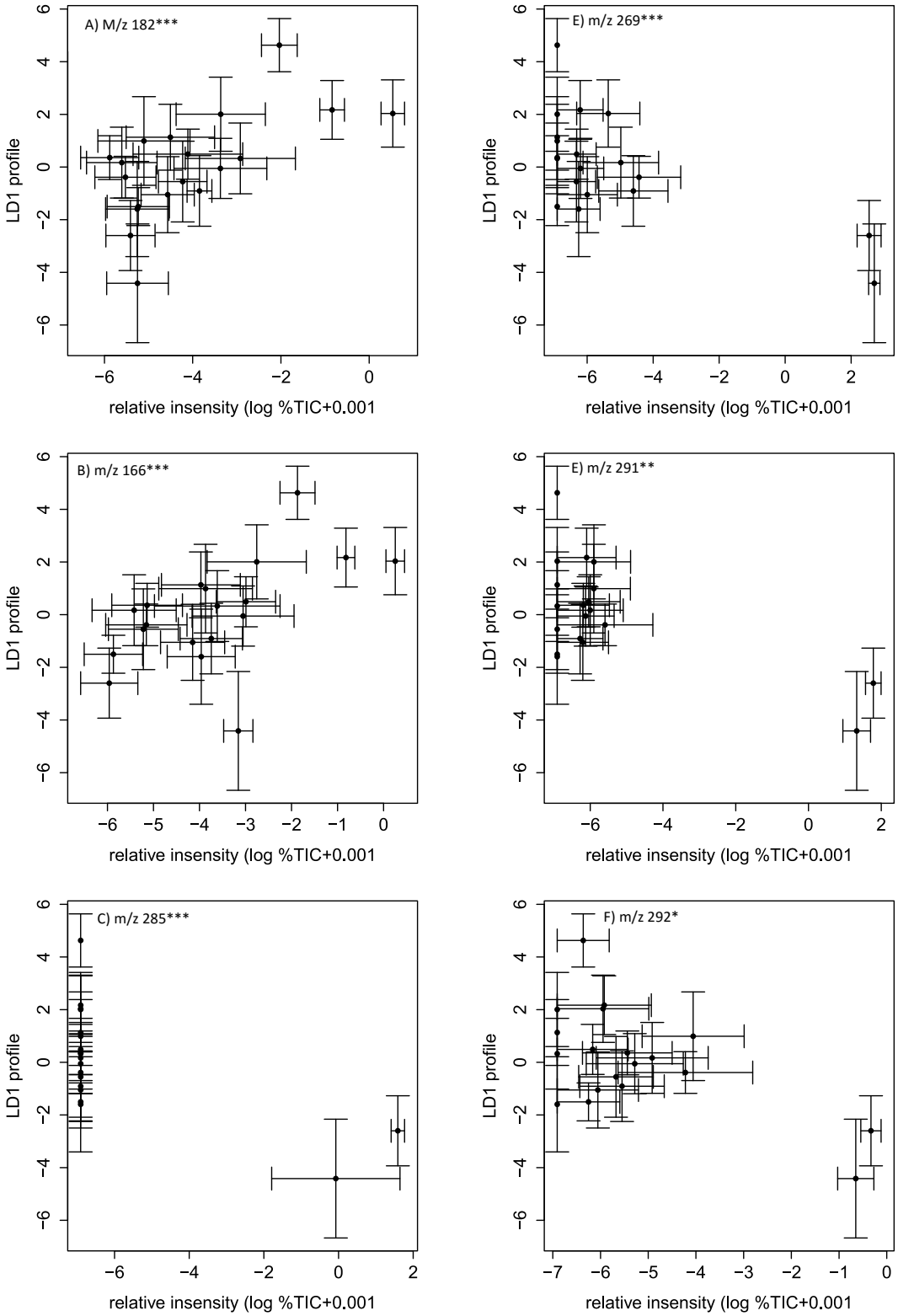


Figure A18.2. Scatter plot of intensity of m/z bin values from plant non-polar mass spectrometry profile against LD1 score of EPG values. Mean ± SEM. Significance tested with Spearman's rank sum test with FDR correction. * P< 0.05 . **P< 0.01, ***P< 0.001.

Appendix 19: Block design for Chapter 4 plant exposure experiment

Table A19.1: Code for the combinations of plant species, aphid used pre-exposure treatments and naive aphid clones used for testing acceptance

code Number	Plant Species	Treatment Aphid	Tested aphid
1	M. sativa	LSR1	LSR1
2	M. sativa	232	232
3	M. sativa	LSR1	232
4	M. sativa	232	LSR1
5	*T. pratense*	LSR1	LSR1
6	*T. pratense*	232	232
7	*T. pratense*	LSR1	232
8	*T. pratense*	232	LSR1

Table A19.2: Block deign for EPG testing of aphid exposed plants

rep	**1**	**rep**	**2**	**rep**	**3**	**rep**	**4**	**rep**	**5**	**rep**	**6**
am	pm	am	pm	am	pm	am	pm	am	pm	am	pm
3	1	5	7	2	6	8	4	7	3	5	1
4	2	6	8	1	5	7	3	8	4	6	2
7	5	1	3	4	8	6	2	1	5	3	7
8	6	2	4	3	7	5	1	2	6	4	8
rep	**7**	**rep**	**8**	**rep**	**9**	**rep**	**10**	**rep**	**11**	**rep**	**12**
am	pm	am	pm	am	pm	am	pm	am	pm	am	pm
3	4	8	7	2	6	5	1	1	5	2	6
1	2	6	5	4	8	7	3	3	7	4	8
7	8	4	3	1	5	8	2	6	2	5	1
5	6	2	1	3	7	6	4	8	4	7	3

Appendix 20: Significance testing for mixed models from Chapter 4.

Table 20.1: Type II Wald chisquare test of LMM and GLMM models that include morning and afternoon (am.pm) as a fixed factor.

	M sativa			T. pratense		
Factors	χ^2	DF	P-value	χ^2	DF	P-value
Duration of E2						
Aphid treatment	9.55	2	<0.001	18.33	2	<0.001
Aphid tested	71.42	1	<0.001	35.27	1	<0.001
Aphid treatment x aphid tested	7.57	2	0.023	1.36	2	0.507
Am.pm	0.17	1	0.684	1.26	1	0.261
Number of Probes						
Aphid treatment	5.60	2	0.061	0.41	2	0.814
Aphid tested	54.40	1	<0.001	4.85	1	0.028
Aphid treatment x aphid tested	2.27	2	0.322	0.23	2	0.893
Am.pm	0.20	1	0.658	1.07	1	0.301
PC1 score						
Aphid treatment	5.11	2	0.078	22.22	2	<0.001
Aphid tested	70.36	1	<0.001	3.48	1	0.062
Aphid treatment x aphid tested	1.37	2	0.504	0.47	2	0.790
Am.pm	1.95	1	0.162	0.37	1	0.541

Appendix 21: Machine setting for LC-ESI-TOF-MS

Machine name:

ABI Sciex Qstar Elite

Scan settings
Scan type = TOF MS
Polarity = positive
Mass range = m/z 40 to m/z 700
Scan length = 1 minutes

Syringe pump method
Diameter = 2.3mm
Flow rate = 10.0 µL/min

Compound
Declustering Potential= 120
FP = 265.0
DP2 = 15.0
IRD = 6.0
IRW = 5.0

Source/ Gas
Ion source Gas 1 (Gs1) = 27.0
Ion source Gas 1 (Gs1) = 0.0
Curtain Gas (CUR) = 20.0
Ion spray voltage (IS) = 3500.0
Temperature (TEM) =0.0
Accumulation time = 0.999942 (sec)

Resolution
Ion energy (IE1) = 1.0
Focusing lens (IQ2) = 8.5
Collision cell rod offset (RO2) = 8.5
DC Quad lens horizontal focus (GR) = 7.8
DC Quad lens vertical focus = (TFO) = 9.8
DC Quad lens steering (TST) = -0.5

Detection
Detector (CEM) = 2500.0

Appendix 22: Tyramine tandem mass spectrometry fragmentation pattern

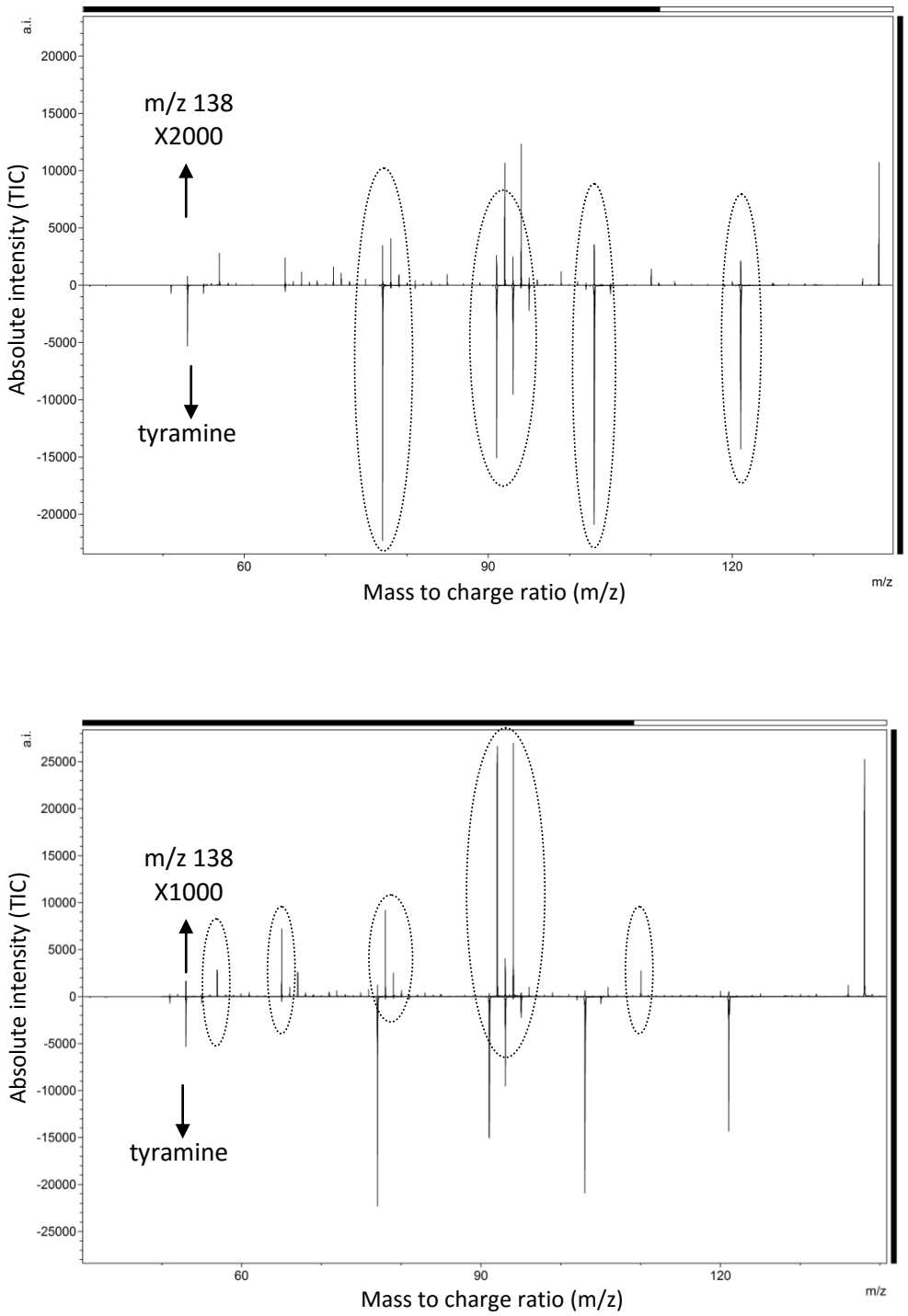


Figure 21.1: Tandem mass spec fragmentations patterns of m/z 138 from two polar samples against tyramine standard. (A) sample showing fragments that match and do not match L- tyramine standard (dotted circles). (B) sample with no matching fragment to tyramine but showing the same unmatched fragments. The unmatched fragments are caused by fragments of another compound or compounds also detected at m/z 138.

Appendix 23: Artificial diet formulation for aphids

Each of the stock solutions was prepared separately (Table A23.1-A23.5) then mixed on the day of experimentation. Once the stock solutions are mixed mixed pH should be between 7- 7.5. All compounds are diluted using diluted water (d.w.).

Table A23.1: Amino acid stock solution. Dissolved in 50 ml d.w. and divide into 10 x 5 ml lots

Amino acid	**g/50ml**
alanine	0.017
asparagine	0.063
aspartate	0.064
cysteine	0.012
glutamate	0.042
glutamine	0.084
Glycine	0.004
proline	0.023
serine	0.020
arginine	0.084
histidine	0.045
isoleucine	0.038
leucine	0.038
lysine	0.042
methionine	0.014
threonine	0.034
tryptophan	0.020
valine	0.013

Table A.23.2: Mineral stock solution. Dissolved in 10 ml d.w. and divided into 0.1 ml lots

Mineral	**mg/0.1 ml**
$FeCl_3.6H_2O$	11
$CuCl_2.4H_2O$	2
$MnCl_2.6H_2O$	4
$ZnSO_4$	17

***Table A23.3: Vitamins stock*.** Dissolved
In 5 ml d.w. and divided into 0.5 ml lots

Vitamin	**mg/0.5 ml**
biotin	0.1
pantothenate	5
folic acid	2
nicotinic acid	10
pyridoxine	2.5
thiamine	2.5
choline	50
myo-inositol	50

***Table A23.4: Sucrose mix*.** Made-up on
day of experimentation. Dissolved into 3 ml d.w.

Compound	**mg/3ml**
ascorbic acid	10
citric acid	1
$MgSO_4.7H_2O$	20
	g/3 ml
Sucrose	0.85

***Table A23.5: Phosphate-mix*.** Made-up on day
of preparation. Dissolved in 1 ml d.w.

Compound	**mg/1ml**
K_2PO_4	115

Appendix 24: Block design used for Chapter 4 choice experiment.

row	Block 1	Block 2	Block 3
1	1a	3b	2b
2	1b	3a	2a
3	2a	1a	3b
4	2b	1b	3a
5	3a	2a	1a
6	3b	2b	1b
7	1b	3a	2a
8	1a	3b	2b
9	2b	1b	3a
10	2a	1a	3b

row	Block 4	Block 5	Block 6	Block 7
1	1a	3a	1b	3b
2	1b	3b	1a	3a
3	2a	2a	2b	2b
4	2b	2b	2a	2a
5	3a	1a	3b	1b
6	3b	1b	3a	1a
7	2b	2b	2a	2a
8	2a	2a	2b	2b
9	3b	1b	3a	1a
10	3a	1a	3b	1b
11	1b	3b	1a	3a
12	1a	3a	1b	3b

Figure A24.1: Block design for choice tests between three diets of differing. *concentration against control diets, for two aphids types.* The three different diet concentrations are represented by the numbers “1”, “2” and “3”. Aphid clones MS_LSR1 are (“a") and Tp_232 (“b”) are tested in alternate

References

Agrawal, A. A. (2011). Current trends in the evolutionary ecology of plant defence. *Functional Ecology, 25,* 420-432.

Agrawal, A. F., Feder, J. L., and Nosil, P. (2011). Ecological divergence and the origins of intrinsic postmating isolation with gene flow. *International Journal of Ecology, 2011, 1-15*.

Akhtar, Y., Isman, M. B., Niehaus, L. A., Lee, C. H., and Lee, H. S. (2012). Antifeedant and toxic effects of naturally occurring and synthetic quinones to the cabbage looper, *Trichoplusia ni*. *Crop Protection*, *31*, 8-14.

Ali, J. G., and Agrawal, A. A. (2014). Asymmetry of plant-mediated interactions between specialist aphids and caterpillars on two milkweeds. *Functional Ecology, 28,* 1404-1412.

Andrews, R. S. and Prideham, J.B. (1965) Structure of DOPA glucoside from *Vicia faba*. *Nature*, *205,* 123-1214.

Arimura, G. I., Matsui, K. and Takabayashi, J. (2009) Chemical and molecular ecology of herbivore-induced plant volatiles: proximate factors and their ultimate functions. *Plant Cell Physiology, 50,* 911-923.

Arimura, G. I., Ozawa, R., and Maffei, M. E. (2011). Recent advances in plant early signaling in response to herbivory. *International Journal of Molecular Sciences*, *12,* 3723-3739.

Awmack, C. S. and Leather, S. R. (2002). Host plant quality and fecundity in herbivorous insects. *Annual Review of Entomology*, 47, 817-844.

Ballhorn, D. J., Kautz, S. and Lieberei, R. (2010). Comparing responses of generalist and specialist herbivores to various cyanogenic plant features. *Entomologia Experimentalis et Applicata*, *134*, 245-259.

Barron, A. B., Søvik, E., and Cornish, J. L. (2010). The roles of dopamine and related compounds in reward-seeking behavior across animal phyla. *Frontiers in Behavioral Neuroscience*, *4*, 163.

Barton, N. H. (2010). What role does natural selection play in speciation?. *Philosophical Transactions of the Royal Society London B: Biological Sciences*, *365*, 1825-1840.

Bates, D. (2005). Fitting Linear Models in R Using the lme4 Package. *R News*, 5, 27–30.

Berlocher, S. H., and Feder, J. L. (2002). Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology*, *47*, 773-815.

Bernays, E. A., and Funk, D. J. (1999). Specialists make faster decisions than generalists: experiments with aphids. *Proceedings of the Royal Society of London B: Biological Sciences*, *266*, 151-156.

Bino, R. J., Hall, R. D., Fiehn, O., Kopka, J., Saito, K., Draper, J. and Sumner, L. W. (2004). Potential of metabolomics as a functional genomics tool. *Trends in Plant Science*, *9*, 418-425.

Bisby, F. (1994). *Phytochemical dictionary of the Leguminosae*. Chapman and Hall/CRC Press, London.

Bonaventure, G. (2012). Perception of insect feeding by plants. *Plant Biology*, *14*, 872-880.

Braendle, C., Caillaud, M. C. and Stern, D. L. (2005). Genetic mapping of aphicarus–a sex-linked locus controlling a wing polymorphism in the pea aphid (*Acyrthosiphon pisum*). *Heredity*, *94*, 435-442.

Brain, K.R. (1976). Accumulation of L-Dopa in Cultures from *Mucuna pruriens*. *Plant Science Letters*, 7, 157-161.

Brisson, J. A. and Stern, D. L. (2006). The pea aphid, *Acyrthosiphon pisum*: an emerging genomic model system for ecological, developmental and evolutionary studies. *Bioessays*, *28*, 747-755.

Bruce, T. J. and Pickett, J. A. (2011). Perception of plant volatile blends by herbivorous insects–finding the right mix. *Phytochemistry*, *72*, 1605-1611.

Bruce, T. J. A., Wadhams, L. J. and Woodcock, C. M., (2005). Insect host location: a volatile situation. *Trends in Plant Science*, 10, 269–274.

Brunissen, L., Cherqui, A., Pelletier, Y., Vincent, C. and Giordanengo, P. (2009). Host-plant mediated interactions between two aphid species. *Entomologia Experimentalis et Applicata*, *132*, 30-38.

Bush, G. L. (1969). Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera, Tephritidae). *Evolution*, *23*, 237-251.

Bush, G. L. (1975). Modes of animal speciation. *Annual Review of Ecology and Systematics*, *6*, 339-364.

Bush, G. L., and Butlin, R. K. (2004). *Sympatric speciation in insects*. In: Dieckmann, U., Doebeli, M., Metz, J. A. J. and Tautz, D., eds *Adaptive Speciation*, Cambridge University Press, Cambridge, pp. 229–248.

Caillaud, M. C., and Via, S. (2000). Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. *The American Naturalist*, *156*, 606-621.

Caillaud, M. C., and Via, S. (2012). Quantitative genetics of feeding behavior in two ecological races of the pea aphid, *Acyrthosiphon pisum*. *Heredity*, *108*, 211-218.

Carolan, J. C., Caragea, D., Reardon, K. T., Mutti, N. S., Dittmer, N., Pappan, K., Cui, F., Castaneto, M., Poulain, J., Dossat, C., Tagu, D., Reese, J.C., Reeck G.R., Wilkinson, T.L. and Edwards, O. R. (2011). Predicted effector molecules in the salivary secretome of the pea aphid (*Acyrthosiphon pisum*): a dual transcriptomic/proteomic approach. *Journal of Proteome Research*, *10*, 1505-1518.

Carolan, J. C., Fitzroy, C. I., Ashton, P. D., Douglas, A. E. and Wilkinson, T. L. (2009). The secreted salivary proteome of the pea aphid *Acyrthosiphon pisum* characterised by mass spectrometry. *Proteomics*, *9*, 2457-2467.

Carrillo, E., Rubiales, D. and Castillejo, M. A. (2014). Proteomic Analysis of Pea (*Pisum sativum* L.) Response During Compatible and Incompatible Interactions with the Pea Aphid (*Acyrthosiphon pisum* H.). *Plant Molecular Biology Reporter*, *32*, 697-718.

Chapman, R. F., Stephen, S. J. and Douglas, A.E. (2013) *The Insects: Structure and Function*. Cambridge University Press, Cambridge.

Chaudhary, R., Atamian, H. S., Shen, Z., Briggs, S. P. and Kaloshian, I. (2014). GroEL from the endosymbiont *Buchnera aphidicola* betrays the aphid by triggering plant defense. *Proceedings of the National Academy of Sciences*, *111*, 8919-8924.

Cherqui, A. and Tjallingii, W. F. (2000). Salivary proteins of aphids, a pilot study on identification, separation and immunolocalisation. *Journal of Insect Physiology*, *46*, 1177-1186.

Comeault, A. A., Soria-Carrasco, V., Gompert, Z., Farkas, T. E., Buerkle, C. A., Parchman, T. L. and Nosil, P. (2014). Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in *Timema cristinae*. *The American Naturalist*, *183*, 711-727.

Compton, S. G. (2002). *Sailing with the wind: dispersal by small flying insects*. In Bullock, J. M. , Kenward, R. E. and Hails, R. S. eds. *Dispersal ecology*. Blackwell, Oxford. pp.113-133.

Coyne, J. A. and Orr, H. A. (2004). *Speciation*. Sinauer Associates, Sunderland.

Cruaud, A., Rønsted, N., Chantarasuwan, B., Chou, L. S., Clement, W. L., Couloux, A., Cousins, B., Genson, G., Harrison, R. D., Hanson, P. E., Hossaert-McKey, M., Jabbour-Zahab R., Jousselin E., Kerdelhue, C., Kjellberg F., Lopez-Vaamonde, C., Peebles J., Peng, Y-Q., Santinelo Pereira, R. A., Schramm T., Ubaidillah, R., Van Noort, S., Weiblen, G. D., Yang, D-R., Yodpinyanee, A., Libeskind-Hadas, R., Cook J. M., Rasplus J-Y. and Savolainen, V. (2012). An extreme case of plant–insect codiversification: figs and fig-pollinating wasps. *Systematic Biology*, 6, 1029-1047.

De Bruyne, M. and Baker, T. C. (2008). Odor detection in insects: volatile codes. *Journal of Chemical Ecology*, *34*, 882-897.

de Ilarduya, O. M., Xie, Q. and Kaloshian, I. (2003). Aphid-induced defense responses in Mi-1-mediated compatible and incompatible tomato interactions. *Molecular Plant-Microbe Interactions*, *16*, 699-708.

de Vos, M., Kim, J. H. and Jander, G. (2007). Biochemistry and molecular biology of Arabidopsis–aphid interactions. *Bioessays*, *29*, 871-883.

Del Campo, M. L., Via, S. and Caillaud, M. C. (2003). Recognition of host-specific chemical stimulants in two sympatric host races of the pea aphid *Acyrthosiphon pisum*. *Ecological Entomology*, *28*, 405-412.

Dixon, A.F.G. (1998) *Aphid ecology: an optimization approach, Second edition*. Chapman and Hall, London.

Donovan, M. P., Nabity, P. D. and DeLucia, E. H. (2013). Salicylic acid-mediated reductions in yield in *Nicotiana attenuata* challenged by aphid herbivory. *Arthropod-Plant Interactions*, *7*, 45-52.

Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K. S., Haines, R. R., Tchalukov, K. A., Labeda, D. P., Kelleher, N.L. and Metcalf, W. W. (2014). A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nature Chemical Biology, 10, 963-968.*

Douglas, A. E., Price, D. R. G., Minto, L. B., Jones, E., Pescod, K. V., François, C. L. M. J., Pritchard J. and Boonham, N. (2006). Sweet problems: insect traits defining the limits to dietary sugar utilisation by the pea aphid, *Acyrthosiphon pisum*. *The Journal of Experimental Biology*, *209*, 1395-1403.

Drès, M., and Mallet, J. (2002). Host races in plant–feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *357*, 471-492.

Dugravot, S., Brunissen, L., Létocart, E., Tjallingii, W. F., Vincent, C., Giordanengo, P. and Cherqui, A. (2007). Local and systemic responses induced by aphids in *Solanum tuberosum* plants. *Entomologia Experimentalis et Applicata*, *123*, 271-277.

Duncan, M. W., Roder, H. and Hunsucker, S. W. (2008). Quantitative matrix-assisted laser desorption/ionization mass spectrometry. *Briefings in Functional Genomics and Proteomics*, *7*, 355-370.

Dunn, P.K. (2014). tweedie: Tweedie exponential family models. R package version 2.2.1. R Foundation for Statistical Computing, Vienna, Austria.

Duvaux, L., Geissmann, Q., Gharbi, K., Zhou, J. J., Ferrari, J., Smadja, C. M. and Butlin, R. K. (2015). Dynamics of copy number variation in host races of the pea aphid. *Molecular Biology and Evolution*, *32*, 63-80.

Eastop, V. F. (1971). Keys for the identification of *Acyrthosiphon* (Hemiptera: Aphididae). *Bulletin of the British Museum (Natural History), B. Entomology*, *26*, 1-115.

Ellison, N. W., Liston, A., Steiner, J. J., Williams, W. M. and Taylor, N. L. (2006). Molecular phylogenetics of the clover genus (*Trifolium*-Leguminosae). *Molecular Phylogenetics and Evolution*, *39*, 688-705.

Emden, H. V. (1972). *Aphid technology. With special reference to the study of aphids in the field*. Academic Press, London.

Facchini, P. J., Yu, M. and Penzes-Yost, C. (1999). Decreased cell wall digestibility in canola transformed with chimeric L-tyrosine decarboxylase genes from opium poppy. *Plant Physiology*, *120*, 653-664.

Farrell, B. D. (1998). " Inordinate fondness" explained: Why are there so many beetles? *Science*, *281*, 555-559.

Feder, J. L., Berlocher, S. H., Roethele, J. B., Dambroski, H., Smith, J. J., Perry, W. L., Gavrilovic, V., Filchak, K.E., Rull, J. and Aluja, M. (2003). Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proceedings of the National Academy of Sciences*, *100*, 10314-10319.

Fenselau, C. C. (2013). Rapid characterization of microorganisms by mass spectrometry—What can be learned and how? *Journal of The American Society for Mass Spectrometry*, *24*, 1161-1166.

Ferrari, J. and Godfray, H. C. J. (2006). The maintenance of intraspecific biodiversity: the interplay of selection on resource use and on natural enemy resistance in the pea aphid. *Ecological Research, 21*, 9-16.

Ferrari, J., Godfray, H. C. J., Faulconbridge, A. S., Prior, K. and 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., Scarborough, C. L. and Godfray, H. C. J. (2007). Genetic variation in the effect of a facultative symbiont on host-plant use by pea aphids. *Oecologia,153*, 323-329.

Ferrari, J., Via, S., and Godfray, H. C. J. (2008). Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. *Evolution, 62*, 2508-2524.

Field, K. J. and Lake, J. A. (2011). Environmental metabolomics links genotype to phenotype and predicts genotype abundance in wild plant populations. *Physiologia Plantarum, 142*, 352-360.

Fitzpatrick, B. M., Fordyce, J. A. and Gavrilets, S. (2008). What, if anything, is sympatric speciation? *Journal of Evolutionary Biology, 21*, 1452-1459.

Fitzpatrick, B. M., Fordyce, J. A. and Gavrilets, S. (2009). Pattern, process and geographic modes of speciation. *Journal of Evolutionary Biology, 22*, 2342-2347.

Frantz, A., Simon J.-C, Plantegenest, M. (2006) Temporal habitat variability and the maintenance of sex in host populations of the pea aphid. *Proceedings of the Royal Society of London B: Biological Sciences*, 273, 2887-2891.

Frantz, A., Plantegenest, M., Simon J.-C. (2010) Host races of the pea aphid Acyrthosiphon pisum differ in male wing phenotypes. *Bulletin of entomological research*, 100, 59-66.

Frost, C.J., Mescher, M.C., Carlson, J.E., C.M., De Moraes (2008a) Plant Defense Priming against Herbivores: Getting Ready for a Different Battle. *Plant Physiology*, 146, 818-824

Frost, C.J., Mescher, M.C., Carlson, J.E., C.M., De Moraes (2008b) Why do distance limitations exist on plant-plant signaling via airborne volatiles? *Plant Signal Behaviour*, 3, 466–468.

Funk, D. J., Filchak, K. E. and Feder, J. L. (2002). *Herbivorous insects: model systems for the comparative study of speciation ecology*. In: Etges, W.T., and Noor, M.A., eds. *Genetics of Mate Choice: From Sexual Selection to Sexual Isolation.* Springer, Netherlands, pp. 251-267.

Furch, A. C., van Bel, A. J., Fricker, M. D., Felle, H. H., Fuchs, M. and Hafke, J. B. (2009). Sieve element Ca^{2+} channels as relay stations between remote stimuli and sieve tube occlusion in *Vicia faba*. *The Plant Cell*, *21*, 2118-2132.

Futuyma, D. J. and Mayer, G. C. (1980). Non-allopatric speciation in animals. *Systematic Biology*, *29*, 254-271.

Gao, L. L., Anderson, J. P., Klingler, J. P., Nair, R. M., Edwards, O. R. and Singh, K. B. (2007). Involvement of the octadecanoid pathway in bluegreen aphid resistance in *Medicago truncatula*. *Molecular Plant-Microbe Interactions*, *20*, 82-93.

Gao, L. L., Klingler, J. P., Anderson, J. P., Edwards, O. R. and Singh, K. B. (2008). Characterization of pea aphid resistance in *Medicago truncatula*. *Plant Physiology*, *146*, 996-1009.

Garcia-Brugger, A., Lamotte, O., Vandelle, E., Bourque, S., Lecourieux, D., Poinssot, B., Wendehenne, D. and Pugin, A. (2006). Early signaling events induced by elicitors of plant defenses. *Molecular Plant-Microbe Interactions*, *19*, 711-724.

Gary, G. T., Kim, B. B., McMurtray, A. M. and Nakamoto, B. K. (2013). Case of Levodopa toxicity from ingestion of *Mucuna gigantea*. *Hawai'i Journal of Medicine and Public Health*, *72*, 157.

Gavrilets, S. (2003). Perspective: models of speciation: what have we learned in 40 years? *Evolution*, *57*, 2197-2215.

Gavrilets S. (2004). *Fitness landscapes and the origin of species*. Princeton University Press, Princeton.

Gavrilets, S. (2014). Models of speciation: where are we now? *Journal of Heredity*, 105(S1), 743-755.

Gavrilets, S., Vose, A., Barluenga, M., Salzburger, W. and Meyer, A. (2007). Case studies and mathematical models of ecological speciation. 1. Cichlids in a crater lake. *Molecular Ecolology*, *16*, 2893–2909.

Giordanengo, P., Brunissen, L., Rusterucci, C., Vincent, C., Van Bel, A., Dinant, S., Girousse, C, Faucher, M. and Bonnemain, J. L. (2010). Compatible plant-

aphid interactions: how aphids manipulate plant responses. *Comptes Rendus Biologies*, *333*, 516-523.

Gish, M., Dafni, A. and Inbar, M. (2012). Young aphids avoid erroneous dropping when evading mammalian herbivores by combining input from two sensory modalities. PloS one, 7(4), e32706-e32706.

Goławska, S., Sprawka, I., and Łukasik, I. (2014). Effect of saponins and apigenin mixtures on feeding behavior of the pea aphid, *Acyrthosiphon pisum* Harris. *Biochemical Systematics and Ecology*, *55*, 137-144.

Gonzales, W. L., Ramirez, C. C., Olea, N., and Niemeyer, H. M. (2002). Host plant changes produced by the aphid *Sipha flava*: consequences for aphid feeding behaviour and growth. *Entomologia Experimentalis et Applicata*, *103*, 107-113.

Gripenberg, S., Mayhew, P. J., Parnell, M., and Roslin, T. (2010). A meta-analysis of preference performance relationships in phytophagous insects. *Ecology Letters*, *13*, 383-393.

Gupta, K. J. Igamberdiev, A.U. (2014). *Reactive Oxygen and Nitrogen Species Signaling and Communication in Plants.* Springer, Heidelberg.

Hafke, J. B., Furch, A. C., Fricker, M. D., and van Bel, A. J. (2009). Forisome dispersion in Vicia faba is triggered by Ca^{2+} hotspots created by concerted action of diverse Ca^{2+} channels in sieve element. *Plant Signaling and Behavior*, *4*, 968-972.

Hagel, J. M., and Facchini, P. J. (2013). Benzylisoquinoline alkaloid metabolism–a century of discovery and a brave new world. *Plant and Cell Physiology*, pct020.

Hall, R., Beale, M., Fiehn, O., Hardy, N., Sumner, L., and Bino, R. (2002). Plant metabolomics the missing link in functional genomics strategies. *The Plant Cell*, *14*, 1437-1440.

Hawthorne, D. J., and Via, S. (2001). Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature*, *412*, 904-907.

Herbert R. B. (1985) *The biosynthesis of the isoquinoline alkaloids*. In: Phillipson J.D., Roberts, M. F. and Zenk M. H. eds. *The chemistry of the isoquinoline alakaloids*. Springer, Berlin, pp. 213-228

Hogenhout, S. A., and Bos, J. I. (2011). Effector proteins that modulate plant–insect interactions. *Current Opinion in Plant Biology*, *14*, 422-428.

Hothorn, T. , Bretz,F., and Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal*, 50, 346-363.

Huang, T., Jander, G., and de Vos, M. (2011). Non-protein amino acids in plant defense against insect herbivores: representative cases and opportunities for further functional analysis. *Phytochemistry*, 72, 1531-1537.

Hultman, J., Waldrop, M. P., Mackelprang, R., David, M. M., McFarland, J., Blazewicz, S. J., Jennifer Harden, J. , Turetsky, .R., McGuire, A. D., Shah, .B., VerBerkmoes, N. C., Lee, L. H., Mavrommatis, K. and Jansson, J. K. (2015). Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature*, 521, 208-212.

International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol*, *8*, e1000313.

Jaenike, J. (1990). Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics*, *21*, 243-273.

Jaquiéry, J., Stoeckel, S., Nouhaud, P., Mieuzet, L., Mahéo, F., Legeai, F., Bernard, B., Bonvoisin, A., Vitalis, R. and Simon, J. C. (2012). Genome scans reveal candidate regions involved in the adaptation to host plant in the pea aphid complex. *Molecular Ecology*, *21*, 5251-5264.

Jaouannet, M., Rodriguez, P. A., Thorpe, P., Lenoir, C. J., MacLeod, R., Escudero-Martinez, C. and Bos, J. I. (2014). Plant immunity in plant–aphid interactions. *Frontiers in Plant Science*, *5*, 1-10.

Joyce, A. R. and Palsson, B. Ø. (2006). The model organism as a system: integrating'omics' data sets. *Nature Reviews Molecular Cell Biology*, *7*, 198-210.

Kaiser, B. D., Li, J., Sanford, J. A., Kim, Y. M., Kronewitter, S. R., Jones, M. B., *et al.* (2013). A multi-omic view of host-pathogen-commensal interplay in *Salmonella*-mediated intestinal infection. *PLoS One*, *8*(6), e67155.

Kidd, N. A. C., Lewis, G. B. and Howell, C. A. (1985). An association between two species of pine aphid, *Schizolachnus pineti* and *Eulachnus agilis*. *Ecological Entomology*, *10*, 427-43.

Kessler, A., Halitschke, R., DiezelIan, C., Baldwin, I.T. (2006) Priming of plant defense responses in nature by airborne signaling between Artemisia tridentata and Nicotiana attenuata.

Kuhlisch, C. and Pohnert, G. (2015). Metabolomics in chemical ecology. *Natural Product Reports, 32*, 937-955.

Klingler, J., Creasy, R., Gao, L., Nair, R. M., Calix, A. S., Jacob, H. S., , Owain, R. E. and 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 Physiology*, 137, 1445-1455.

Kordan, B., Lahuta, L., Dancewicz, K., Sądej, W. and Gabryś, B. (2011). Effect of lupin cyclitols on pea aphid probing behaviour. *Journal of Plant Protection Research, 51*, 171-178.

Koul, O. (2008). Phytochemicals and insect control: an antifeedant approach. *Critical Reviews in Plant Sciences, 27*, 1-24.

Kulma, A. and Szopa, J. (2007). Catecholamines are active compounds in plants. *Plant Science, 172*, 433-440.

Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. *R News, 2,* 18-22.

Leiss, K. A., Maltese, F., Choi, Y. H., Verpoorte, R. and Klinkhamer, P. G. (2009). Identification of chlorogenic acid as a resistance factor for thrips in chrysanthemum. *Plant Physiology, 150*, 1567-1575.

Leroy, P. D., Wathelet, B., Sabri, A., Francis, F., Verheggen, F. J., Capella, Q., Thonart, P. and Haubruge, E. (2011). Aphid-host plant interactions: does aphid honeydew exactly reflect the host plant amino acid composition? *Arthropod-Plant Interactions*, 5, 193-199.

Li, Q., Xie, Q. G., Smith-Becker, J., Navarre, D. A. and Kaloshian, I. (2006). Mi-1-mediated aphid resistance involves salicylic acid and mitogen-activated protein kinase signaling cascades. *Molecular Plant-Microbe Interactions, 19*, 655-664.

Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2, 18-22.

Lin, Y., Sun, X., Yuan, Q. and Yan, Y. (2014). Extending shikimate pathway for the production of muconic acid and its precursor salicylic acid in *Escherichia coli*. *Metabolic Engineering, 23*, 62-69.

Linn, C., Feder, J. L., Nojima, S., Dambroski, H. R., Berlocher, S. H. and Roelofs, W. (2003). Fruit odor discrimination and sympatric host race

formation in *Rhagoletis*. *Proceedings of the National Academy of Sciences*, *100*, 11490-11493.

Lokvam, J., Brenes-Arguedas, T., Lee, J. S., Coley, P. D. and Kursar, T. A. (2006). Allelochemic function for a primary metabolite: the case of L-tyrosine hyper-production in *Inga umbellifera* Fabaceae). *American Journal of Botany*, *93*, 1109-1115.

Lokvam, J., Clausen, T. P., Grapov, D., Coley, P. D. and Kursar, T. A. (2007). Galloyl depsides of L-tyrosine from young leaves of *Inga laurina*. *Journal of Natural Products*, *70*, 134-136.

Loxdale, H. D., Hardie, J. I. M., Halbert, S., Foottit, R., Kidd, N. A. and Carter, C. I. (1993). The relative importance of short-and long-range movement of flying aphids. *Biological Reviews*, *68*, 291-311.

Ma, R., Reese, J. C., Black, W. C. and Bramel-Cox, P. (1990). Detection of pectinesterase and polygalacturonase from salivary secretions of living greenbugs, *Schizaphis graminum* (Homoptera: Aphididae). *Journal of Insect Physiology*, *36*, 507-512.

MacNeil, S. M., Johnson, W. E., Li, D. Y., Piccolo, S. R. and Bild, A. H. (2015). Inferring pathway dysregulation in cancers from multiple types of omic data. *Genome Medicine*, *7*, 1-12.

Macoy, D. M., Kim, W. Y., Lee, S. Y. and Kim, M. G. (2015). Biotic stress related functions of hydroxycinnamic acid amide in plants. *Journal of Plant Biology*, *58*, 156-163.

Matsubayashi, K. W., Ohshima, I. and Nosil, P. (2010). Ecological speciation in phytophagous insects. *Entomologia Experimentalis et Applicata*, *134*, 1-27.

Matzkin, L. M., Watts, T. D., Bitler, B. G., Machado, C. A. and Markow, T. A. (2006). Functional genomics of cactus host shifts in *Drosophila mojavensis*. *Molecular Ecology*, *15*, 4635-4643.

Maureira-Butler, I. J., Pfeil, B. E., Muangprom, A., Osborn, T. C. and Doyle, J. J. (2008). The reticulate history of *Medicago* (Fabaceae). *Systematic Biology*, *57*, 466-482.

Maureira-Butler, I. J., Pfeil, B. E., Muangprom, A., Osborn, T. C. and Doyle, J. J. (2008). The reticulate history of *Medicago* (Fabaceae). *Systematic Biology*, *57*, 466-482.

Mescher, M.C. and De Moraes, C.M. (2015) Role of plant sensory perception in plant-animal interactions. *Journal of Experimental Botany*. *66*, 425-433.

Mittler, T. E. and Dadd, R. H. (1963a). Studies on the artificial feeding of the aphid *Myzus persicae* (Sulzer)—II. Relative survival, development, and larviposition on different diets. *Journal of Insect Physiology*, *9*, 741-757.

Mittler, T. E. and Dadd, R. H. (1963b). Studies on the artificial feeding of the aphid *Myzus persicae* (Sulzer)—I. Relative uptake of water and sucrose solutions. *Journal of Insect Physiology*, *9*, 623-645.

Mohase, L. and van der Westhuizen, A. J. (2002). Salicylic acid is involved in resistance responses in the Russian wheat aphid-wheat interaction. *Journal of Plant Physiology*, *159*, 585-590.

Mutti, N. S., Louis, J., Pappan, L. K., Pappan, K., Begum, K., Chen, M. S., Park, Y., Dittmer, N., Marshall, J., Reese, J. C. and Reeck, G. R. (2008). A protein from the salivary glands of the pea aphid, *Acyrthosiphon pisum*, is essential in feeding on a host plant. *Proceedings of the National Academy of Sciences*, *105*, 9965-9969.

Nagasawa, A., Kamada, Y., Kosaka, Y., Arakida, N. and Hori, M. (2014). Catechol- an oviposition stimulant for cigarette beetle in roasted coffee beans. *Journal of Chemical Ecology*, *40*, 452-457.

Newman, M. A., von Roepenack-Lahaye, E., Parr, A., Daniels, M. J. and Dow, J. M. (2001). Induction of hydroxycinnamoyl-tyramine conjugates in pepper by *Xanthomonas campestris*, a plant defense response activated by hrp gene-dependent and hrp gene-independent mechanisms. *Molecular Plant-Microbe Interactions*, *14*, 785-792.

Nguyen, D. D., Wu, C. H., Moree, W. J., Lamsa, A., Medema, M. H., Zhao, X., *et al.* (2013). MS/MS networking guided analysis of molecule and gene cluster families. *Proceedings of the National Academy of Sciences*, *110*, E2611-E2620.

Nishida, R. (2014). Chemical ecology of insect–plant interactions: ecological significance of plant secondary metabolites. *Bioscience, Biotechnology, and Biochemistry*, *78*, 1-13.

Nosil, P. (2012). *Ecological speciation*. Oxford University Press, Oxford.

Nosil, P., Crespi, B. J. and Sandoval, C. P. (2002). Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature*, *417*, 440-443.

Nouhaud, P., Peccoud, J., Mahéo, F., Mieuzet, L., Jaquiéry, J. and Simon, J. C. (2014). Genomic regions repeatedly involved in divergence among plant-specialized pea aphid biotypes. *Journal of Evolutionary Biology*, *27*, 2013-2020.

Patil, R. R., Gholave, A. R., Jadhav, J. P., Yadav, S. R. and Bapat, V. A. (2015). *Mucuna sanjappae* Aitawade et Yadav: a new species of *Mucuna* with promising yield of anti-Parkinson's drug L-DOPA. *Genetic Resources and Crop Evolution, 62*, 155-162.

Oerke, E-C, (2006) Crop losses to pests. *Journal of Agricultural Science*, 144, 31-43.

Peccoud, J., Huerta, M., Bonhomme, J., Laurence, C., Outreman, Y., Smadja, C. M. and Simon, J. C. (2014). Widespread host-dependent hybrid unfitness in the pea aphid species complex. *Evolution, 68*, 2983-2995.

Peccoud, J., Ollivier, A., Plantegenest, M. and Simon, J. C. (2009a). A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences,106*, 7495-7500.

Peccoud, J. and Simon, J. C. (2010). The pea aphid complex as a model of ecological speciation. *Ecological Entomology, 35*(s1), 119-130.

Peccoud, J., Simon, J. C., McLaughlin, H. J. and Moran, N. A. (2009b). Post-Pleistocene radiation of the pea aphid complex revealed by rapidly evolving endosymbionts. *Proceedings of the National Academy of Sciences, 106*, 16315-16320.

Peccoud, J., Simon, J. C., von Dohlen, C., Coeur d'acier, A., Plantegenest, M., Vanlerberghe-Masutti, F. and Jousselin, E. (2010). Evolutionary history of aphid-plant associations and their role in aphid diversification. *Comptes Rendus Biologies, 333*, 474-487.

Piasecka, A., Jedrzejczak-Rey, N. and Bednarek, P. (2015). Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. *New Phytologist, 206*, 948-964.

Pickett, J. A., Wadhams, L. J., Woodcock, C. M. and Hardie, J. (1992). The chemical ecology of aphids. *Annual Review of Entomology, 37*, 67-90.

Powell, G. and Hardie, J. (2001). The chemical ecology of aphid host alternation: How do return migrants find the primary host plant? *Applied Entomology and Zoology, 36*, 259-267.

Powell, G., Tosh, C. R. and Hardie, J. (2006). Host plant selection by aphids: behavioral, evolutionary, and applied perspectives. *Annual Review of Entomology, 51*, 309-330.

Prado, E. and Tjallingii, W. F. (2007). Behavioral evidence for local reduction of aphid-induced resistance. *Journal of Insect Science, 7*, 48.

Premont, R. T., Gainetdinov, R. R. and Caron, M. G. (2001). Following the trace of elusive amines. *Proceedings of the National Academy of Sciences*, *98*, 9474-9475.

Prosser, W. A. and Douglas, A. E. (1992). A test of the hypotheses that nitrogen is upgraded and recycled in an aphid (*Acyrthosiphon pisum*) symbiosis. *Journal of Insect Physiology*, *38*, 93-99.

Prosser, W. A., Simpson, S. J. and Douglas, A. E. (1992). How an aphid (*Acyrthosiphon pisum*) symbiosis responds to variation in dietary nitrogen. *Journal of Insect Physiology*, *38*, 301-307.

Ramsey, J. S., Elzinga, D. A., Sarkar, P., Xin, Y. R., Ghanim, M. and Jander, G. (2014). Adaptation to nicotine feeding in *Myzus persicae*. *Journal of Chemical Ecology*, *40*, 869-877.

Ramsey, J. S., Rider, D. S., Walsh, T. K., De Vos, M., Gordon, K. H. J., Ponnala, L., Macmil, S. L., Roe, B.A. and Jander, G. (2010) Comparative analysis of detoxification enzymes in *Acyrthosiphon pisum* and *Myzus persicae*. *Insect Molecular Biology,* 19 (Suppl. 2), 155–164.

Rehr, S. S., Janzen, D. H. and Feeny, P. P. (1973). L-DOPA in legume seeds: a chemical barrier to insect attack. *Science*, *181*, 81-82.

Rundle, H. D. and Nosil, P. (2005). Ecological speciation. *Ecology Letters*, *8*, 336-352.

Sadeghi A., Van Damme E.M. and Smagghe, G. (2009) Evaluation of the susceptibility of the pea aphid, *Acyrthosiphon pisum*, to a selection of novel biorational insecticides using an artificial diet. *Journal of Insect Science*, 9, 65.

Sarria, E., Cid, M., Garzo, E. and Fereres, A. (2009). Excel Workbook for automatic parameter calculation of EPG data. *Computers and Electronics in Agriculture*, *67*, 35-42.

Sauvion, N., Charles, H., Febvay, G. and Rahbé, Y. (2004). Effects of jackbean lectin (ConA) on the feeding behaviour and kinetics of intoxication of the pea aphid, *Acyrthosiphon pisum*. *Entomologia Experimentalis et Applicata*, *110*, 31-44.

Schwartzberg, E. G., Böröczky, K. and Tumlinson, J. H. (2011). Pea aphids, *Acyrthosiphon pisum*, suppress induced plant volatiles in broad bean, *Vicia faba*. *Journal of Chemical Ecology*, *37*, 1055-1062.

Schwartzberg, E. G. and Tumlinson, J. H. (2014). Aphid honeydew alters plant defence responses. *Functional Ecology*, *28*, 386-394.

Schwarzkopf, A., Rosenberger, D., Niebergall, M., Gershenzon, J. and Kunert, G. (2013). To feed or not to feed: plant factors located in the epidermis, mesophyll, and sieve elements influence pea aphid's ability to feed on legume species. *PloS one*, *8*, e75298.

Schweiger, R., Heise, A. M., Persicke, M. and Müller, C. (2014). Interactions between the jasmonic and salicylic acid pathway modulate the plant metabolome and affect herbivores of different feeding types. *Plant, Cell and Environment*, *37*, 1574-1585.

Sempruch, C., Leszczyński, B., Wójcicka, A., Makosz, M., Chrzanowski, G. and Matok, H. (2009). Changes in activity of triticale L-tyrosine decarboxylase caused by grain aphid feeding. *Polish Journal of Environmental Studies*, *18*, 901-906.

Sempruch, C., Marczuk, W., Leszczyński, B., Kozak, A., Zawadzka, W., Klewek, A. and Jankowska, J. (2013). Effect of pea aphid infestation on activity of amino acid decarboxylases in pea tissues. *Acta Biologica Cracoviensia Series Botanica*, *55*, 45-50.

Sharma, A., Khan, A. N., Subrahmanyam, S., Raman, A., Taylor, G. S. and Fletcher, M. J. (2014). Salivary proteins of plant-feeding hemipteroids–implication in phytophagy. *Bulletin of Entomological Research*, *104*, 117-136.

Simmonds, M. S. (2003). Flavonoid–insect interactions: recent advances in our knowledge. *Phytochemistry*, *64*, 21-30.

Simon, J. C., Boutin, S., Tsuchida, T., Koga, R., Le Gallic, J. F., Frantz, A., *et al.* (2011). Facultative symbiont infections affect aphid reproduction. *PLoS One*, *6*(7), e21831-e21831.

Simon, J. C., d'Alençon, E., Guy, E., Jacquin-Joly, E., Jaquiéry, J., Nouhaud, P., Peccoud, J., Sugio, A. and Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in Functional Genomics*, elv015.

Smadja, C. and Butlin, R. K. (2009). On the scent of speciation: the chemosensory system and its role in premating isolation. *Heredity*, *102*, 77-97.

Smadja, C. M. and Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, *20*, 5123-5140.

Smadja, C. M., Canbäck, B., Vitalis, R., Gautier, M., Ferrari, J., Zhou, J. J. and Butlin, R. K. (2012). Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution*, *66*, 2723-2738.

Smadja, C., Shi, P., Butlin, R. K. and Robertson, H. M. (2009). Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrthosiphon pisum*. *Molecular Biology and Evolution*, *26*, 2073-2086.

Stork, N. E. (2007). World of insects. *Nature*, 448, 657-658.

Smith, M. A. H. and MacKay, P. A. (1989). Seasonal variation in the photoperiodic responses of a pea aphid population: evidence for long-distance movements between populations. *Oecologia*, *81*, 160-165.

Snyder, M. J., & Glendinning, J. I. (1996). Causal connection between detoxification enzyme activity and consumption of a toxic plant compound. *Journal of Comparative Physiology A*, *179*, 255-261.

Sugimoto, M., Kawakami, M., Robert, M., Soga, T. and Tomita, M. (2012). Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. *Current Bioinformatics*, *7*, 96.

Takemoto, H., Uefune, M., Ozawa, R., Arimura, G. I. and Takabayashi, J. (2013). Previous infestation of pea aphids *Acyrthosiphon pisum* on broad bean plants resulted in the increased performance of conspecific nymphs on the plants. *Journal of Plant Interactions*, *8*, 370-374.

Takemura, M., Nishida, R., Mori, N. and Kuwahara, Y. (2002). Acylated flavonol glycosides as probing stimulants of a bean aphid, *Megoura crassicauda*, from *Vicia angustifolia*. *Phytochemistry*, *61*, 135-140.

Tanaka, S., Han, X. and Kahmann, R. (2015). Microbial effectors target multiple steps in the salicylic acid production and signaling pathway. *Frontiers in Plant Science*, *6*, 349.

Thompson, G. A. and Goggin, F. L. (2006). Transcriptomics and functional genomics of plant defence induction by phloem-feeding insects. *Journal of Experimental Botany*, *57*, 755-766.

Tjallingii, W. F. (1978). Electronic recording of penetration behaviour by aphids. *Entomologia Experimentalis et Applicata*, *24*, 721-730.

Tjallingii W. F. (1987) *Electrical recording of stylet penetration activities*. In: Minks, A.K., Harrewijn, P. eds. *Aphids: Their Biology, Natural Enemies, and Control, Vol 2B*. Elsevier, Amsterdam, pp. 95-108.

Tjangilii, W. F. (1992) *Regulatory Mechanisms in Insect Feeding.* In: Chapman, R. F., de Boer, G. eds. *Regulatory Mechanisms in Insect Feeding*. Chapman and Hall, London, pp. 190-209.

Tjallingii, W. F. (2006). Salivary secretions by aphids interacting with proteins of phloem wound responses. *Journal of Experimental Botany*, *57*, 739-745.

Tjallingii, W. F. and Esch, T. H. (1993). Fine structure of aphid stylet routes in plant tissues in correlation with EPG signals. *Physiological Entomology*, *18*, 317-328.

Tjallingii, W. F. and Gabryś, B. (1999). *Anomalous stylet punctures of phloem sieve elements by aphids*. In: Simpson, S.J., Mordue, A.J., Hardie, J. eds. *Proceedings of the 10th International Symposium on Insect-Plant Relationships*. Springer, Netherlands, pp. 97-103.

Tohge, T. and Fernie, A. R. (2009). Web-based resources for mass-spectrometry-based metabolomics: a user's guide. *Phytochemistry*, *70*, 450-456.

Torchiano M. (2015). *effsize: Efficient Effect Size Computation*. R package version 0.5.4., http://CRAN.R-project.org/package=effsize. R Foundation for Statistical Computing, Vienna, Austria.

Tosh, C. R., Powell, G. and Hardie, J. (2002). Maternal reproductive decisions are independent of feeding in the black bean aphid, *Aphis fabae*. *Journal of Insect Physiology, 48*, 619-629.

Tzin, V. and Galili, G. (2010). The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana*. *The Arabidopsis book/American Society of Plant Biologists*, *8*, e0132.

Van Alstyne, K. L., Nelson, A. V., Vyvyan, J. R. and Cancilla, D. A. (2006). Dopamine functions as an antiherbivore defense in the temperate green alga *Ulvaria obscura*. *Oecologia*, *148*, 304-311.

Van Emden, H. (2009). Artificial diet for aphids-thirty years' experience. *Redia*, 92, 163-167.

Vavricka, C., Han, Q., Huang, Y., Erickson, S. M., Harich, K., Christensen, B. M. and Li, J. (2011). From L-dopa to dihydroxyphenylacetaldehyde: a toxic biochemical pathway plays a vital physiological function in insects. *PloS one*, *6*, e16124.

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S-PLUS. Fourth Edition.* Springer Science + Business Media, New York.

Via, S. (1991). The genetic structure of host plant adaptation in a spatial patchwork - demographic variability among reciprocally transplanted clones. *Evolution*, 45, 827-852.

Via, S. (1999). Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution*, 1446-1457.

Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology and Evolution*, *16*, 381-390.

Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, *106* (Supplement 1), 9939-9946.

Via, S. Bouck, A. C. and Skillman, S. (2000). Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution*, 54, 1626–1637.

Via, S. and Hawthorne, D. J. (2002). The genetic architecture of ecological specialization: correlated gene effects on host use and habitat choice in pea aphids. *The American Naturalist*, *159* (S3), S76-S88.

Via, S. and West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, *17*, 4334-4345.

Visnevschi-Necrasov, T., Faria, M. A., Cunha, S. C., Harris, J., Meimberg, H. W., Curto, M. A., Pereira, G., Oliviera, M. B. P. P. and Nunes, E. (2013). Isoflavone synthase (IFS) gene phylogeny in *Trifolium* species associated with plant isoflavone contents. *Plant Systematics and Evolution*, *299*, 357-367.

Vogel, K. J. and Moran, N. A. (2011). Sources of variation in dietary requirements in an obligate nutritional symbiosis. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*, 115-121.

Vranova, V., Rejsek, K., Skene, K. R. and Formanek, P. (2011). Non-protein amino acids: plant, soil and ecosystem interactions. *Plant and Soil*, *342*, 31-48.

Walling, L. L. (2000). The myriad plant responses to herbivores. *Journal of Plant Growth Regulation*, *19*, 195-216.

Walling, L. L. (2008). Avoiding effective defenses: strategies employed by phloem-feeding insects. *Plant Physiology*, *146*, 859-866.

Walters D. R. (2011). *Plant defense: warding off attack by pathogens, herbivores and parasitic plants.* Wiley-Blackwell, Chichester.

Webster, B., Bruce, T., Dufour, S., Birkemeyer, C., Birkett, M., Hardie, J. and Pickett, J. (2008). Identification of volatile compounds used in host location by the black bean aphid, *Aphis fabae*. *Journal of Chemical Ecology*, *34*, 1153-1161.

Webster, B., Bruce, T., Pickett, J. and Hardie, J. (2010). Volatiles functioning as host cues in a blend become nonhost cues when presented alone to the black bean aphid. *Animal Behaviour*, *79*, 451-457.

Wichers, H. J., Visser, J. F., Huizing, H. J. and Pras, N. (1993). Occurrence of L-DOPA and dopamine in plants and cell cultures of *Mucuna pruriens* and effects of 2, 4-d and NaCl on these compounds. *Plant Cell, Tissue and Organ Culture*, *33*, 259-264.

Wilkinson, T. L. and Douglas, A. E. (1998). Plant penetration by pea aphids (*Acyrthosiphon pisum*) of different plant range. *Entomologia experimentalis et applicata*, *87*, 43-50.

Will, T. and van Bel, A. J. (2008). Induction as well as suppression: How aphid saliva may exert opposite effects on plant defense. *Plant Signaling and Behavior*, *3*, 427-430.

Will, T., Furch, A. C. and Zimmermann, M. R. (2013). How phloem-feeding insects face the challenge of phloem-located defenses. *Frontiers in Plant Science*, *4*,336.

Will, T., Kornemann, S. R., Furch, A. C., Tjallingii, W. F. and van Bel, A. J. (2009). Aphid watery saliva counteracts sieve-tube occlusion: a universal phenomenon?. *Journal of Experimental Biology*, *212*, 3305-3312.

Will, T., Tjallingii, W. F., Thönnessen, A. and van Bel, A. J. (2007). Molecular sabotage of plant defense by aphid saliva. *Proceedings of the National Academy of Sciences*, *104*, 10536-10541.

Wink, M. (1992) *The role of quinolizidine alkaloids in plant-insect interactions*. In: Bernays, E.A. ed. *The role of quinolizidine alkaloids in plant-insect interactions.* CRC Press, Boca Raton, pp. 133-169.

Wink, M. (2003). Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry*, *64*, 3-19.

Wink, M. (2013). Evolution of secondary metabolites in legumes (Fabaceae). *South African Journal of Botany*, *89*, 164-175.

Wink, M. and Mohamed, G. I. A. (2003). Evolution of chemical defense traits in the Leguminosae: mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred from nucleotide sequences of the rbcL gene. *Biochemical Systematics and Ecology, 31*, 897-917.

Wojciechowski, M. F., Lavin, M. and Sanderson, M. J. (2004). A phylogeny of legumes (Leguminosae) based on analysis of the plastid matK gene resolves many well-supported subclades within the family. *American Journal of Botany, 91*, 1846-1862.

www.gbif.org/species/2077503, Accessed: 20/09/2015

www.genome.jp/kegg/pathway.html, Accessed: 20/09/2015

www.biocyc.org, Accessed: 20/09/2015

www.epgsystems.eu, Accessed: 20/09/2015

www.epgsystems.eu/downloads, Accessed: 20/09/2015

www.epgsystems.eu/epg/measuring-systems, Accessed: 20/09/2015

www.gbif.org, Accessed: 20/09/2015

Zacarés, L., López-Gresa, M. P., Fayos, J., Primo, J., Bellés, J. M. and Conejero, V. (2007). Induction of p-coumaroyldopamine and feruloyldopamine, two novel metabolites, in tomato by the bacterial pathogen *Pseudomonas syringae*. *Molecular Plant-Microbe Interactions, 20*, 1439-1448.